University of
Southern
Queensland

ENHANCED ARTIFICIAL INTELLIGENCE-BASED CYBERSECURITY FOR THE DETECTION OF CYBER FRAUD IN THE BANKING INDUSTRY

A Thesis submitted by

Eyad Abdel Latif Marazqah Btoush
(MIT)

For the award of

Doctor of Philosophy

2024

ABSTRACT

The financial and banking industries have acknowledged the significance of creating credit card cyber fraud detection systems for a considerable period. Despite these efforts, businesses are still struggling with an increase in credit card cyber fraud. These incidents are fuelled by the ongoing technological revolution, which relies heavily on key enabling technologies such as Artificial Intelligence (AI)/Machine Learning (ML), big data, cloud computing, and the Internet of Things (IoT). Cybersecurity has become a paramount concern in the banking industry owing to the widespread occurrence of breaches and crimes. The growing prevalence of cybersecurity data breaches has undermined the efficacy of existing cyber fraud detection systems in detecting intricate criminal activities. The challenge of detecting credit card cyber fraud is exacerbated by two factors: the absence of enriched datasets and class imbalance. Credit card cyber fraud is a significant cybersecurity concern for the banking system worldwide, particularly as the number of financial transactions using credit cards increases. Diverse methodologies have been used to counteract these threats. Conventional anomaly detection is frequently employed; however, it tends to be time-consuming, resource-intensive, and imprecise. Artificial intelligence has the potential to significantly improve the accuracy of cyber fraud detection. It is imperative to promptly utilise experimental approaches that integrate machine learning and deep learning (DL) techniques to identify fraudulent activities and conduct factor analysis on anonymised credit card data. This will facilitate a more comprehensive comprehension of the interrelationships among various features and provide invaluable perspectives for counteracting credit card cybercrime. This research assesses several machine learning and deep learning techniques that are routinely employed to address credit card cyber fraud and binary cybersecurity problems. We created and evaluated three innovative cyber fraud detection models, each specifically designed to enhance the accuracy and efficiency of credit card transaction cyber fraud detection algorithms. These include a novel hybrid ML model, a novel hybrid DL model (CNN-BiLSTM), and a novel hybrid ML+DL model. The experiments demonstrate that the novel hybrid ML model, combined with the stacking ensemble, outperforms other individual ML models in detecting credit card cyber fraud. It also achieved the highest performance in credit card cyber fraud detection compared to other individual ML models. In addition, the novel hybrid DL model (CNN-BiLSTM) surpasses other individual DL models in cyber fraud detection. Furthermore, the novel hybrid ML+DL model with a stacking ensemble surpassed the hybrid ML model, hybrid DL model, and all individual models. This research further presents a theoretical structure based on empirical evidence from the actual world. The innovative ensemble hybrid ML + DL model represents the highest level of cyber fraud detection capabilities, leading to improved security in financial transactions. In summary, the results of the experiments clearly demonstrate that the proposed models were capable of producing accurate outcomes for the detection of cyber fraud in credit card systems and could be implemented in banking systems. As prominent artificial intelligence tools that enable the more accurate detection of cyber fraud, the newly developed models make significant contributions to the banking industry. As a result, these methods could be implemented to safeguard financial transactions through the implementation of a more coherent, accurate, and efficient methodology.

CERTIFICATION OF THESIS

I Eyad Marazqah Btoush declare that the Thesis entitled *Enhanced AI-based Cybersecurity For The Detection Of Cyber Fraud In The Banking Industry* is not more than 100,000 words in length including quotes and exclusive of tables, figures, appendices, bibliography, references, and footnotes. The thesis contains no material that has been submitted previously, in whole or in part, for the award of any other academic degree or diploma. Except where otherwise indicated, this thesis is my own work.

Date: 12-08-2024

Endorsed by:

Prof. Xujuan Zhou
Principal Supervisor

Prof. Raj Gururajan
Associate Supervisor

Dr. KC Chan
Associate Supervisor

Student and supervisors' signatures of endorsement are held at the University.

ACKNOWLEDGEMENTS

I want to express my deep gratitude to the individuals and institutions whose unwavering support and contributions have been pivotal to the successful completion of this doctoral thesis.

Foremost, I extend my sincere thanks to my advisor, Prof. Xujuan Zhou, for the invaluable guidance, expertise, and dedication throughout this academic journey. Their mentorship has played a crucial role in shaping the research and my academic growth.

I am profoundly appreciative of the contributions of the supervisory team of my thesis, Prof. Raj Gururajan and Dr. KC Chan. Their insightful feedback, constructive criticism, and unwavering support have significantly enriched the quality of this work.

I would also like to acknowledge the faculty and staff at UniSQ for fostering an academic environment that encourages intellectual exploration and excellence.

My family, especially my wife and my sons, who have stood by me with unwavering support, love, and patience, deserves my deepest gratitude. Their understanding and encouragement have been a constant source of strength.

This research journey has been marked by both challenges and triumphs, and I am thankful to all those who have contributed to this endeavour. Your support and encouragement have been instrumental in bringing this thesis to fruition.

DEDICATION

To my beloved wife, my sons, and my entire family. Your support and encouragement have been a source of hope. Nouf, your patience has been my source of strength. Jawad and Ahmad, your understanding and cheerfulness have brightened my days. To my whole family, your collective support has made this achievement possible.

KEYWORDS

Fraud; Cyber fraud; Credit card fraud; Credit fraud; Machine learning; Deep learning; Cybersecurity; Banking industry; Financial transactions; Cybercrime; Supervised learning; Unsupervised learning; Fraud detection; Cyber fraud detection.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

NO	Abbreviation	Description
1	Adaptive Boosting	AdaBoost
2	AEs	Autoencoders
3	AI	Artificial Intelligence
4	ANN	Artificial Neural Network
5	AR	Action Research
6	BBN	Bayesian Belief Networks
7	BGRU	Bidirectional Gated Recurrent Unit
8	BiLSTM	Bidirectional Long Short-Term Memory
9	BPNN	Back-Propagation Neural Network
10	CNN-ELM	CNN with Extreme Learning Machine
11	CNNs	Convolutional Neural Networks
12	ConvNets	Convolutional Neural Networks
13	CS	Cuckoo search
14	DAEGAN	Dual Autoencoders Generative Adversarial Networks
15	DBN	Deep Belief Network
16	DBSCAN	Density Based Spatial Clustering of Application with Noise
17	DCL	Dilated Convolutional Layer
18	DCNN	Deep Convolutional Neural Networks
19	DL	Deep Learning
20	DNN	Deep Neural Network
21	DT	Decision Tree
22	ENN	Edited Nearest Neighbour
23	GA	Genetic algorithm
24	GAN	Generative Adversarial Network
25	GAORF	Genetic Algorithm with RF model Optimisation
26	GB	Gradient Boosting
27	GBM	Gradient Boosting algorithm
28	GRU	Gated Recurrent Unit
29	HAS	Harmony Search
30	HMM	Hidden Markov Model
31	HOBA	Homogeneity-Oriented Behaviour Analysis
32	HS	Hybrid Sampling
33	IForest	Isolation Forest
34	IoT	Internet of Thing
35	IT	Information Technology
36	KNN	K-Near Neighbour
37	LightGBM	Light Gradient Boosting Machine
38	LMT	Logistic Model Tree
39	LR	Logistic regression
40	LSTM	Long Short-Term Memory
41	ML	Machine Learning
42	MLP	Multi-Layer Perceptron Learning
43	NB	Naive Bayes

NO	Abbreviation	Description
44	NN	Neural Networks
45	NNHS	Harmony Search with ANN
46	PCA	Principal Component Analysis
47	PSO	Particle Swarm Optimisation
48	RBM	Restricted Boltzmann machine
49	RF	Random Forest
50	RNN	Recurrent Neural Networks
51	ROS	Random Oversampling
52	RUS	Random Undersampling
53	RUSBoost	Random Undersampling Boost
54	SAL	Similarity Attention Layer
55	SMOTE	Synthetic Minority Oversampling Technique
56	SOM	Self-organising map
57	SVM	Support Vector Machine
58	T-SNE	T-distributed stochastic neighbour embedding
59	TVIWDA	Time Varying Inertia Weight Base Dragonfly Algorithm
60	UQ	Uncertainty Quantification
61	VAE	Variational Autoencoder
62	VAEGAN	Variational Autoencoder Generative Adversarial Network
63	WFSVM	Weighted Feature Support Vector Machine
64	XGB	Extreme Gradient Boosting

CHAPTER 1: INTRODUCTION

1.1 Background

The technological revolution is developing rapidly owing to several key enabling technologies, such as Artificial Intelligence (AI), Internet of Things (IoTs), and big data. Given the widespread adoption of the ever-evolving Internet technology. Banks are implementing new technology and digital platforms to increase both their client base and revenue (Carbo-Valverde et al., 2020). The banking industry has greatly benefited from the development of technologies and has undergone considerable transformation. However, the rapid increase in technology usage has exacerbated cyber fraud using credit cards. Credit card cyber fraud refers to the act of committing illegal financial transactions by fraudsters as a means of identity theft.

Banking plays a critical part in financial transactions today, as people must engage with banks, either physically or electronically. The implementation of a banking information system has greatly enhanced the efficiency and financial gains. The majority of banking application system transactions are currently conducted using credit cards or online net banking (Patil et al., 2018), leading to increased vulnerability to new attacks and techniques. Thus, cyber fraud is far more serious in the banking industry (Mienye & Sun, 2023). As cybercrimes of increasing sophistication become increasingly prevalent in infiltrating digital lives, there is an imperative need to innovate credit card cyber fraud detection. This innovation should be built around the sophisticated techniques utilised in cybersecurity and firmly grounded in the fundamental concepts of safeguarding against cyber threats (Almarshad et al., 2023).

Credit card cyber fraud is a major challenge for the banking industry, resulting in billions of dollars in annual losses (Krishna et al., 2023). The detection of criminal activity can be enhanced by the implementation of credit card cyber fraud detection techniques, which also reduces the occurrence of false positives and false negatives. Since the early 2010s, large banks have detected deviations and abnormalities by using anomaly detection and traditional approaches. However, such tactics and technologies have changed over time. Several monitoring and detection systems have been devised to identify cyber frauds in credit card. However, as the threat landscape constantly evolves, it is essential to equip banks with smart and novel technologies for threat management (Seera et al., 2024). AI plays a crucial role in identifying credit card cyber frauds. Utilising AI approaches can enhance the probability of monitoring systems to accurately identify anomalies and mitigate possible threats (Ala'raj et al., 2021). AI techniques will enable comprehensive behavioural monitoring and individual profiling of consumers, which will track their behaviour and deliver actionable knowledge to help reduce cyber fraud and risks (Priya, 2021; Sarker et al., 2021). Machine learning (ML) techniques have been employed in the domain of credit card cyber fraud detection to facilitate data processing and analysis of data (Soni et al., 2021). Both unsupervised and supervised learning algorithms are employed in these approaches to identify and classify fraudulent transactions. The rapid transition towards a digital society has unavoidably led to a significant rise in fraudulent activities, reaching levels never experienced.

Current research has focused on employing sophisticated techniques aiming to facilitate the instantaneous examination of online transactions conducted by consumers. The transition encompasses sophisticated methodologies that enhance the efficacy of identifying fraudulent behaviours. Nevertheless, the escalating prevalence of fraudulent activities necessitates a progressively more comprehensive and astute examination of transactional data.

1.1.1 Credit card cyber fraud and cybersecurity

The banking industry has been significantly and profoundly impacted by the development of information technology (IT). However, since it progressed, several methods by which individuals might become victims of diverse attacks have also developed. Banks holding large amounts of client data have made themselves a prime target for hackers, and as a result, banks have been at the forefront of enterprise cybersecurity (Abadi et al., 2016). Cybercrime's meteoric rise to prominence in the twenty-first century is a primary threat to financial institutions. Cybercrime targeting the banking industry includes phishing, spoofing, identity theft, spyware, blackmail, and denial of service attacks (Creado & Ramteke, 2020). The attacker's purpose may be to harm the victims' reputations, create a political stir, or extract wealth from them.

In the past decade, the lack of cybersecurity has become more problematic in the banking industry. During this period, the cybersecurity market grew approximately 35 times. In 2019, the cybersecurity sector had an expenditure of around 40.8 billion. The expenditure on cybersecurity technology rose to 71.1 billion U.S. dollars in 2022 (Research & Department, 2023). In 2023, the expenditure on the cybersecurity sector amounted to approximately 80 billion U.S. dollars. Projections indicate that the market will surpass 87 billion U.S. dollars by 2024. The global expenditure on cybersecurity has been steadily rising since 2017.

Interestingly, the growth of the digital economy coincided with an increase in digital crime. The proliferation of online and mobile interactions has generated several avenues for attack, resulting in data breaches that pose a threat to both individuals and corporations. Based on the existing pace of expansion, it is estimated that the financial damage caused by cyberattacks will reach almost $10.5 trillion per year by 2025. This represents a threefold increase compared with the levels observed in 2015 (McKinsey & Company, 2022). Figure 1.1 shows spending on cybersecurity worldwide from 2017 to 2022 (Statista Research Department, 2023).

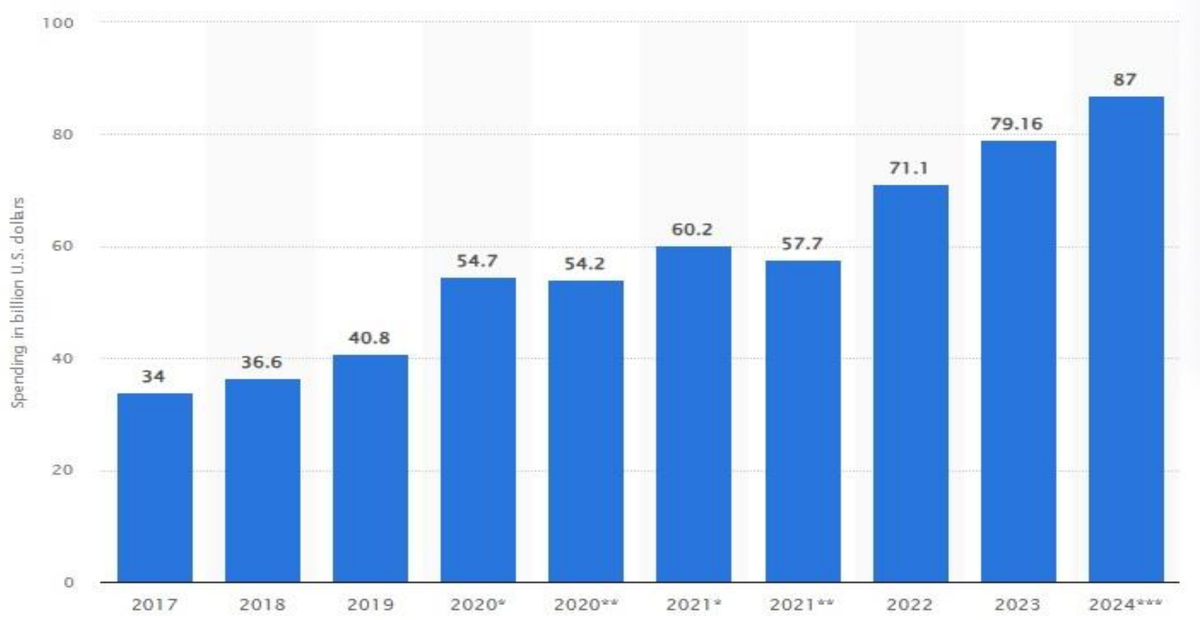


Figure 1.1 *Spending on cybersecurity worldwide from 2017 to 2024(in billion U.S. dollars). (Statista Research Department 2023)* [Cybersecurity spending worldwide 2024 | Statista]

Individual modes of payment have shifted dramatically as a result of advancements in modern technology. The use of online payment methods, such as online banking, debt cards, and credit cards, has grown. Credit card cyber fraud has grown extremely prevalent in the modern day, with several cases reported in recent years owing to the rise in cybercrime. A type of identity theft is credit card cyber fraud carried out when an individual other than the person making the payment uses their credit card or account information without authorisation. The term "cyber fraud" pertains to individuals engaged in cybercriminal activities, utilising technological resources to execute cyberattacks that lead to the compromise of card data. Subsequently, these compromised data may be employed for fraudulent purposes in either physical or virtual environments. Credit card cyber fraud may arise from counterfeit, neglected, or stolen cards. Credit card cyber fraud is one of the most serious hazards facing individuals and the banking industry worldwide, particularly as the volume of financial transactions involving credit cards continues to expand. Credit card cyber fraud is expanding at an alarming rate and has become a significant issue in the banking industry (Karthika & Senthilselvi, 2023).

Credit card cyber fraud occurs in various forms. The first type of cyber fraud is physically stealing a card. The second type of cyber fraud is theft of private credit card information (Btoush et al., 2021). A third cyber fraud occurs when a user uses his or her own credit card even though there is no money on the card, and the bank is required to pay by issuing a bill to the address. Further, cyber fraud is committed online when an item is purchased by inputting the card details of any credit card without notifying the owner (Trivedi et al., 2020).

The utility and accessibility of conducting commercial transactions on online platforms have been considerably improved by the emergence of digital transformation. Simultaneously, the advent of digital transformation has brought about a shift in the tactics employed by cybercriminals to exploit data, leading to a rise in supplementary risks associated with cybercrimes, namely, in the field of cyber fraud. While certain individuals engaged in cybercriminal activities prioritise the infiltration of networks and the disruption of Internet servers, others concentrate on the unauthorised access of extensive datasets that include confidential information, including credit card details, which may be exploited for personal gain or traded on digital platforms.

The credit card cyber fraud problem refers to the criminal use of credit cards to engage in fraudulent activities, such as the unauthorised purchase of products, services, or funds. Fraudsters can illegally appropriate credit cards, replicate credit card information, or assume control of individuals' credit card accounts (Bagga et al., 2020). Cybercrime refers to criminal activities perpetrated using digital technologies and the Internet. There is a wide array of cybercrimes, with a significant proportion of them being associated with the illegal use of credit cards. For example, those engaged in cybercriminal activities can employ social engineering techniques within the context of a social network, thereby convincing users to transmit funds over a platform by exposing their credit card information. Cybercriminals possess the ability to illegally acquire numerical data and use that information to engage in unauthorised online transactions. Fraudsters possess the ability to illegally breach the database of financial institutions or corporations, thereby acquiring and subsequently vending personally identifiable information, which may include sensitive credit card particulars.

Criminals in the field of cyber fraud are constantly enhancing their techniques for avoiding detection, and are now integrating novel methods to circumvent credit cyber fraud detection systems and other measures employed for cyber fraud detection (Itoo et al., 2021). In addition to heightening the probability of fraudulent activities, the cyber domain is also instrumental in protecting individuals and organisations from the threats associated with credit card cyber fraud. Effective utilisation of cyber defence measures can aid in the reduction of credit card fraud (Potula et al., 2023). By adopting this technique, it is possible to integrate credit card cyber fraud prevention software as an essential component within the current array of cybersecurity solutions. This integration serves the purpose of identifies instances of credit card cyber fraud and mitigates fraudulent transactions. Effective solutions are those that are capable of recognising trends and alerting users to potentially fraudulent transactions in advance, thereby preventing them from becoming problematic.

In the current era of digital technology, the prevalence of credit card cyber fraud has increased, necessitating immediate attention and the resolve of the challenges associated with its detection. The use of credit card cyber fraud detection systems can effectively mitigate the adverse consequences of financial losses incurred by fraudulent transactions (Abd El-Naby et al., 2023). Recognised or unrecognised are the respective classifications of fraudulent transactions. Fraudulent transactions, which refer to potentially illegal activities that have taken place previously, can be detected by rule-based systems that have been appropriately coded. This detection process often necessitates training systems that use both legitimate and fraudulent transaction data. The detection of unidentified fraudulent transactions typically poses more challenges, necessitating the training of the system on regular transactions to effectively recognise abnormal transactions and detect new instances of cyber fraud.

The growing popularity of credit and debit cards has significantly reduced the need for cash for in-person transactions, leading to a substantial revolution in payment processing. Nevertheless, the widespread use of credit cards, which currently stands at over 2.8 billion globally, has also raised the potential of fraudulent operations taking advantage of this trend. Significantly, the 2024 global credit card cyber fraud figures reveal that the United States accounts for 46% of such fraudulent activities. Projections suggest that the global total of credit card cyber fraud will increase to $43 billion by 2026. The latest statistics indicate a significant increase of 46% in credit card cyber fraud attempts compared to the previous year, underscoring the urgent requirement for enhanced security measures. There has been a substantial increase of 140% in the past three years in credit card cyber fraud assaults against online retailers in the US, with ecommerce being a prime target for criminals. These data highlight the urgent need for increased awareness and security measures to combat the growing problem of credit card cyber fraud, as the landscape of credit card transactions continues to change (Merchantcostconsulting, 2024) . Figure 1.2 illustrates the worldwide losses resulting from credit card cyber fraud.

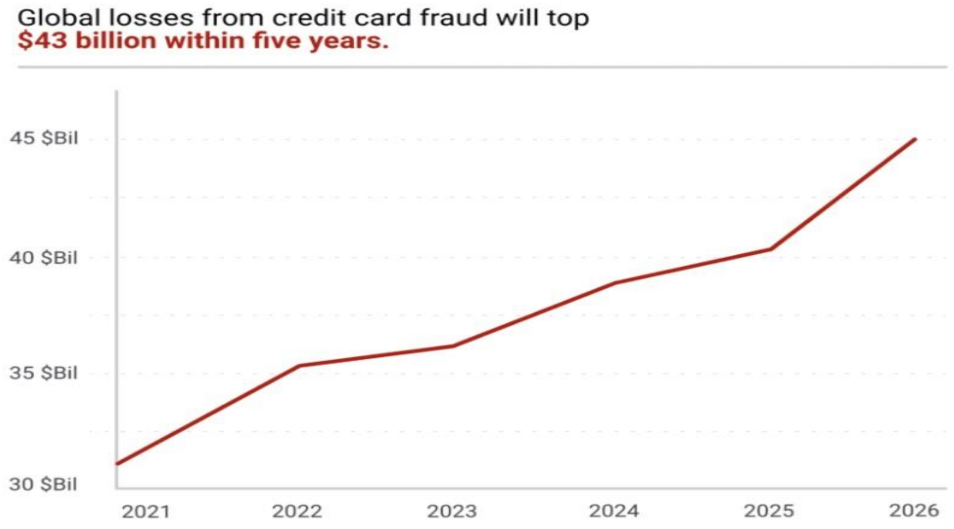


Figure 1.2 *Global losses from credit card cyber fraud.*

Cybercriminals perpetrate credit card cyber fraud through various methods. One such method is identity cyber fraud, which involves the impersonation of another individual to make unauthorised purchases using their credit card. Another method involves capturing data from physical credit cards during legitimate transactions. Cybercriminals frequently employ data breaches to obtain credit card information online. It is worth noting that data breaches have grown increasingly prevalent because of digital transformation. Cybercriminals access the computer networks of organisations, thereby obtaining unauthorised access to sensitive personal data. In addition to employing alternative tactics, cybercriminals employ other techniques to obtain credit card information, such as engaging in phishing attacks or deploying malware.

Threat attacks have seen a discernible metamorphosis in recent years, with the primary objective shifting from just interrupting services or generating cyber disorder to actively pursuing monetary gain. This movement has happened in the context of current times. Malicious software is designed to stealthily capture personal information and then transmit that information to those who are engaged in fraudulent activities.

In Australia, as reported by the Australian Bureau of Statistics, in 2022-23, an estimated 8.7% of persons (1.8 million) experienced card cyber fraud, which was higher than the rate in 2021-22 (8.1%). Figure 1.3 shows the Card cyber fraud victimisation rate, 2020-21 to 2022-23. The victimisation rate refers to the number of persons who experienced card cyber fraud. (Statistics, 2024).

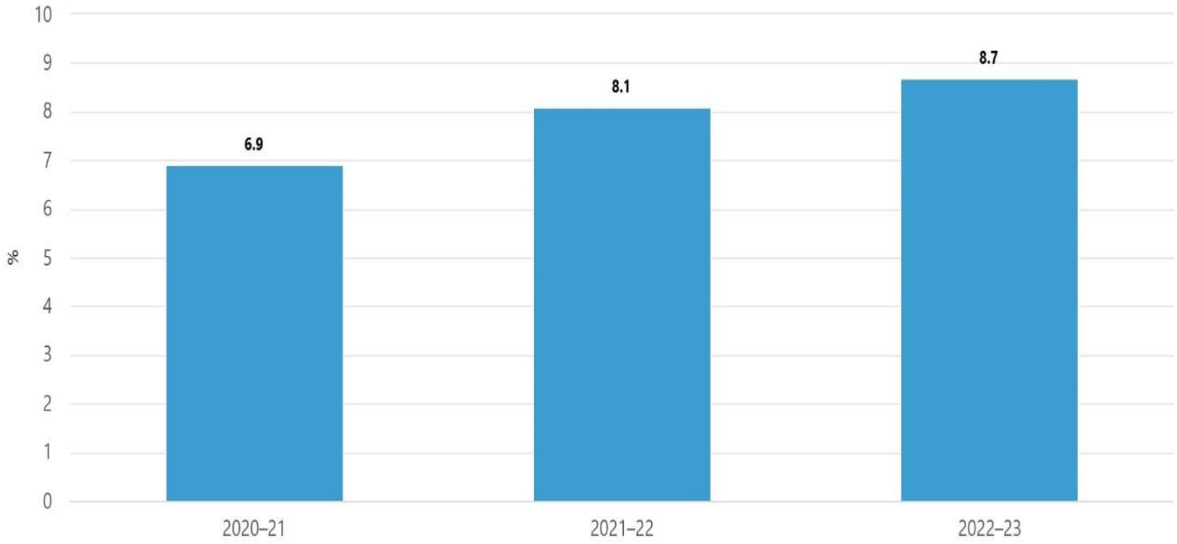


Figure 1.3 *Card cyber fraud victimisation rate, 2020-21 to 2022-23.*

Australian Bureau of Statistics 2022-23-financial-year, Personal Fraud, ABS, viewed 15 May 2024, https://www.abs.gov.au/statistics/people/crime-and-justice/personal-fraud/latest-release

In the most recent incident, almost all individuals who fell victim to card cyber fraud in 2022-23 reported the issue to an authority (98%), with banks or financial institutions being the most prevalent choice (92%). Figure 1.4 shows the reporting rate.

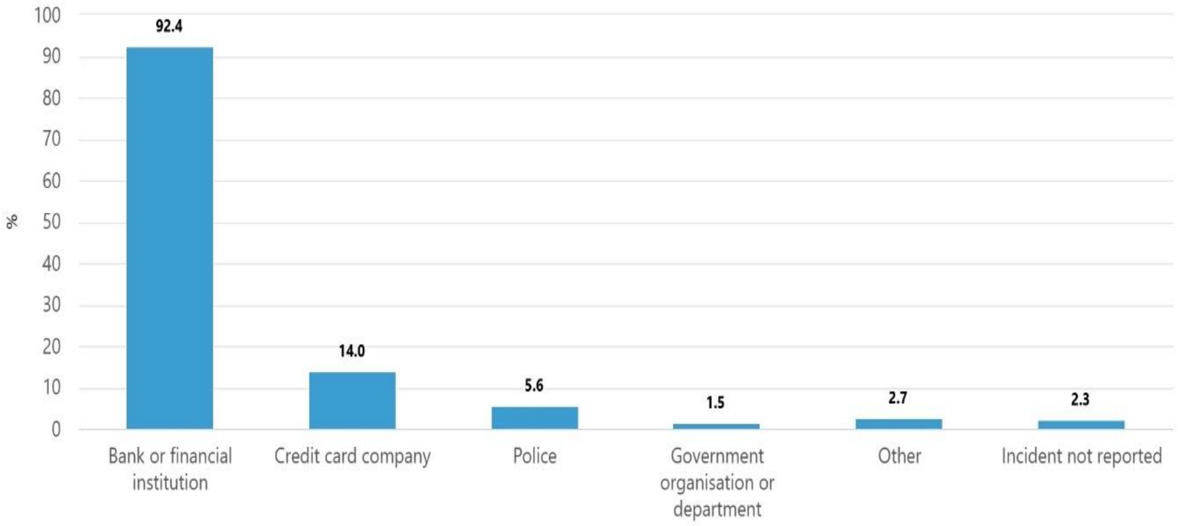


Figure 1.4 *Reporting rate.*

Australian Bureau of Statistics 2022-23-financial-year, Personal Fraud, ABS, viewed 15 May 2024, https://www.abs.gov.au/statistics/people/crime-and-justice/personal-fraud/latest-release

The incidence of card cyber fraud among Australians saw a substantial increase in 2023 compared to the previous year. According to a report from the Australian Bureau of Statistics (ABS), an estimated gross amount of $2.2 billion was lost due to card cyber fraud. According to this report, card cyber fraud refers to the illegal utilisation of credit, debit, or EFTPOS card information for the purpose of making illegal purchases or withdrawing cash. The proportion of Australians affected by card cyber fraud has increased from 6.9 percent in 2020-21 to 8.7 percent in 2022-23, with individuals aged 45-54 being the most frequently targeted. Figure 1.5 shows the change in Australians' affected by personal cyber fraud from 2020 to 2023.

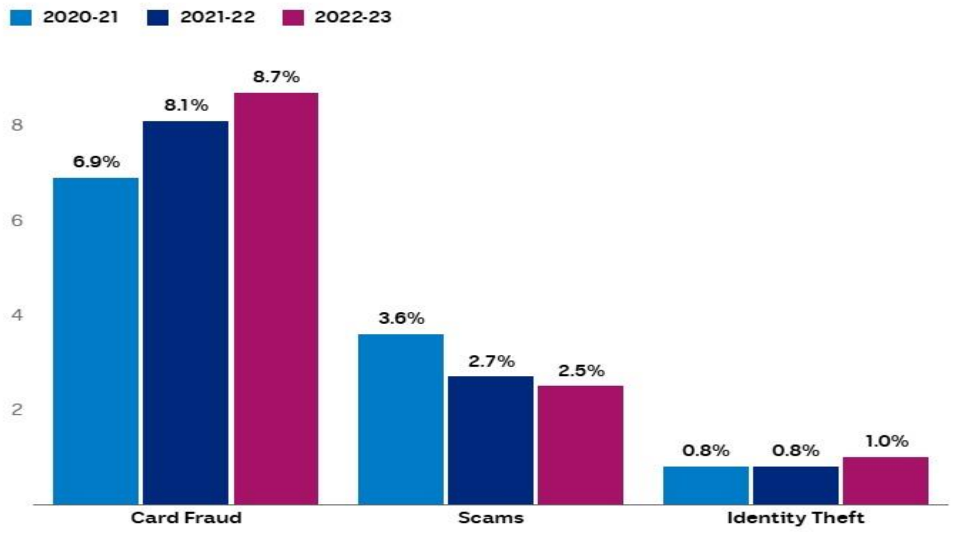


Figure 1.5 *Change of Australians affected by personal fraud.*

Australian Bureau of Statistics 2022-23-financial-year, Personal Fraud, ABS, viewed 15 May 2024, Australian Bureau of Statistics reveal details of 'sizeable' increase in card fraud as Australians lose $2.2 billion in 2023 - ABC News

Credit cards are widely used to ease financial transactions. Cybercriminals may attempt to compromise information systems that contain card data with the intention of using them for their own financial benefits. Figure 1.6 illustrates the rising incidence of Google searches linked to credit card cyber fraud and cybersecurity over the past 10 years. There is an expectation that the positive trend witnessed in the subsequent year will continue because of the growing importance of cybersecurity in addressing the challenges related to cyber fraud detection systems. In addition, the incorporation of innovative cybersecurity monitoring techniques into operational security systems is anticipated to provide a valuable contribution to current endeavours focused on mitigating credit card cyber fraud.

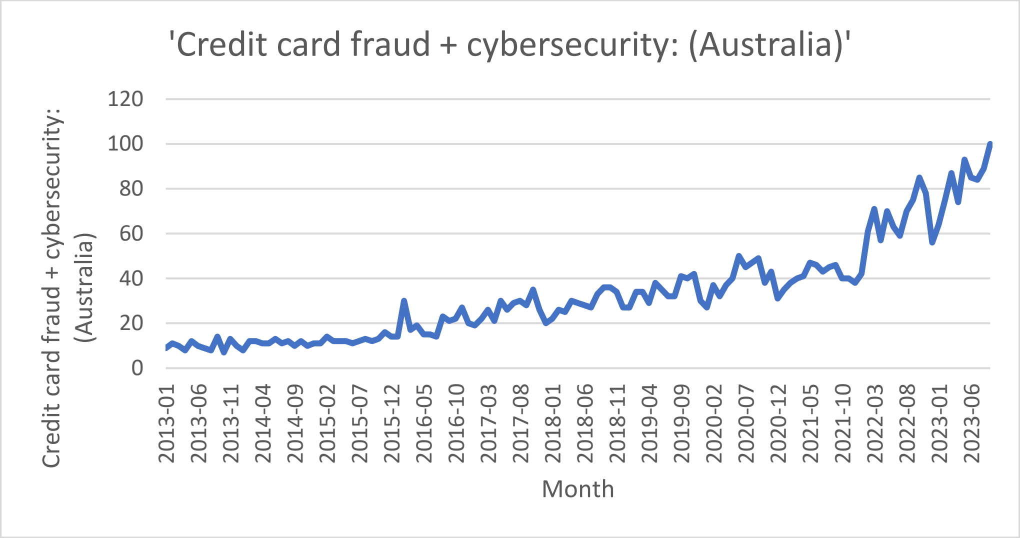


Figure 1.6 *Credit Card Fraud + Cybersecurity Term Trends in Australia*.

1.1.2 Credit card cyber fraud detection techniques

The primary objective of credit card cyber fraud detection is to identify anomalies that exhibit a low occurrence rate or demonstrate substantial deviation from most credit card transaction records (Khalid et al., 2024). Recent research has been particularly active in the area of credit card cyber fraud detection, as it is crucial to the expansion of critical sectors such as cybersecurity. Among other approaches, ML has been extensively explored in this context. The decreasing costs and increasing accessibility of infrastructure and computing capacity have facilitated the utilisation of ML techniques for the purpose of processing, analysing, and categorising instances of fraudulent behaviour within extensive datasets (Aftab et al., 2023).

Machine learning is an approach that is applicable to a wide range of challenges, particularly in industries that necessitate data analysis and processing. ML, which is classified as supervised, unsupervised, or reinforced ML, is essential for the resolution of unbalanced datasets (Btoush et al., 2023). ML approaches are highly effective in cyber fraud detection and prevention due to their ability to automatically identify forms across immense quantities of data. The differentiation between legitimate and fraudulent behaviour is facilitated by the implementation of appropriate ML models (Khan et al., 2022). Over time, these sophisticated systems may adjust to new, previously unknown cyber deception schemes. For this to be feasible, it is necessary to execute thousands of computations accurately within milliseconds. Both supervised and unsupervised technologies are essential for the detection of cyber fraud and must be incorporated into the next iteration of cyber fraud safeguards.

The training technique of supervised learning is used to train ML algorithms on labelled datasets. Data that is configurable and has known variable targets. Supervised learning encompasses classification, regression, and inference. The most prevalent ML techniques in all disciplines are supervised models that have been trained on a large number of precisely labelled transactions. Each transaction is classified as either legitimate or fraudulent. The models are trained by supplying numerous volumes of labelled transaction data to identify patterns that most closely resemble genuine behaviour. Unsupervised learning is the process of training a machine learning algorithm on a dataset that contains ambiguous target variables. The model endeavours to identify the most significant patterns in the data. Dimension reduction and cluster segmentation are among the unsupervised learning techniques. Semi-supervised learning integrates supervised and unsupervised learning with a training model for unlabelled data. The unsupervised learning attribute is employed in this method to ascertain the optimal data representation, while the directed learning attribute is employed to analyse the relationships within that representation and subsequently generate predictions. These techniques have been implemented in cybersecurity to investigate security vulnerabilities, including malware and phishing attacks. Furthermore, they may be employed to mitigate the constraints associated with conventional detection methods.

The identification among cyber fraud using credit cards is constantly evolving as a result of the dynamic nature of credit card characteristics. Several ML approaches have been employed to develop advanced detection systems for analysing credit card transactions. Nevertheless, the identification of credit card fraud poses a complex problem that has engrossed the domains of ML and AI due to several compelling factors. For example, credit card cyber fraud datasets are substantially skewed, as the number of fraudulent transactions is far smaller than the number of valid transactions. As a result, many standard classifiers fail to distinguish items belonging to minority classes in highly skewed datasets. On the other hand, credit card cyber fraud detection systems must be extremely responsive to real-world circumstances. Another crucial factor is the evolution of the data conditional distribution over time because of seasonality and new attack techniques (Benchaji et al., 2021). Various issues in the field of credit card cyber fraud detection must be addressed, including the massive amount of data that is stored on a regular basis, and the design architecture must be swift enough to react appropriately to cyber fraud (Al Rubaie, 2021). Unbalanced or misclassified data may also be a significant concern, as not all fraudulent conduct has been discovered or reported. Adaptive strategies employed by fraudsters against the system include real-time detection, abrupt changes in transaction type, and a high percentage of false positives. In addition, there is complexity, feature name obscurity, and real-world dataset scarcity (Adebayo et al., 2023; Khalid et al., 2024).

The prevalence of credit card cyber fraud is increasing substantially in parallel with advancements in modern technology; hence, the field of cyber fraud detection has become crucial. The aim of this study was to enhance the development of machine learning methods, specifically in the field of cybersecurity. We examined the use of machine learning algorithms in real-world datasets. Our methodology aims to reveal latent patterns inside a real-world dataset and make inferences about these concealed aspects to develop a framework for detecting credit card cyber fraud. The chosen research methodologies for addressing the issue of credit card cyber fraud detection are validated from an independent perspective. In order to differentiate between legitimate and fraudulent credit card transactions, the approach incorporates a number of machine learning techniques, taking inspiration from ways in which these techniques have been used in the realm of cybersecurity. Cybersecurity solutions have the ability to make use of the data from credit cards, which includes information such as the amount of the transaction, the time, and the location. Data were utilised by academics to determine the distinguishing features that indicate whether a transaction is genuine or fraudulent.

1.2 Research problem

Cyber fraud is becoming an increasingly serious problem in the banking industry (Diwan, 2021). Banks respond to client demands via a range of channels, including online, mobile, and web channels, all of which pose new vulnerabilities (Mohanty et al., 2023). In addition, banks hold a large amount of client data, making them a prime target for hackers. Criminals are developing novel forms of attacks that are difficult to detect using traditional cyber fraud detection systems (Ahmadi, 2023; Elluri et al., 2023; Sharma et al., 2024).

Cyber Fraud detection is critical for banks to avoid financial loss. Conventional anomaly detection and rule-based techniques are insufficient for detecting increasingly sophisticated threats (Aschi et al., 2022; Shabbir et al., 2022). Although numerous cyber fraud detection techniques are available, no cyber fraud detection systems have been able to deliver high efficiency (Al-Hashedi & Magalingam, 2021; El Hlouli et al., 2024; Nguyen et al., 2020; Patel, 2023). Thus, it is necessary to have a high-performance, adaptable, and flexible cyber-defence system. To achieve this goal, the banking industry is currently transitioning away from the traditional approach to cyber fraud detection via AI-based solutions owing to AI’s popularity, efficiency, and accuracy (Fang et al., 2021; Hassan et al., 2023; Mytnyk et al., 2023; Priya, 2021).

Banks require a system that is both adaptive and dynamic to detect cyber fraud. Recently, cybersecurity experts have begun to examine how AI can enhance cyber security in the banking industry. These experts spent 50.1 billion dollars on AI in 2020, and is predicted to spend 110 billion dollars by 2024 (Salameh & Lutfi, 2021). AI-based solutions for detecting and preventing cyber fraud can facilitate self-learning to discover complex hidden patterns (Priya, 2021).

Detecting cyber fraud using credit cards poses significant challenges due to many factors. The data on credit card cyber fraud exhibit significant skewness, with a far lower count of fraudulent transactions compared to legitimate transactions (Carcillo et al., 2021). The design architecture must be sufficiently quick to react appropriately to cyber fraud, and the vast quantity of data that is stored on a regular basis must be addressed, among other issues in the detection of credit card cyber fraud. Unbalanced data may also be a significant concern, as not all fraudulent conduct has been identified (Kalid et al., 2020; Khalid et al., 2024; Wang et al., 2021). Fraudsters' adaptive strategies necessitate real-time detection, precipitous changes in transaction type, and a high percentage of false positives in order to overcome the system (Chen et al., 2024). Thus, there is a need to create a model that fits well, predicts with more accuracy, and can also adapt to new cyber fraud patterns (Voican, 2021). ML and DL are the leading and most widely utilised technologies because of their numerous applications, low time requirements, and high accuracy of results (Janiesch et al., 2021).

Credit card cyber fraud detection is a crucial component of cybersecurity in the banking industry. With the rapid growth in financial transactions carried out using credit cards, the risk of credit card cyber fraud is increasing at an alarming rate. To effectively address this imminent threat, it is crucial to have a strong cyber-defence system that can adapt to the dynamic nature of credit cards. The field of credit card cyber fraud detection is continually changing as credit cards are constantly evolving. Although a range of ML approaches are utilised to create complex detection systems specifically designed for analysing credit cards, these methods face several problems. These obstacles involve problems such as the complexity of the data, large amounts of data, imbalanced data, unclear feature names, and a shortage of real-world datasets.

Worldwide, credit card cyber fraud is a significant and growing concern. Despite the active involvement of research communities in applying ML and DL methodologies to combat fraudulent activities, comprehensive and varied approaches to credit card cyber fraud detection (Hilal et al., 2022; Zioviris et al., 2022). From a cybersecurity perspective, it is crucial to acquire a thorough understanding of the fundamental parameters that regulate credit cards. This understanding forms the foundation for developing a robust system that continuously monitors, detects, and promptly reports instances of credit card cyber fraud. The constant urgency to improve detection rates is emphasised by the ever-changing nature of online transactions and the increasing difficulties in enhancing credit card cyber fraud detection strategies in the context of a surge in cyberattacks.

To effectively navigate this ever-changing environment, It is essential to develop and implement creative approaches that are specifically designed to detect credit cards. The prevalence of credit card payments in the context of online purchasing is on the rise, the focus of cybersecurity and information systems communities is progressively shifting towards the development and reinforcement of strategies aimed at mitigating credit card cyber fraud. Fundamentally, the domain of credit card cyber fraud detection requires continual attention as well as a constant search for novel, efficient approaches to proactively counter emergent risks within the constantly changing cybersecurity sphere.

This PhD thesis developed and addressed the following research questions:

1- What is the most suitable AI technique for extracting and selecting features from banking data to build a cyber fraud detection model?

2- What are the most effective and efficient ML models for detecting cyber fraud in the banking industry?

3- How effective and efficient are the developed models to detecting cyber fraud in banking transactions?

4- What patterns in banking transaction dataset can be identified that would indicate potential cyber fraud?

5- How can patterns be analysed to reduce false positives and increase cyber fraud detection efficiency and accuracy in the banking industry?

1.3 Aim of the PhD thesis

This PhD thesis aim to develop, validate, and evaluate an innovative machine learning and deep learning model designed to detect credit card cyber fraud. This model was developed and assessed to examine financial transaction data and to efficiently identify fraudulent activities. The research was centred on addressing the previously mentioned problems and enhancing the existing ML and DL techniques used in credit card cyber fraud detection.

To overcome the inherent difficulties of ML and DL credit card cyber fraud detection systems, this thesis proposes a variety of solutions. Through the utilisation of sophisticated statistical score metrics and the execution of visual analyses of the outcomes, the suggested advancements were thoroughly verified. The research encompasses an extensive examination of previous studies in the field of credit card cyber fraud detection, concluding with the development of a novel model intended to improve the accuracy and efficiency of financial transaction cyber fraud detection procedures. This thesis makes a significant contribution by applying and improving ML and DL techniques to the unique challenge of credit card cyber fraud detection. In doing so, it provides valuable insights and solutions for addressing fraudulent activities within the domain of financial transactions.

This research emphasises the relevance of cybersecurity in preventing credit card cyber fraud and the necessity to improve detection systems using ML and DL. Consequently, we propose a novel model that integrates the ML and DL techniques for credit card cyber fraud detection and prediction. The main objective of this Ph.D. thesis is to develop a novel model consisting of ML and DL algorithms to detect cyber fraud in the banking industry.

From a cybersecurity standpoint, it is important to have a comprehensive understanding of the primary credit card characteristics subject to surveillance. Considering the escalating frequency of cybersecurity attacks, This research recognises the obstacles that are linked to the enhancement of credit card cyber fraud detection systems. Previously, many separate models were developed to tackle credit card cyber fraud detection using machine learning techniques. Nonetheless, the complex purchasing patterns of credit card customers, and the issue of class imbalance have impeded machine learning classifiers from attaining optimal performance. The pressing need now is to create a hybrid model that effectively addresses these challenges, aiming to minimise the risks linked to fraudulent transactions through skilled classification and prediction (Karthik et al., 2022; Mienye & Sun, 2023). Academics are currently conducting research on the detection of credit card cyber fraud, which is a substantial concern. In order to mitigate cyber fraud, they implemented machine learning and deep learning methodologies. However, there is a scarcity of research that employs multiple approaches. We will investigate novel approaches for enhancing the accuracy of credit card detection.

In this research, we investigated novel approaches aimed at enhancing the performance of credit card cyber fraud detection. A composite approach has been used in previous studies, which incorporates a robust deep learning approach that includes long short-term memory (LSTM), gated recurrent units (GRU), and neural networks (NNs) as base learners in a stacking ensemble framework. A multilayer perceptron (MLP) is used as the meta-learner (Mienye & Sun, 2023). Meanwhile, the hybrid synthetic minority oversampling technique and edited nearest neighbour (SMOTE-ENN) method are employed to balance the class distribution in the dataset, with the aim of enhancing the rate of cyber fraud detection. A hybrid model proposed by Esenogho et al. (2022) is an alternative hybrid model that uses a neural network ensemble classifier and a hybrid data resampling method. The ensemble classifier was obtained using the LSTM neural network as the base learner in the adaptive boosting (AdaBoost) technique. Hybrid resampling was achieved using the SMOTE-ENN method. This technique combines many ML classifiers to enhance prediction accuracy. Alarfaj et al. (2022) proposed a hybrid model that includes several ML and DL classifiers with three architectures based on a convolutional neural network (CNN) to improve cyber fraud detection performance and reduce cyber fraud losses.

In light of the recent trend toward the integration of hybrid or combined frameworks to improve the accuracy of cyber fraud detection, we have also developed a new hybrid learning models. In order to mitigate the effects of class imbalances on credit card datasets, these models integrate the capabilities of ML and DL.

This research endeavour, conducted to address the problems outlined in the preceding section, yields the following objectives:

1- To critically review the literature on cybersecurity, ML, DL, AI, cyber fraud, and cybercrime in the banking industry.
2- To investigate the relationship between AI, cybersecurity, and cyber fraud detection in banking. To examine the differences and similarities between ML and DL in the detection of cyber fraud.

3- Build multi-level novel models to address the issue of class imbalance, increase the detection rate of cyber fraud, and decrease false positive rates based on convolutional neural networks and recurrent neural networks.

4- Training and testing a novel model to detect and indicate potential cyber fraud in the banking industry.

5- To evaluate a novel model for detecting cyber fraud in the banking industry by applying measurement techniques and comparing it with baseline models.

1.4 Significant contributions

In the rapidly evolving digital landscape, the banking industry is facing an increasingly urgent need to enhance the security of credit card services. This pressing need underscores the critical necessity for banks to fortify their information systems using the latest technological advancements. As banks expand their information security infrastructure, cybersecurity has emerged as a pivotal enabler of seamless business processes, particularly those involving credit card transactions.

Cybersecurity plays a paramount role in augmenting the security posture of the banking industry processes related to credit card processing (Darem et al., 2023; Patel, 2023). Positioned at the forefront of selecting information systems and merchant partners for financial transactions, banks must implement robust security solutions to safeguard their processes. The challenge for banks is the strategic deployment of technology that is in alignment with specific business functions, which requires an in-depth understanding of the business landscape. It is imperative to maintain ongoing endeavours to identify areas for improvement, thereby safeguarding the banking industry's financial stability.

It is imperative to employ ML techniques in cybersecurity contexts to detect anomalous activities, as financial transactions are predominantly conducted online, particularly credit card processing. The objective of this research is to develop a credit card cyber fraud identification system by utilising ML and DL methods. In this context, Credit card transaction history must be included to cybersecurity logs and events to categorise transactions. Data-driven machine learning to understand credit card cyber fraud is another highlight of this research. An exploratory model that explains data variables' associations is proposed to give a new viewpoint on credit card cyber fraud detection.

In spite of the fact that machine learning techniques are effective in identifying instances of credit card cyber fraud, there is an immense disparity in the understanding of the characteristics that are obscured in real-world datasets, as shown by research that are now in existence. The efficacy of individual ML methods was the primary focus of previous studies, which neglected the nuances of credit card cyber fraud features in actual datasets. The ongoing surge in data breaches and credit card cyber fraud incidents, attributed to the evolution of advanced technologies, underscores the need to address these challenges (Alamri & Ykhlef, 2024; Cherif et al., 2023; Ileberi et al., 2021).

This research aims to fill this gap by developing a model that leverages ML and DL algorithms to predict and detect cyber fraud with heightened accuracy and reduced false rates. The anticipated contributions of this research extend to the credit card cyber fraud detection literature, providing valuable insights into understanding and mitigating cyber fraud risks. If adopted by the banking industry, the proposed AI-driven model promises enhanced efficiency, reduced time consumption, and increased accuracy in detecting and safeguarding transactions against cyber fraud.

The financial industry is significantly aided by deep learning algorithms in the detection of credit card cyber fraud with greater accuracy and efficacy. In order to improve their estimation of cyber fraud detection, expedite client service procedures, and safeguard clients from cybercrime and fraudulent activities, banks necessitate an AI tool that is secure, reliable, time-saving, and accurate.

ML and DL approaches have become mainstream techniques with the capacity to handle various nonlinear modelling tasks, particularly in the banking and financial sectors, involving classification and feature extraction from complex datasets. Transfer learning and pretraining techniques for DL have shown success in feature extraction. Although the effectiveness of DL models has been demonstrated in various domains, obtaining an accurate algorithm to detect cyber fraud at multiple levels remains challenging and requires improvement (Chen et al., 2024; Cherif et al., 2023; El Hlouli et al., 2024; Khalid et al., 2024).

Khalid et al. (2024) argued that investigating the feasibility of integrating DL models into a system is a valuable endeavour. By integrating traditional ML techniques with the investigation of CNNs and Recurrent Neural Networks (RNNs), cyber fraud detection solutions that are more accurate and adjustable can be generated. The authors of this research also propose that additional research be conducted on techniques that seek to increase the resilience of the proposed model against novel or adversarial attacks. Adversarial attacks possess the capability to capitalise on weaknesses present in ML models; therefore, investigating methodologies that mitigate these threats would be of enormous value. Future studies may evaluate the model's capacity to scale in terms of managing larger datasets and satisfying increasing computational requirements.

This research work aims to address the issues outlined in Chapter 1, Section 2 and respond to the research questions, resulting in the following novel contributions:

1. A novel feature extraction method was developed to extract the most significant features by utilising the Principal Component Analysis (PCA) dimension reduction method and other techniques, including Random Forest (RF), XGBoost, and Permutation Feature Importance, which are based on transfer learning.
2. Three novel and effective hybrid models, the hybrid ML model, hybrid DL, and hybrid ML+ DL model, were designed to automatically detect fraudulent activity.
3. The individual ML and DL models extend into a newly developed hybrid and stacked ensemble technique, proving to be a more effective and accurate cyber fraud classifier than other models.
4. Cyber fraud detection frameworks can be easily implemented as AI algorithms in banking industry platforms, such as mobile applications or web portals, to autonomously regulate the level of fraudulent activity and be practically applicable to cyber fraud detection tasks.

1.5 Thesis organisation

This thesis is organised as followings:

Chapter 1 presents the background and problem statements of the present research. The chapter begins with an introductory section on the topic of credit card cyber fraud detection, emphasising the importance of an effective system for detecting such fraudulent activities. Furthermore, the research delves into the numerous challenges and issues associated with credit card cyber fraud detection while also identifying the research topics that will be addressed in this thesis. The objectives and novel contributions of this thesis are comprehensively outlined.

Chapter 2 provides a thorough examination of the current literature on the use of ML and DL techniques to detect credit card cyber fraud. The research encompasses several aspects, including dataset pre-processing feature extraction techniques, and classifiers employed in the ML and DL approaches. The results presented in this chapter offer insights into addressing research question 2. Additionally, they contribute to the identification of issues, development of improved ML and DL models for credit card cyber fraud detection, and identification of appropriate datasets for training and evaluating the suggested algorithms.

Chapter 3 discusses the research methodology devised in this thesis. Research methodology refers to the systematic approach employed to address a specific topic and encompasses a detailed account of the procedures and techniques used in conducting the research. The research technique employed in this thesis integrates a scientific approach and action research to accomplish the research objectives effectively. This research presents a framework for identifying cyber fraud in credit card transactions. The framework illustrates several stages and their interconnections. This chapter provides an explanation of the datasets used for training and testing the suggested algorithms as well as the evaluation metrics and validation methodologies utilised. In addition, the experimental settings are described.

Chapter 4 introduces a novel hybrid stacking ML model for the detection of cyber fraud in credit card transactions. This chapter provides a detailed explanation of the proposed model and its constituent elements. It covers the data pre-processing techniques employed in the model, the newly devised feature extraction algorithm, and recently developed ML and DL classifiers. Furthermore, this chapter discusses the results obtained from the model and compares them with the outcomes of state-of-the-art models.

In Chapter 5, a novel hybrid CNN-BiLSTM Model is introduced to detect cyber deception in credit card transactions. An exhaustive account of the proposed Hybrid DL models and their constituent components is provided in this chapter. Furthermore, this chapter discusses the results obtained from the models and compares them with the outcomes of state-of-the-art models.

Chapter 6 explains the novel hybrid stacking ML+DL model using a stacking ensemble, which consists of suggested models designed to detect instances of cyber fraud. This chapter expands on the model given in Chapters 4 and 5 by employing ensemble ML and DL techniques to assess its efficacy in the context of cyber fraud detection. The findings indicate a high degree of efficiency when compared with both state-of-the-art models and other baseline models. This section details the outcomes that were collected as well as the assessment and comparison of these data.

Chapter 7 provides a summary of the findings of this thesis, including the conclusions drawn, the limitations encountered, and potential areas for further research. Figure 1.7 shows the structure of the subsequent research content and the relationship between the chapters.

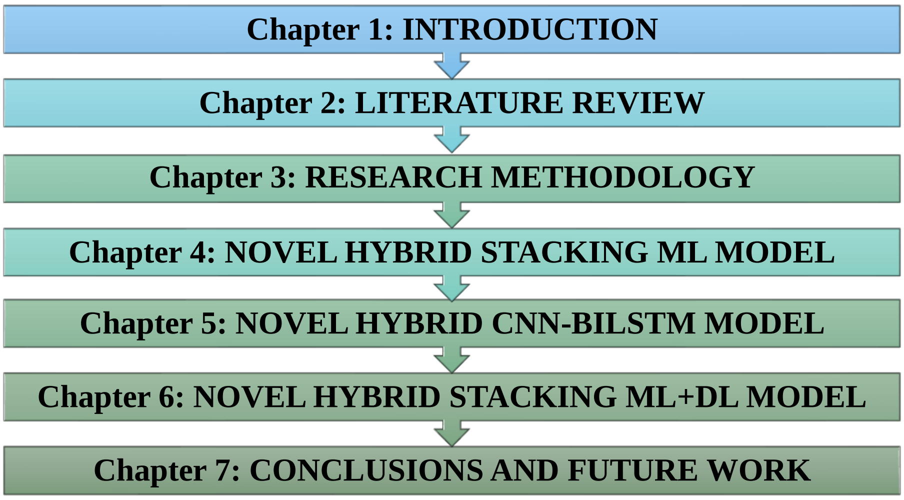

Chapter 1: INTRODUCTION
Chapter 2: LITERATURE REVIEW
Chapter 3: RESEARCH METHODOLOGY
Chapter 4: NOVEL HYBRID STACKING ML MODEL
Chapter 5: NOVEL HYBRID CNN-BILSTM MODEL
Chapter 6: NOVEL HYBRID STACKING ML+DL MODEL
Chapter 7: CONCLUSIONS AND FUTURE WORK

Figure 1.7 *Thesis structure.*

CHAPTER 2: LITERATURE REVIEW

In this chapter, the essential background and current state of the methods are presented and reviewed. This review is organised into two main sections. Section 1 explores the topic of credit card cyber fraud. Section 2 examining credit card cyber fraud with cybersecurity and gives relevant research conducted in this field, which facilitates the identification of appropriate algorithms and the creation of models. Additionally, this research examines the literature related to the application of ML and DL techniques in the detection and prevention of credit card cyber fraud.

2.1 Credit card cyber fraud

Cyber fraud is a criminal act characterised by the intention to steal funds. Cyber fraud may be characterised as the intentional exploitation or theft of an organisation's resources or assets by an individual in a position of authority, with the aim of personal gain. Cyber fraud, as a criminal act, exerts a significant influence on several aspects of the economy, legal framework, and individuals inside the societal fabric. Considerable resources have been dedicated to mitigating fraudulent activities within the financial sector, resulting in the development of several techniques aimed at preventing and detecting instances of cyber fraud. These techniques are incorporated into cyber fraud prevention systems and cyber fraud detection systems . Neither cyber fraud prevention systems nor cyber fraud detection systems appear to be formally defined terminology. Rather, they are broad phrases employed by the industry to include various processes aimed at preventing and detecting fraudulent activities.

The cyber fraud prevention system serves as the initial line of defence and is implemented to safeguard technological systems from fraudulent activities. The techniques present in this layer serve to limit, suppress, dismantle, obliterate, regulate, eliminate, or hinder the manifestation of cyber-attacks in hardware and software systems. Examples of these techniques include firewalls, encryption methods, and electronic signatures. Transitioning to the cyber fraud detection system, the subsequent layer is responsible for actively identifying and detecting fraudulent activities upon entry into the system. Two primary techniques are employed in cyber fraud detection systems: anomaly based cyber fraud detection and misuse-based cyber fraud detection (Abdallah et al., 2016; Arfeen & Khan, 2023; Omair & Alturki, 2020).

Anomaly based cyber fraud detection involves the examination of individual client behaviour, wherein the system is designed to detect and alert to potential instances of fraudulent activity when deviations from regular behaviour patterns are observed. This methodology is based on data mining, a process that encompasses the application of statistical, mathematical, AI, and DL methodologies to extract potentially significant insights from extensive databases. In contrast to anomaly based cyber fraud detection, misuse-based cyber fraud detection is characterised by a fixed setup. Fraudulent behaviours are pre-learned by the system, resulting in consistent behaviour across all customers. This implies that if a client engages in a behaviour that has been previously taught or is deemed analogous, the system will provide an alert indicating a potentially fraudulent activity. This technique uses rule-based statistics to identify potentially suspicious events. The protection system mechanisms for countering cyber-based fraud are shown in Figure 2.1.

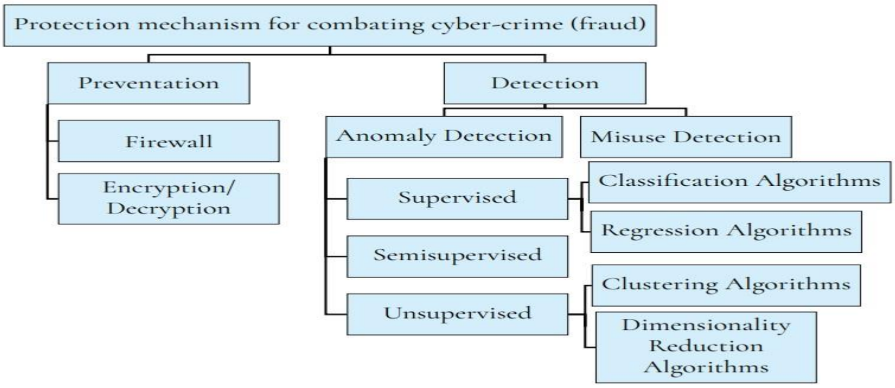


Figure 2.1 *Protection system mechanisms for combating cyber based fraud*. (Abdallah et al., 2016).

The prevalent approach to cyber fraud detection involves the utilisation of pattern-recognition techniques. These patterns encompass several aspects of the cardholder's behaviour, including the transaction amount, time elapsed since the last purchase, and specific day of the week, among others. When evaluating new spending behaviour, it is customary to compare it with an individual's past patterns. If the behaviour appears to deviate from the cardholder's established profile, it is deemed suspect (Al-Hashedi & Magalingam, 2021).

The utilisation of misuse detection systems is seldom due to the dynamic nature of both legal and fraudulent behaviour. However, proponents contend that anomaly-based cyber fraud detection systems are more prevalent for three primary reasons: There are many key points to consider in relation to cyber fraud detection: (i) Cyber fraud patterns may be derived automatically from data. (ii) By extracting patterns of fraudulent behaviour, the system can provide alerts when similar behaviour is discovered. This allows for the prioritisation of potentially fraudulent activities. (iii) The system has the capability to identify previously undetected fraudulent behaviours.

Cyber fraud detection systems are frequently developed using ML techniques, including supervised learning, unsupervised learning, and semi-supervised learning. The techniques are further elaborated upon in Chapter 2, Section 2.1.

2.1.1 Areas of cyber fraud

The occurrence of cyber fraud is almost common in technology systems that entail monetary transactions. Moreover, researchers have conducted a comprehensive mapping of the prevalent domains associated with fraudulent activities, including banking, insurance, telecommunication, and Internet marketing cyber fraud, as shown in Figure 2.2 (Abdallah et al., 2016).

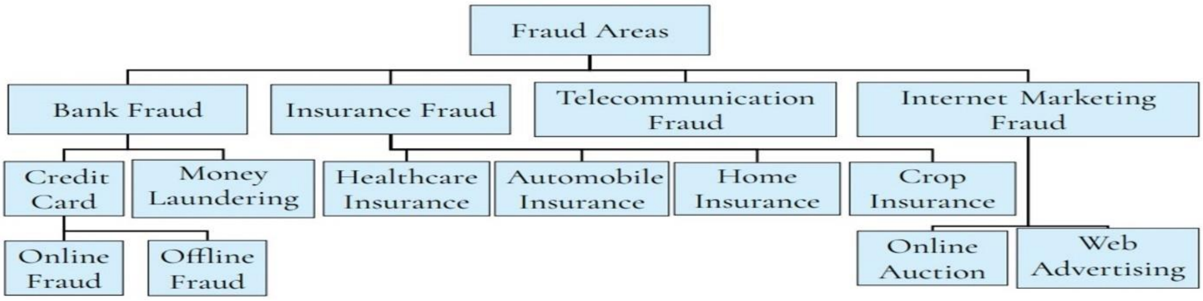


Figure 2.*2 The most common areas of fraud.*

Among the four primary domains depicted in Figure 2.2, bank cyber fraud has garnered the most extensive scholarly investigation (Abdallah et al., 2016; Hilal et al., 2022).

In contemporary society, credit cards have gained significant popularity because of their inherent convenience in comparison to physical currency as well as their ability to facilitate the monitoring and recording of expenditures. Despite the availability of several authorisation procedures, such as signatures, credit card numbers, identity numbers, and cardholder addresses, it remains feasible for fraudulent individuals to compromise the security of their cards. Credit card cyber fraud can be classified into two distinct categories: online credit card cyber fraud, wherein fraudsters acquire and subsequently utilise stolen credit card information, and offline credit card cyber fraud, wherein the cardholder's personal details are obtained through methods such as skimming, site cloning, credit card generators, or phishing. The primary objective of this thesis is to examine the phenomenon of bank cyber fraud, with a specific emphasis on online credit card cyber fraud.

2.1.2 Challenges with cyber fraud detection systems

Working with cyber fraud detection systems presents several complex difficulties, typically encompassing two primary areas of discussion: there are two main factors that need to be considered in this context. First, the class distribution is skewed, meaning that the distribution of data points across different classes is unbalanced. Second, there was a substantial volume of data available for the analysis.

A. The data has a skewed distribution

The issue of skewed class distribution is a significant challenge encountered by cyber fraud detection systems and pertains to the presence of significantly unbalanced data (Cherif et al., 2023; Mittal & Tyagi, 2020; Rtayli & Enneya, 2020). There is a substantial discrepancy between the number of fraudulent and non-fraudulent transactions in the dataset, which includes credit card transactions. This discrepancy results in a skewed dataset characterised by substantial variations in the quantity of data points for each class. The dataset included in this research had a significant degree of skewness. The presence of skewed datasets poses challenges for accurately and effectively detecting fraudulent transactions (Muaz et al., 2020; Unogwu & Filali, 2023). This phenomenon occurs primarily because of the use of a classification model, commonly referred to as a classifier, for imbalanced data. The consequences of this action may manifest as an uneven distribution of performance across several classes, or in extreme cases, the classifier may entirely disregard the minority class. There are two primary techniques to address this problem: data-level and algorithmic-level approaches (Karthik et al., 2022; Krawczyk, 2016; Singh et al., 2022).

Commencing with data-level methodologies, the procedure involves pre-processing the dataset to enhance its balance before inputting it into the classifier (Saraf & Phakatkar, 2022). Consequently, this facilitates the classifier's accurate classification of the distinct classes. Moreover, techniques at the data level may be categorised into two subgroups: undersampling and oversampling (Kraiem et al., 2021; Taha et al., 2021; Yuan et al., 2023).

The undersampling technique involves reducing the size of the majority class by eliminating a portion of its observations until the dataset achieves a balanced distribution (Alam et al., 2020; Alamri & Ykhlef, 2024). As illustrated in Figure 2.3, It is most effective when the dataset is enormous, and the performance and storage issues can be substantially improved by reducing the preponderance of samples. Random Undersampling (RUS) and Direct

Undersampling. The RUS technique involves random removal of data instances belonging to the majority class. To achieve a balanced dataset, plurality samples were eliminated randomly using this method. When training data is of substantial significance, this methodology is most advantageous. By decreasing the frequency of most samples, both latency and storage issues can be improved. Nevertheless, employing this methodology may result in the omission of valuable data during the elimination of most of the samples. Consequently, the accuracy of the classifier prediction may be compromised. In the context of direct undersampling, the removal of data from the majority class is not conducted randomly, resulting in the knowledge of which observations are deleted (Ghaleb et al., 2023; Leevy et al., 2023). Tomek Link Removal consists of two instances of distinct classes that are the closest Neighbours to one another. A Tomek Link is a pair (E1, E2) if given two instances E1 and E2 from independent classes, there is no sample E3 such that the distance from E1 to E3 is shorter than the distance from E1 to E2 or E2 to E3. Eliminating Tomek links can be regarded as an undersampling technique in which the plurality of samples contained within the link is removed (Pereira et al., 2020).

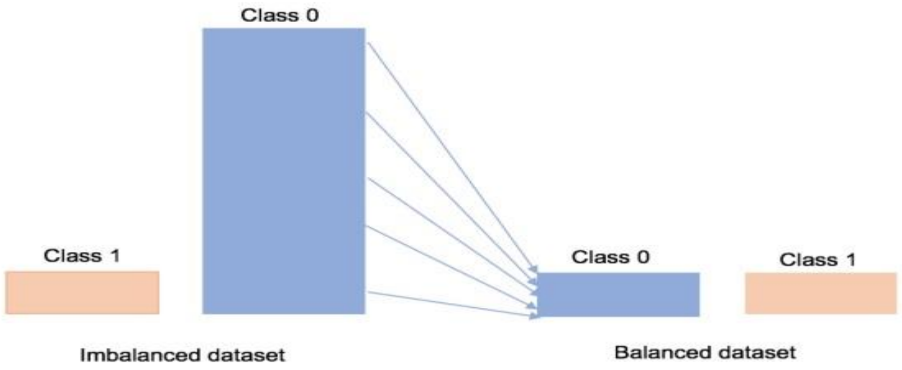


Figure 2.3 *Undersampling technique balances class sizes by reducing the majority class to match the minority class, resulting in a balanced dataset.*

The oversampling strategy increases the representation of the minority class by generating additional observations for that specific class. Nevertheless, the act of balancing the dataset in this manner may lead to the model being excessively tailored to the minority class owing to the up sampling of this class, resulting in an increased number of data points. Figure 2.4 shows the Oversampling technique. The Synthetic Minority Over Sampling Technique (SMOTE) is a widely employed method for addressing class imbalances in datasets (Dang et al., 2021; Douzas et al., 2018; Elreedy et al., 2023). The objective was to generate synthetic cases for the minority class by employing interpolation techniques among the nearest minority examples, as shown in Figure 2.5, instead of oversampling with replacement. Consequently, this mitigates the issue of overfitting the training data. The selection of nearest Neighbours for minority cases is conducted randomly, with the number of Neighbours chosen depending on the desired level of oversampling (Mansourifar & Shi, 2020; Muaz et al., 2020; Ramisetty et al., 2022). The limitation of SMOTE is that it generates synthetic data points without considering the characteristics of the majority class (Elreedy et al., 2023). The presence of overlapping classes may result in unclear examples. In Random Oversampling (ROS), Minority samples were replicated at random to achieve a balance in the dataset. Overfitting the data is a significant possibility because of the replication of the minority samples.

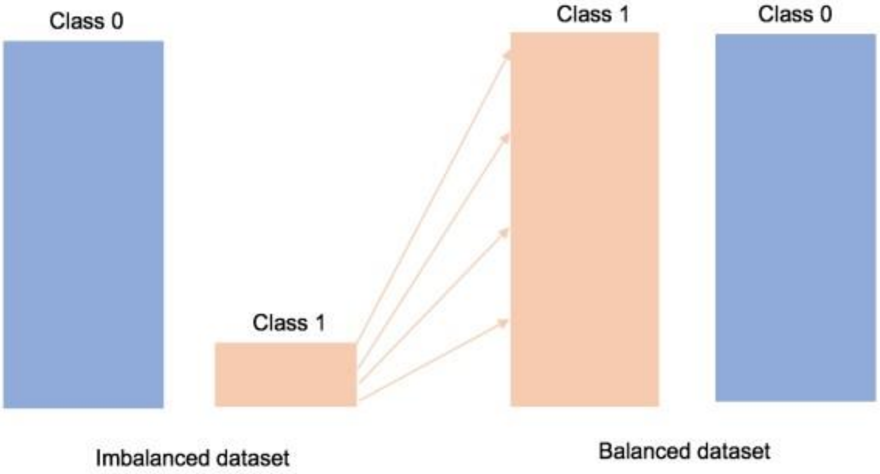


Figure 2.4 *Oversampling technique balances class sizes by duplicating instances from the minority class, resulting in a balanced dataset.*

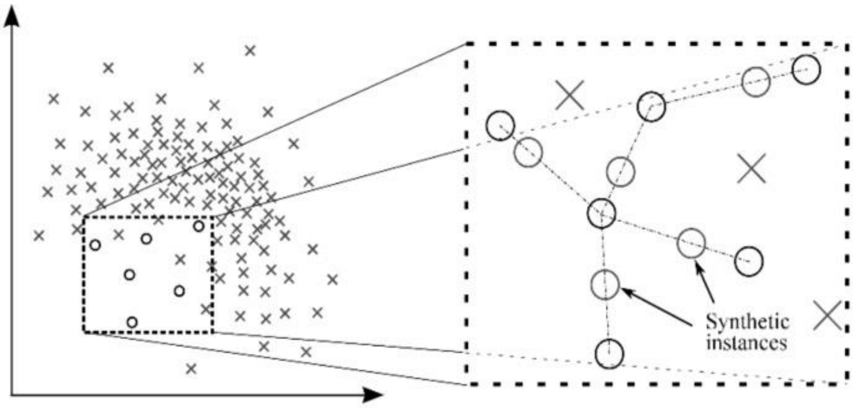


Figure 2.5 *SMOTE generates synthetic examples by interpolating between minority class instances*.

Combining SMOTE and Tomek Link removal, SMOTE effectively addresses dataset class imbalances. Nevertheless, throughout the process of generating novel synthetic instances for the minority class, there is a possibility that the cluster representing the minority class may encroach onto the region occupied by the majority class. The inclusion of such data in a model has the potential to result in overfitting. Therefore, to address this issue, it is recommended to employ both SMOTE and the Tomek Link elimination technique to obtain a balanced distribution of classes. The initial training dataset underwent oversampling using SMOTE to address the class imbalance. Subsequently, Tomek Link elimination was used on the oversampled dataset to provide a balanced dataset (Rtayli & Enneya, 2020).

The data-level method was examined, wherein resampling techniques were employed to rectify the imbalances in class distributions. This section examines the ensemble approach. The ensemble technique involves adapting pre-existing classification algorithms to address imbalanced class distribution. Broadly, an ensemble method refers to a learning algorithm that constructs a collection of classifiers and utilises their collective predictions for classification purposes of classification (Islam et al., 2023; Khalid et al., 2024). In general, ensemble techniques can be categorised into two main types: bagging and boosting.

Bagging, also known as Bootstrap Aggregation, is an ML ensemble technique that combines bagging, short for bootstrap aggregation, and is an ensemble approach that is both straightforward and very effective. This approach employs a bootstrapping technique wherein fresh training samples are generated from the original training set through a process of replacement (Janapareddy & Yenduri, 2023). The newly introduced training samples are referred to as the bootstrap training samples. In the bootstrap method, each sample is utilised for training individual models in isolation and thereafter employed for prediction purposes. The projections from each bootstrapped model are pooled by averaging regression output or voting classification results. Figure 2.6 provides comprehensive representation of this ensemble technique. This approach facilitates a reduction in overfitting. Decision trees are commonly employed as foundational models within bagging techniques.

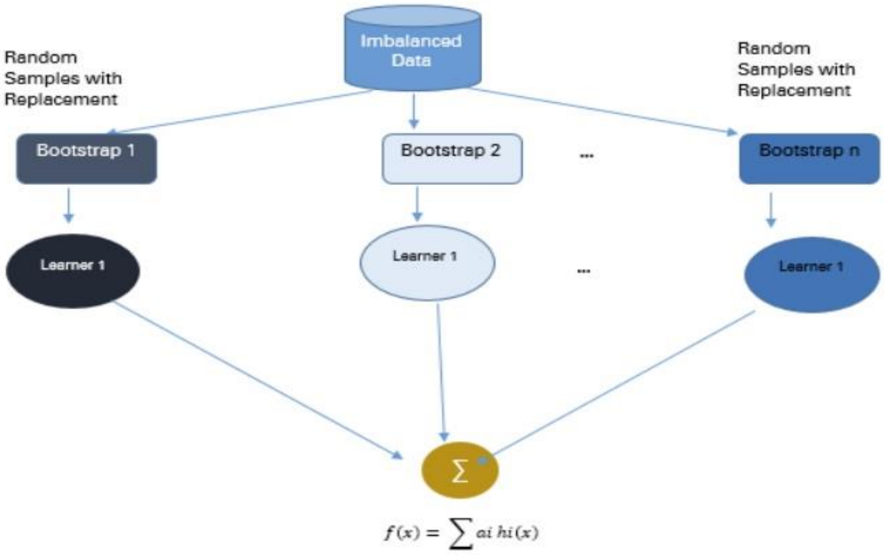


Figure 2.6 *Bagging ensemble technique enhances model robustness by combining predictions from multiple independent models trained on bootstrapped datasets.*

Boosting is a highly effective ensemble technique. The process entails amalgamating weak learners, also known as base learners, to construct a robust learner that can yield superior outcomes compared to those produced by an individual learner (Javaid et al., 2022). In contrast to the concurrent execution of models in bagging, boosting sequentially trains weak learners. Every learner attempts to correct its predecessor through assigning higher weights to previously misclassified data. Hence, in the future, poor learners would prioritise their attention towards the instances that have been incorrectly identified. The diagram in Figure 2.7 provides a more comprehensive representation of the concept of boosting ensemble technique. Bootstrapping was implemented to reduce the risks of variance and overfitting. There are numerous examples of boosting algorithms, including XGBoost, Gradient Boosting, and AdaBoost.

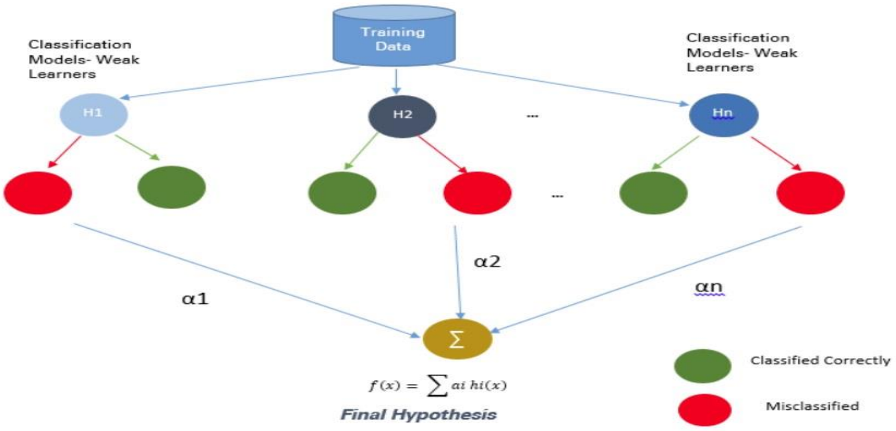
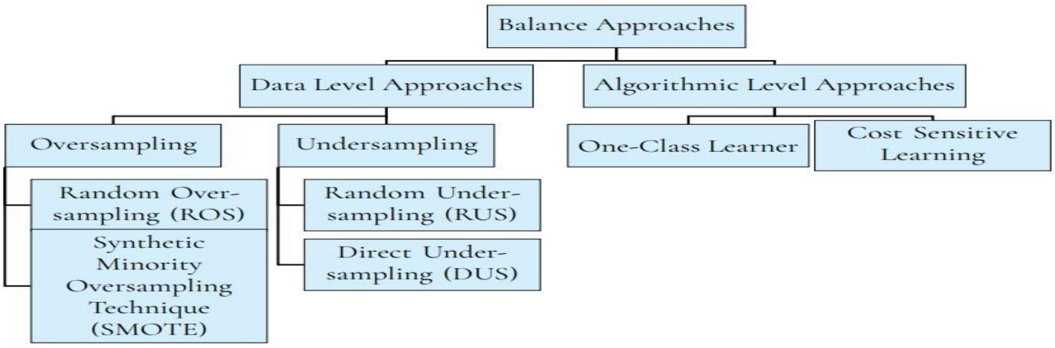


Figure 2.7 *Boosting ensemble technique improves performance by sequentially training models, with each focusing on correcting errors from the previous one.*

In contrast to techniques at the data level, algorithmic level approaches do not involve any pre-processing of the data. Krawczyk (2016) presented an algorithmic method to adjust the model. This may be achieved by utilising cost-sensitive learning or by changing the classification algorithms to effectively handle the detection of minority classes, such as the One-Class Learner. Figure 2.8 presents a comprehensive overview of the many methods employed to address the issue of imbalanced data in a systematic manner (Abdallah et al., 2016).



Balance Approaches
Data Level Approaches
Algorithmic Level Approaches
Oversampling
Undersampling
One-Class Learner
Cost Sensitive Learning
Random Over-sampling (ROS)
Synthetic Minority Oversampling Technique (SMOTE)
Random Under-sampling (RUS)
Direct Under-sampling (DUS)

Figure 2.8 *Balance approaches to handle skewed data and achieve class balance.*

B. Large amount of data

Furthermore, the considerable volume of data and its extensive dimensionality contribute to the intricacy of data mining and detection processes. Hence, it is customary to employ data reduction methodologies such as dimensionality and numerosity reduction techniques. PCA is a widely used technique for reducing data dimensionality of data (Aayushi Agarwal et al., 2021; Gewers et al., 2021).

Complications may arise when dealing with datasets characterised by a substantial volume of data and a high number of dimensions, such as datasets, including credit card transaction information. One potential approach for addressing the issue of excessive dimensionality within a dataset is to reduce the number of dimensions by decreasing the number of features. PCA is a method that may be employed for this purpose. PCA, as other dimensionality reduction methods, offers several benefits. First, it leads to quicker computations owing to the reduced complexity of data. Additionally, the process of data mining becomes more manageable as there are fewer parameters and dimensions to consider.

PCA is a statistical technique that is used to reduce the dimensionality of a dataset by identifying and describing it using a set of orthogonal vectors. These vectors were positioned in the direction with the largest variance, capturing the most critical information from the data. The vectors in question are commonly referred to as principal components. The initial principal component is oriented in the direction exhibiting the greatest variation, whereas the subsequent components display decreasing variance as their order increases. As the order of the vectors increases, the representativeness of the data decreases. These components are identified by computing the covariance matrix of the features, followed by the determination of the eigenvectors and eigenvalues. In the context of linear algebra, each eigenvector can be regarded as a principal component. Notably, the eigenvector corresponding to the eigenvalue with the greatest magnitude is designated as the first principal component.

Once the main components have been calculated, each observation is projected onto a vector. Instead of representing the data based on their original characteristics, they can be represented using principal components. The process described not only decreases the dimensionality of the data but also enhances its security (Alhowaide et al., 2020). By transforming the features, the inherent nature of the characteristics is concealed, resulting in anonymised data that helps safeguard privacy.

2.2 Cybersecurity and machine learning

Cybersecurity is a set of activities designed to protect computers, computer networks, related hardware, and device software, and the information they contain and communicate, from attacks and disruptions (Bansal et al., 2024; Perwej et al., 2021). Cybersecurity is concerned with the identification of various cyber threats or cyberattacks and the development of matching defence methods to prevent them and eventually safeguard the system’s confidentiality, integrity, and availability (Asaad & Saeed, 2022). Cyber-attacks are a continual threat to the banking industry, making cybersecurity a key priority (Al-Alawi & Al-Bassam, 2020). Identity thieves, fraudsters, and hackers are developing new forms of cyberattack and fraudulent activities to avoid being caught by existing cyber fraud detection systems. To overcome this problem, banks must strengthen their cyber security measures to ensure that online scams are prevented in the future.

ML techniques are primarily employed for cybercrime detection. ML algorithms can be used to overcome the limitations and restrictions encountered by traditional detection methods (Shaukat et al., 2020). Pattern identification and anomaly detection are two critical areas in which both machine learning and deep learning have found widespread applications in the field of cybersecurity. The purpose of pattern detection is to discover hidden characteristics that are present in the data and then make use of those characteristics to provide information to an algorithm that is able to comprehend other variants of the data that display characteristics that are comparable. This is particularly relevant in the context of malware analysis and detection. Patterns were deduced or derived from data used for training. The objective of anomaly detection is to develop a conceptual representation of a baseline that accurately characterises a dataset. Anomalies, such as those seen in intrusion and spam detection systems, are identified as deviations from the baseline.

There is a clear connection between the rising popularity of credit cards for spending and the expanding prevalence of fraudulent activities. The issue has been increased by the growing technological prowess of hackers, who exploit vulnerabilities to obtain unauthorised access to credit data to engage in illegal activity. Credit card cyber fraud refers to the unauthorized use of credit card information by criminals to execute illegal transactions or to provide sensitive data to perpetrators engaged in criminal activities.

Supervised learning forms the foundation for most cybersecurity activities. Supervised learning aims to ascertain the benign or detrimental nature of an event or its occurrence. Scholars specialising in cybersecurity have begun to label cyber fraud as a cybersecurity concern. A variety of algorithms, such as Support Vector Machines (SVM), Logistic Regression (LR), Random Forest (RF), and Deep Neural Networks (DNN) have been implemented to resolve credit card cyber fraud concerns in cybersecurity detection systems. As a result, it appears that credit cyber fraud detection research methodologies and cybersecurity machine learning and deep learning techniques overlap substantially. The detection of credit card cyber fraud is a cybersecurity concern that arises from data vulnerabilities, which have the potential to have detrimental effects on both individuals and businesses (Aschi et al., 2022; Li et al., 2020). Credit card cyber fraud detection is a significant subject in the realm of cybersecurity because it can be impeded by a determined assailant seeking to evade detection. To adequately confront inherent risks, it is vital to replicate the offensive strategies employed by the adversary and develop resilient ML models.

An area of research that has not been thoroughly examined is credit card cyber fraud, particularly regarding the application of diverse experimental methodologies, such as ML and DL. This research provides a scholarly examination of credit card cyber fraud, emphasising its importance as a subset of the wider area of cybersecurity. The absence of exhaustive publicly accessible datasets significantly impedes the progress of future research in this field. Consequently, there is a scarcity of observational research and research findings.

2.2.1 Machine learning techniques

ML, a field that enables computers to perform tasks without explicit programming, has the potential to achieve accurate predictions of risk and anomalous behaviour inside datasets, including instances of credit card theft (Sharifani & Amini, 2023). The rationale behind this is that ML facilitates the ability to forecast future events and identify trends through analysis of extensive volumes of data.

As mentioned earlier, numerous types of ML techniques are available and can be categorised into four fundamental categories. Supervised learning refers to an ML paradigm in which the training data provided to an algorithm are accompanied by corresponding labels. Unsupervised learning refers to the utilisation of unlabelled training data in the algorithm. Semi-supervised learning is a hybrid strategy that combines elements of both supervised and unsupervised learning. Reinforcement learning is a learning approach that engages in interactions with the environment through the generation of actions and subsequent identification of faults or rewards.

When applying ML algorithms to identify credit card cyber fraud, the classification task, which involves distinguishing between fraudulent and non-fraudulent transactions, can be classified into three categories: supervised, unsupervised, and semi-supervised learning, depending on the manner in which the dataset has been created. The prevailing approach for constructing credit card cyber fraud detection systems in data mining involves the utilisation of a classification model, which falls under the category of supervised learning techniques.

Data analysis and processing are among the numerous applications of machine learning. Balanced datasets necessitate machine learning. ML algorithms significantly enhance the detection and prevention of cyber fraud by autonomously identifying patterns across vast data sets. Accurate machine learning models facilitate the differentiation between fraudulent and legitimate conduct. New cyber fraud techniques may be accommodated by these intelligent systems. This necessitates thousands of accurate computations in milliseconds. In order to detect cyber deception, future defences must incorporate both supervised and unsupervised technology.

ML systems are trained through supervised learning, which employs labelled and customisable data with known variable objectives. Classification, regression, and inference comprise supervised learning. The most common ML methods in all fields are supervised models that have been trained on a large number of precisely labelled transactions. Each transaction is either legitimate or fraudulent. The algorithms are trained to identify patterns that resemble actual behavior by utilising a significant amount of classified transaction data. Unsupervised learning is a method that trains machine learning algorithms on uncertainty-prone target variables. The primary objective of this model is to identify the most significant data patterns. Cluster segmentation and dimension elimination are unsupervised learning techniques. Semi-supervised learning is a method that employs both supervised and unsupervised learning to train unlabelled data. This method employs the directed learning attribute to analyse the relationships and make predictions, while the unsupervised learning attribute is employed to identify the optimal data representation. Numerous studies have implemented semi-supervised, unsupervised, and supervised machine learning.

2.2.1.1 Supervised techniques

Supervised learning is an approach to ML wherein the model being trained is provided with both input and output labels (Harikrishna et al., 2022). The supervised model trains on labelled input and output data to extract patterns. The extracted patterns were incorporated into subsequent evaluations. Formally, supervised learning can be described as Y = f(x). Y is equal to f(x), where x is a mapping function, Y represents an output variable, and x represents input variables. The objective is to approximate the mapping function such that it accurately predicts the output variable (Y) when unobserved input is provided. Classification and regression are two subcategories of supervised learning(Saranya et al., 2020). The output variable of a classification problem is a category. The output variable of a regression problem is a genuine value.

A. Classification techniques

The classification challenge in the field of ML pertains to the objective of accurately predicting the class label associated with certain data items. One such instance is the identification of cyber fraud detection, which may be characterised as a classification challenge. The objective of this scenario is to forecast whether a transaction is fraudulent or legitimate. In general, classification can be categorised into three types: binary classification, multiclass classification, and multilabel classification (Weng et al., 2020). Binary classification involves assigning one of the two output labels to a given input (e.g., determining whether a transaction is fraudulent or genuine). Multi-class classification, on the other hand, involves assigning one of multiple output labels to an input. Finally, multi-label classification involves assigning multiple target labels to each data sample, where the labels are not mutually exclusive. The binary classification problem is the primary focus of this thesis, in which the output label is classified as either non-fraud or fraud.

The most prevalent approach for identifying credit card cyber fraud is through the application of supervised algorithms (Afriyie et al., 2023). A variety of supervised models are employed in this discipline. SVM was employed to classify data samples into two categories by employing a maximum margin hyperplane (Cervantes et al., 2020). It employs a labelled dataset for each category to classify new data points. The kernel of SVM is composed of mathematical functions that transform input data into a high-dimensional space. Consequently, SVM is capable of classifying linear and nonlinear data by employing the kernel function (Li et al., 2021; Sasikala et al., 2022).

SVM is a classification technique that seeks to identify the most favourable hyperplane that optimises the separation between data points belonging to distinct classes. In the domain of credit card cyber fraud detection, the SVM function discerns the optimal decision boundary that effectively distinguishes genuine transactions from fraudulent transactions. This process determines the support vectors, which refer to the data points that are in close proximity to the decision border. SVM exhibits a high level of efficacy when applied to datasets characterised by many dimensions. Consequently, an SVM is particularly suitable for analysing credit card data that encompass a multitude of attributes. SVMs can effectively discern distinct decision boundaries even in regions with several dimensions. Consequently, SVMs are highly suitable in scenarios where data segregation is not readily apparent. Moreover, SVM can accommodate non-linear data patterns by employing kernel functions, thereby augmenting its capacity to capture intricate linkages inherent in credit card transactions.

A novel approach to credit card cyber fraud detection, which involves a comparison between the Support Vector Classifier and a tree-specific decision tree algorithm, was conceptualised and implemented in a study by Reddy and Sriramya (2023). In detecting credit card cyber fraud datasets, the results indicate that the novel decision tree classifier achieves 94.86% accuracy with a significance level of 0.000 (p0.05) for two-tailed tests, whereas the support vector classifier predicts the same with 98.59% accuracy. This study provides empirical evidence that the Support Vector Classifier algorithm outperforms the novel decision tree Classifier algorithm in terms of credit card cyber fraud detection accuracy.

Linear, radial, polynomial, and sigmoid are the four types of kernel functions utilised by Li et al. (2021), and SVM is used to detect credit card cyber fraud. The SVM parameters (PSO) were optimised using the Cuckoo search algorithm (CS) and genetic algorithm (GA) with a Particle Swarm Optimisation technique. The linear kernel function has been demonstrated to be the most effective in experiments. A radial basis function was employed to optimise the kernel function. CS-SVM and GA-SVM were outperformed by PSO-SVM in terms of overall efficacy.

Zhang et al. (2020) utilised a weighted SVM algorithm. Results of experiments indicated that this model substantially improved performance. Weighted feature-based SVM (WFSVM) with a time-varying inertia weight-based dragonfly algorithm (TVIWDA) was proposed by Arun and Venkatachalapathy (2020). In order to enhance the accuracy of detection, the TVIWDA-optimised property was selected. After that, the WFSVM classifier and the specified characteristics are employed to conduct classification. It is evident from the results that the proposed model surpasses the existing random-tree-based technique. With lesser datasets, the WFSVM is more efficient. To quantify the uncertainty associated with credit card cyber fraud detection and classification, Nama (2023) introduced a novel model. An SVM classifier was implemented using the MLP technique. The research findings indicated that the SVM and MLP techniques achieved an accuracy of 94.59% and 91.21%, respectively. The experimental outcomes demonstrate that SVM and MLP classify credit cyber fraud transactions with an accuracy of over 90%.

Decision trees (DT) are hierarchical data structures that are commonly employed for classification or regression problems (Khalid et al., 2024; Kırelli et al., 2020). Decision trees are composed of nodes that represent feature tests and are connected by branches that lead to many possible outcomes (Lim et al., 2021). Ultimately, the decision tree structure culminates in a final choice. Decision trees operate by iteratively partitioning the data based on the values of the features, resulting in a diagram that resembles a tree structure. This diagram consists of nodes and branches that employ if-then-else logic to categorise the transactions. These tools provide transparency and a user-friendly graphical representation of the decision-making process, rendering them highly beneficial for comprehending the rationale underlying cyber fraud detection determinations. Nevertheless, decision trees are susceptible to overfitting, a phenomenon in which they become too tailored to the training data, thereby constraining their ability to generalise. The performance of a tree is influenced by its depth and complexity, and achieving an optimal balance is essential to ensure efficient cyber fraud detection.

The DT approach has attracted significant research interest.(Bandyopadhyay et al., 2021) employed the DT classifier to identify financial cyber fraud. The DT algorithm demonstrated the highest accuracy among the other classifiers, with a score of 99%. DT using the boosting technique applied by Barahim et al. (2019). The results show that the model achieved the highest accuracy of 98.3%. Choubey and Gautam (2020) utilised a combination of supervised algorithms such as DT, RF, LR, Naive Bayes (NB), and K-near Neighbour (KNN) algorithms. The study observed that the hybrid classifier DT with KNN performed better than any other single classifier. In (Hammed & Soyemi, 2020), the utilisation of the DT algorithm enhanced by regression analysis. The results indicate enhanced performance. This approach is accurate, with a misclassification error rate of 18.4%, and the system successfully validated all the inserted incursions used for testing.

Among the ML approaches, the C4.5, algorithm acts as a DT classifier. The decision was based on certain data occurrences. New model applied C4.5 in Mijwil and Salem (2020). The study revealed that C4.5 is the best classifier compared to other ML techniques. Credit card cyber fraud detection using a C4.5 DT classifier with a bagging ensemble was applied in Husejinovic (2020). This study revealed that bagging with C4.5 DT was the best algorithm. A logistic model tree (LMT) was used in the DT for classification. Hussein et al. (2021) apply LMT to cyber fraud classification and detection. The results show that applying the LMT algorithm to the classification of cyber fraud is better than applying other techniques. LMT model obtained 82.08% accuracy. The intuitionistic fuzzy logic-based DT (IFDTC4. 5) applied by Askari and Hussain (2020) to transaction cyber fraud detection. The results show that IFDTC4.5 outperforms other techniques and can detect cyber fraud proficiently.

The Randon Forest (RF) algorithm is an ensemble learning technique that involves a combination of many decision trees. Every individual tree is built by utilising a randomly selected portion of the available data and a randomly selected subset of the available characteristics. The ultimate forecast was derived by consolidating the forecasts generated by each individual tree. Random Forest is a flexible ensemble of decision trees that is commonly used in the field of credit card cyber fraud detection. It uses a bagging strategy to enhance its performance. Each decision tree is constructed in isolation, utilising a distinct collection of data. These decision trees operate synergistically to provide a collective forecast. RF has exceptional proficiency in managing datasets with a high number of dimensions and intricate feature interdependencies, a characteristic frequently observed in credit card transactions. The utilisation of an ensemble technique in this context strengthens the resilience of the system against overfitting, thus significantly increasing its efficacy in the detection and identification of fraudulent actions. The capacity of RF to effectively capture the non-linear correlations present in the data is a notable benefit, particularly in the context of credit card cyber fraud, where complicated and dynamic patterns are frequently observed. In addition, this tool offers ratings for feature relevance, which can assist in the process of selecting variables and interpreting the model. This capability enables a thorough examination of significant factors in the context of cyber fraud detection. The diagram depicted in Figure 2.9 provides a more comprehensive representation of the RF algorithm.

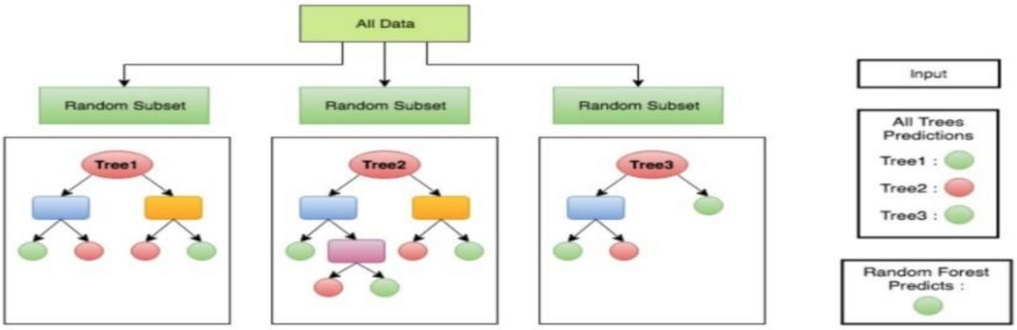


Figure 2.9 *Random Forest algorithm.*

One of the most powerful techniques in credit card cyber fraud detection is RF. According to the literature, RF is the most prevalent credit card cyber fraud detection method. Amusan et al. (2021) applied RF for cyber fraud detection on skewed data. Results indicated that RF had the highest accuracy (95.19%) compared to KNN, LR, and DT. Furthermore, RF was applied with other techniques such as SVM, NB, and KNN in Ata and Hazim (2020). The results showed that the RF algorithm performed better than other techniques. A hybrid model or combination of supervised classifiers was proposed by Choubey and Gautam (2020). Several techniques such as RF, KNN, and LR have been applied. The results show that RF with KNN performed better than RF with a single classifier.

The new model applies RF in Meenakshi et al. (2019). The study revealed that the RF algorithm performed better with more training data; however, the testing and application speeds decreased. Jonnalagadda et al. (2019) applied RF in their study. The recommended value for the highest level of RF precision is 98.6%. The proposed module is suitable for larger datasets and yields more precise results. The RF algorithm performed better with more training data. In Hema and Muttipati (2020), LR, RF, and CatBoost were applied to discover cyber fraud. The results show that RF with CatBoost provides a high accuracy. RF provided the best result with an accuracy 99.95%. The RF with SMOTE was applied by Ahirwar et al. (2020). The results obtained using the RF algorithm showed that this approach was successful in real-time. This model is intended to provide insights into the identification of cyber fraud. Rai and Dwivedi (2020) implemented an RF classifier-based method to detect cyber fraud in credit card systems. Comparing the work with the existing classifiers, LR, and NB. The efficacy of the systems was assessed using a range of metrics, including Accuracy, Precision, Recall, F1 Score, and Specificity, on multiple credit card system datasets. Random Forest performs better than its competitors. Random Forest achieved an accuracy of 99.95%, whereas LR and NB achieved 91.16% and 89.35% accuracy, respectively.

Naive Bayes is a statistical method that bases its decisions on the utmost possible probability in accordance with Bayesian theory. Bayesian probability approximates uncertain probabilities using known values. Uncertain statements are analysed through the application of logic and prior knowledge. This methodology operates under the conditional independence assumption that features in the data are independent (Bagga et al., 2020).

$$P(c_i/f_k) = \frac{P(f_k/c_i)*P(c_i)}{P(f_k)} \qquad (2\text{-}1)$$

$$P(f_k/c_i) = \prod_{i=1}^{n} P(f_k/c_i), k = 1,2,\dots,n \qquad (2\text{-}2)$$

where n is the maximum number of features, $P(f_k/c_i)$ is the probability of generating feature value f_k given class c_i, $P(c_i/f_k)$ is the probability of feature value fk being in class c_i, $P(c_i)$ and $P(f_k)$ are the probabilities of occurrence of class c_i and feature value f_k occurring respectively. The following classification rules were used by the classifier to perform binary classification. The classification was c_i if $P(c_1/f_k)$ > $P(c_2/f_k)$. The classification was c_2 if $P(c_1/f_k)$ < $P(c_2/f_k)$.

Credit card cyber fraud detection has been the subject of several studies that have employed NB and Bayesian belief networks (BBN). Detection of credit card cyber fraud via

NB and robust scaling approaches as described by Borse et al. (2021). The results indicated that the NB classifier with a robust scaler was the most effective in predicting fraudulent activity in the dataset. NB using robust scaling got accuracy 97.78%. In Divakar and Chitharanjan (2019), the NB classifier and other classifiers were applied. NB did not obtain the best results compared with the other classifiers. In (Gupta et al., 2021), among ML algorithms such as LR, RF, and SVM, the performance of the NB algorithm is remarkable. BBN was applied in M. D. Kumar et al. (2020) to detect cyber fraud in credit cards. The results showed that the BBN classifier was more accurate than the NB classifier. This is disturbed by using the conditional dependence between the attributes in the Bayesian network, but this requires more calculation and training processes. The transaction of data value is available in a dataset that is trained with their results as cyber fraud or genuine transactions, which is predicted by a testing data value for individual transactions.

The K-nearest Neighbour classifier is an instance-based learning method that employs similarity measures such as Euclidean, Minkowski, and Manhattan distances. The Minkowski distance is most effectively applied to categorical variables, whereas Euclidean and Manhattan distances are more suitable for continuous variables. Di j represents the Euclidean distance between the two input vectors (Xi, Xj) (Bagga et al., 2020).

$$D_{ij} = \sqrt{\sum_{k=1}^{n} \left(x_{ik} - x_{jk}\right)^2} \ , k = 1,2, \dots , n \qquad (2\text{-}3)$$

For each data point in the dataset, the Euclidean distance between the present input and a different input data point is computed. After arranging the calculated Euclidean distances in ascending order, k items are chosen based on the shortest distance to the input data. The classifier assigns the input point the classification that is most prevalent among k data points.

Various studies have used the k-nearest Neighbour (KNN) technique to detect credit card cyber fraud. KNN uses neighbouring samples to identify class labels. The KNN technique is best for overlapping sample sets (Yao et al. (2019). Chowdari and Chowdari (2021)reported that KNN is a stronger classifier for detecting cyber fraud in credit cards than other techniques such as DT, LR, and RF. In DeepaShree et al. (2019), R. Kumar et al. (2020), the KNN classifier applied for credit card fraudulent transaction detection, comparing with RF and NB, KNN showed the highest accuracy than the RF algorithm and NB. Vengatesan et al. (2020) compared the KNN technique with many other techniques, such as SVM, LR, DT, and RF

XGBoost. The KNN model was found to be the most precise. The KNN model achieved an accuracy score: 99.95%. A new ML approach to detecting anonymous cyber fraud patterns was proposed by Manlangit et al. (2019), who proposed SMOTE with KNN. The results revealed that the proposed model performed satisfactorily. The KNN model achieved a precision of 98.32% and 97.44%.

B. Regression techniques

Logistic regression (LR) is a statistical strategy that models a binary dependent variable using a logistic function. Logistic regression determines the probability of a binary response using a functional approach and various features. It employs a nonlinear sigmoid function to determine the parameters that provide the best fit. The sigmoid function (sigma) and its corresponding input (x) are as follows (Bagga et al., 2020):

$$\sigma(x) = \frac{1}{(1+l^{-x})} \tag{2-4}$$

$$x = w_0 z_0 + w_1 z_1 + \cdots + w_n z_n \tag{2-5}$$

The optimal coefficients w and vector z, representing the input data, were obtained by multiplying each element individually. The result of adding these values is a numerical value that ultimately determines the classification score of the target class. If the sigmoid value is less than 0.5, it is considered to be zero; otherwise, it is 1.

The LR model estimates the likelihood of a binary event using a linear combination of input variables, which are subsequently translated into a probability score using the logistic function. LR is employed in the realm of credit card cyber fraud detection to ascertain the probability of a transaction being fraudulent, utilising the interconnections among input characteristics as a basis for estimation. This system has advantages in terms of simplicity, computing efficiency, and interpretability. LR is a statistical method that offers valuable insight into the impact of specific factors on the probability of fraudulent activities. This characteristic renders it well suited for decision-making processes that prioritise transparency. Nevertheless, the assumption made by this model is that a linear relationship exists between the characteristics and target variable. This assumption may impose limitations on its ability to effectively identify and capture intricate non-linear patterns associated with fraudulent activities. To optimise its performance, one may explore the use of feature engineering techniques and explore correlations between variables. This approach enables the system to effectively adapt to the intricate patterns that may exist in credit card transactions. Logistic regression is recognised as a prominent ML technique employed for classification purposes. Despite the inclusion of the term "regression" in its name, this method does not belong to the family of regression algorithms.

LR derives its nomenclature from its foundation on linear regression, a widely employed ML technique designed for regression tasks. LR entails expressing prediction as the likelihood of the outcome being associated with each class. Real-valued outputs are predicted by the linear regression model through the combination of input variables (x) and their corresponding weights. For enhancing clarity, it is important to acknowledge the presence of a singular input or independent variable denoted as 'x’ alongside a dependent variable referred to as 'y’. The hypothesis of linear regression can be mathematically represented as

$$y = a0 + a1 * x \qquad (2\text{-}6)$$

In the given equation, a0 represents the bias component, whereas a1 denotes the weight assigned to the individual input variable x. These weights were acquired through the training process. In this scenario, the hypothesis has the potential to assume values that are either less than 0 or larger than 1.

LR used a linear equation. Nevertheless, to estimate the likelihood of an individual's classification into each category, a sigmoid or logistic function, as seen in the equation, is employed to compress the projected continuous values within the interval of 0 to 1.

$$sigm(z) = \frac{1}{(1+e^{-z})} \qquad (2\text{-}7)$$

Figure 2.10 depicts a visual representation of the sigmoid function graph. Logistic regression may be mathematically written as equation In this classification problem, there is just one independent variable, denoted as 'x', and one dependent variable, denoted as 'y'. LR often employs a default threshold of 0.5, wherein probabilities below this threshold are assigned to class 0, while probabilities above it are assigned to class 1. The threshold can be modified based on specific requirements.

$$P(y = 1) = sigm(a0 + a1 * x) \qquad (2\text{-}8)$$

The parameters of the logistic regression model, denoted a0 and a1, were acquired during the training process. Hence, by setting a threshold value of 0.5, the anticipated outcome can be expressed in a subsequent manner.

$$y = 1 \ \ if \ P(y = 1) \geq 0.5$$

$$y = 0 \ \ if \ P(y = 1) < 0.5$$

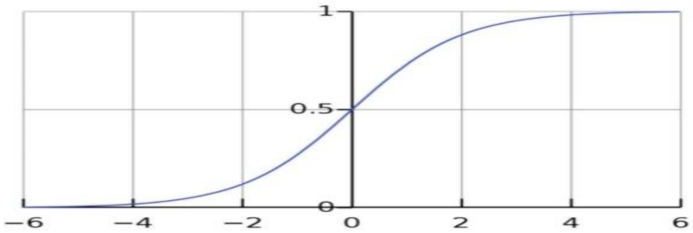


Figure 2.10 *Sigmoid function graph.*

Adityasundar et al. (2020) Utilised logistic regression (LR) on a dataset with a significant class imbalance. This work produced a very robust classification model using imbalanced data. This new system uses LR to build the classifier proposed by Alenzi and Aljehane (2020). Comparative analysis of the LR-based classifier with KNN and voting classifiers. The results indicate that the LR-based method yields the most precise conclusions, with a success percentage of 97.2%. Itoo et al. (2021) compared LR, NB, and KNN for cyber fraud detection. The results showed that LR achieved optimal performance. LR achieved greater accuracy than KNN and NB did. The LR attained accuracy of 95%, while the NB achieved 91%, and the KNN achieved 75%. Karthik et al. (2019) proposed a new approach that employs a stacking classifier that applies LR as a meta-classifier, which is the most promising method, followed by SVM, KNN, and LR. Soh and Yusuf (2019) suggested four models for detecting cyber fraud using imbalanced data. The results show that RF and KNN overfit. Therefore, only DT and LR were compared. The results show that LR with stepwise splitting rules has outperformed the DT, with only 0.6% error rate. Sujatha (2019) used a single hybrid model of undersampling and oversampling. The analysis demonstrated that LR outperformed all other algorithms. The results indicate that the LR and NN methods presented perform better than the DT method.

C. Ensemble techniques

The Random Forest (RF) typically outperforms a single Decision Tree (DT) by creating an ensemble of DTs throughout the training process. New research conducted in 2021 revealed that RF outperformed K-means and SVM Al Rubaie (2021).

Another ensemble method is bagging, a collection of different estimators created using a particular learning process to enhance a single estimator. Bagging reduces the DT classifier variance. This approach created random subsets from the training sample. The bagging method has been applied in several studies,(Alias et al., 2019; Husejinovic, 2020; Karthik et al., 2022; Lin & Jiang, 2021; Mijwil & Salem, 2020). Husejinovic (2020) applied C4.5 DT, NB, and bagging ensemble to predict cyber fraud. The results showed that the best algorithm was bagging with C4.5 DT.

Khalid et al. (2024) introduced an innovative ensemble model that combines KNN, SVM, RF, Bagging, and Boosting classifiers. By incorporating undersampling and SMOTE into a subset of ML algorithms, this ensemble model addresses the issue of dataset imbalance that is prevalent in most credit card datasets. The performance of this ensemble was impressive and demonstrated the effectiveness of merging multiple classifiers to improve the accuracy of cyber fraud detection. The results validated the efficacy of the proposed ensemble model in reducing false positives and false negatives, which are critical obstacles in the field of credit card cyber fraud detection.

Boosting includes AdaBoost, Random Undersampling Boost (RUSBoost), Gradient Boosting (GBM), LightGBM, and Extreme Gradient Boosting (XG Boost) algorithms. Several studies used boosting techniques in the reviewed articles. AdaBoost was developed by Barahim et al. (2019). In this study, DT, NB, and SVM were combined with AdaBoost. The results show that AdaBoost with DT outperforms other techniques. A compared different ensemble methods to predict cyber fraud in credit cards has been done by Faraj et al. (2021). The experiment shows that XGBoost performs better than other ensemble methods and neural networks.

Stacking is a technique in ensemble learning that merges many classification or regression systems. In stacking, a single model is used to exactly integrate predictions from contributing models, but in boosting, a series of models are utilised to enhance the predictions of earlier models. In contrast to bagging, the complete dataset was utilised as compared to portions of the training dataset. Several studies have used stacking to learn classifiers for detecting cyber fraud in credit cards (Karthik et al., 2019; Muaz et al., 2020; Prabhakar et al., 2019; Veigas et al., 2021). The stacked ensemble approach has demonstrated potential for detecting fraudulent transactions. The stacked ensemble exhibits the best performance at 0.78 after trained for sampled datasets (Muaz et al., 2020).

Mienye and Sun (2023) introduced a novel approach to efficiently identify credit card cyber fraud by employing a DL-based stacking ensemble combined with data resampling. As base learners, the stacking ensemble employs LSTM and GRU neural networks, whereas the meta-learner is an MLP. Resampling of the data was accomplished using the hybrid SMOTE-ENN method. Sensitivity, specificity, and AUC values for the proposed method were 1.000, 997%, and 1.000, respectively, surpassing those of the baseline classifiers (AdaBoost, random forest, MLP, LSTM, and GRU). Compared with the performance of other scholarly works, the proposed method demonstrates an exceptional level of performance. Which exhibits superior performance compared to other extensively utilised ML classifiers and methodologies documented in the literature. Subsequent investigations, according to Mienye and Sun (2023), should attempt to incorporate diversity into the foundational models by employing classifiers trained using various techniques, including the integration of LSTM and random forest, logistic regression, or SVM. In addition, future research endeavours may wish to incorporate risk factor analysis and feature importance.

2.2.1.2 Unsupervised techniques

Clustering is the process of categorising similar instances into identical groups. The clustering methods were utilised far less than the classification methods in the reviewed literature. The Hidden Markov Model was used to model the probability distribution across observation sequences. It comprises hidden states and observable outputs. Das et al. (2020) the HMM model to detect cyber-attacks. The results demonstrate the excellent performance of the proposed system, demonstrating the advantage of learning the cardholder’s spending behaviour. Singh and Jain (2019) suggested a method to identify cardholders’ spending profile and then attempted to determine the observation symbols, which will help in an initial estimate of the model parameters. Thus, the HMM can detect whether a transaction is genuine or fraud. The performance of this model was relatively superior for all of the hidden states.

The K-means algorithm, a non-hierarchical technique, can be employed to achieve data clustering. This algorithm uses a simple method. Therefore, K-means classifies a given dataset into a specified number of clusters or K-clusters. Abdulsalami et al. (2019) applied K-mean with a Back-Propagation Neural Network (BPNN). The BPNN model achieved great accuracy with fewer false alarms than the K-means model.

Isolation Forest (IForest) is an unsupervised ensemble. No point-based distance calculations or profiling of regular instances were performed. Instead, the isolated forest builds an ensemble of DTs. The IForest concept is to split anomalies for the purpose of isolation. The data points with the shortest average path length were deemed anomalous, and an ensemble of DTs was generated for a specific data collection. Cyber fraud detection in mobile banking transactions was proposed by Bwalya and Phiri (2023) as an effective implementation of artificial intelligence data mining algorithms by Bwalya and Phiri (2023). This study developed a mobile banking cyber fraud detection system that is robust, cost-effective, efficient, and precise. By employing an AI framework that incorporates K-means clustering and Isolation Forest anomaly detection algorithms, the proposed application was capable of promptly distinguishing between fraudulent and non-fraudulent transactions. A 5% detection rate was achieved for anomalies utilising Isolation Forest and K-means clustering on the volume of transactions analysed. To detect credit card cyber fraud, Jaiswal et al. (2021) proposed a novel model that employs isolated forest and local outlier factor algorithms. Accuracy of Local Outlier Factor was 97%, whereas that of the isolated forest was 76%.

The IForest algorithm is an effective unsupervised method for anomaly detection. Numerous methods can be implemented to enhance the functionality of an isolated forest model. Dwivedi (2021) introduced a novel model to identify credit card cyber fraud. The model implements a variety of ML methodologies. The results demonstrated that the isolated forest achieved greater efficiency. LR: 90%, DT: 94.3%, and RF: 95.5%. 100% IForest. The 99.69% local outlier factor. Oluwasanya and Braimah Joachim (2023) introduced a model by combining the LR and IForest. The accuracy score of the logistic regression algorithm for training data was 99.91%, and for testing data, it was 78%. The corresponding precision, recall, and F1-score were 95%, 56%, and 70%, respectively. Additionally, testing data yielded an accuracy score of 74% and the training data 99.82% for the isolation forest algorithm. The findings derived from the evaluation of the dataset indicate that the isolation forest algorithm performed less effectively than the logistic regression algorithm.

Self-organising map (SOM) is unsupervised NN learning. SOM are appropriate for building and analysing customer profiles to detect cyber fraud. The SOM and NN in the hybrid approach were applied by Harwani et al. (2020). Compared with using SOM and ANN alone, the model achieved better accuracy. DEB et al. (2021) presented three unsupervised algorithms: SOM, K-means clustering using PCA, and T-distributed stochastic Neighbour embedding (T-SNE). For the detection of cyber fraud on credit cards, this model obtained an accuracy rate of 90%. The results also show that K-means clustering combined with PCA is much better than simple k-means clustering. In addition, T-SNE is much better than PCA because PCA is highly affected by outliers.

2.2.1.3 Semisupervised techniques

A hybrid approach that integrates both supervised and unsupervised learning techniques. The unsupervised learning attribute is employed to determine the most efficient method of data representation, while the supervised learning attribute is employed to analyse the relationships within the representation prior to making predictions. Semi-supervised learning is highly advantageous in situations when there is an imbalance in the data gathering. Several studies have employed semi-supervised methods to detect cyber fraud on credit cards (Dzakiyullah et al., 2021; Shekar & Ramakrisha, 2021). Dzakiyullah et al. (2021) presented a combination of semi-supervised learning and autoencoders to detect fraudulent transactions. Results obtained are helpful because credit card cyber fraud will be easily classified at 98%.

2.2.1.4 Deep learning techniques

Deep learning is a relatively new field of machine learning research. It uses a system like that of the human brain to understand data, such as sights, sounds, and words. DL is an ML technique that is based on data analysis. Employing certain representations makes it easier to learn tasks from examples. The learning models developed using various learning frameworks are distinct. The advantage of DL is that it enables the efficient manual replacement of features through the use of unsupervised or semi-supervised feature learning and hierarchical feature extraction (Aayushi Agarwal et al., 2021; Xin et al., 2018). The primary advantage of DL over traditional ML is its superior performance on large datasets. The use of DL for cybersecurity research and intrusion detection is highly important because most attacks use invasive software families that can be detected and classified (Aleesa et al., 2020; Azam et al., 2023). The utilisation of DL is prevalent in the field of pattern recognition. Furthermore, the issue of classification, has also shown efficiency when DL is used (Fang et al., 2021; Ozbayoglu et al., 2020). DL algorithms such as CNN and LSTM are associated with image processing and NLP, respectively. Using these methods for credit card cyber fraud detection has yielded better performance than traditional algorithms (Alarfaj et al., 2022; Alshingiti et al., 2023; Nguyen et al., 2020).

An Artificial Neural Network (ANN) utilises cognitive computing to facilitate the creation of machines that are capable of utilising self-learning algorithms, such as data mining, natural language processing, and pattern recognition. An ANN presents more accurate results because it learns from the patterns of authorised behaviour and thus distinguishes between 'fraud' and 'non-fraud' in credit card transactions (Alarfaj et al., 2022).

ANNs are exceptionally flexible computational architectures that comprise hidden, output, and input layers. By utilising connected neurons, ANNs can accurately represent the complex interactions that occur within datasets. In the domain of credit card cyber fraud detection, ANNs operate by gaining knowledge and generating representations of the intricate relationships between various attributes. By modifying the quantity of layers and neurons, the complexity of their architecture can be tailored to correspond with the intricacy of the cyber fraud detection problem at hand. ANNs exhibit a significant level of adaptability, which enables them to accurately detect and comprehend a wide range of fraudulent patterns, including those that are simple to extremely intricate. They possess remarkable expertise in capturing complex non-linear connections, which are commonly encountered in cases involving credit card cyber fraud. Figure 2.11 shows the Architecture of the Artificial Neural Networks (ANN).

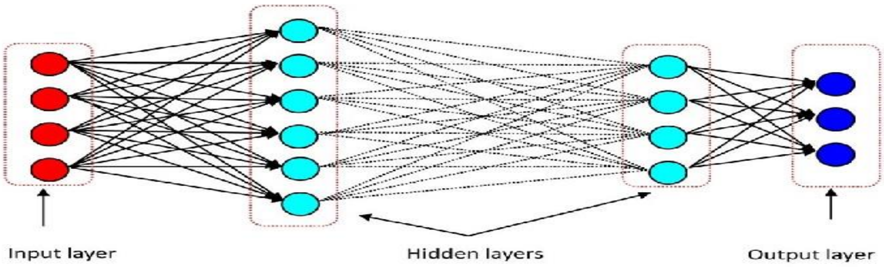


Figure 2.11 *Architecture of the Artificial Neural Networks (ANN).*

Abhishek Agarwal et al. (2021) an ANN for identity theft detection. The proposed model uses different layers in an NN to determine cyber fraudulent transactions. The results show that applying an ANN yields accuracy nearly equal to 100%. The results show that ANN is best suited for determining whether a transaction is fraudulent. A recent study applied ANN to detect cyber fraud. The ANN technique was then compared with ML algorithms such as SVM and KNN. The results show that ANN provides more accuracy than other ML algorithms, and the suggested model is optimal for detecting credit card cyber fraud (Asha & KR, 2021).

Abdulsalami et al. (2019) applied BPNN and K-means. The results indicate that the BPNN algorithm is more accurate than the k-means algorithm. BPNN achieved accuracy of 79.9%. The results also indicate that K-means reduced the prediction time and provided an advantage over BPNN. Daliri (2020) applied the Harmony Search algorithm with an ANN (NNHS) to improve cyber fraud detection in banking systems. The results demonstrate an acceptable capability for cyber fraud detection based on customer information. Oumar and Augustin (2019) applied an ANN with LR for cyber fraud detection. Backpropagation decreases the error function and enables the model to discriminate between fraudulent and legitimate transactions. The proposed model was 99.48% accurate in its predictions and highly reliable.

Al Balawi and Aljohani (2023) introduced a novel model for detecting credit card cyber fraud that incorporates an ANN and a CNN. An ANN model can acquire knowledge of any nonlinear function as well as weights that correspond to any given input value and output value. Furthermore, the activation function acts as a crucial criterion for adapting nonlinear functions and facilitates the network's acquisition of knowledge regarding intricate associations between the input and output values. The deep neural network model that was devised exhibits effective classification accuracy. The accuracy of the proposed system was 99.81 %.

The MLP technique is often regarded as the most common method in ML due to its exceptional accuracy in approximating nonlinear functions. An MLP is a feed-forward ANN composed of a collection of neurons connected via linkage weights. A set of inputs is transformed into outputs intended by the MLP. The architectural design of the MLP is depicted in Figure 2.12, comprising three fundamental components: an input layer, concealed layer, and output layer. After receiving the data, the input layer transmits it to the first covert layer, which in turn passes it to the output layer. Each stratum is comprised of a specific number of neurons. Neurons are linked between layers (Ramchoun et al., 2017).

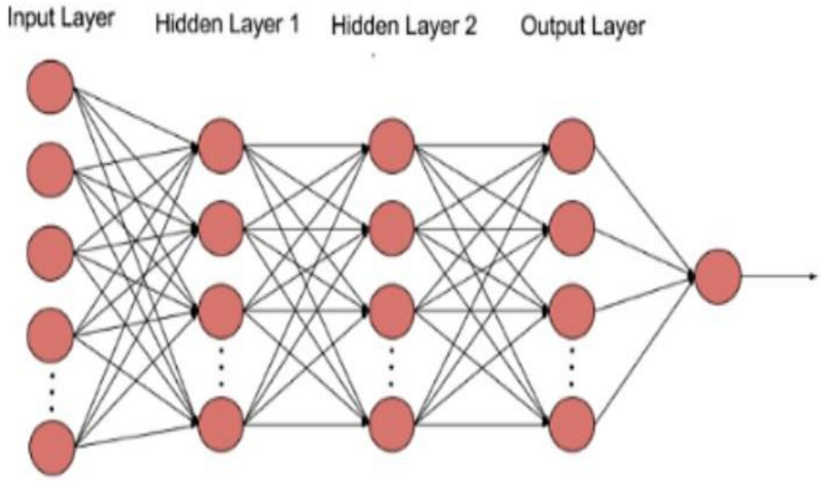


Figure 2.12 *Architecture of the Multilayer perceptron (MLP).*

The value of output (O_j) for each artificial neuron j situated in the hidden layer is determined by the equation given:

$$O_j = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \qquad (2\text{-}9)$$

The parameter values for the sigmoid activation function (f), n (number of neurons in the previous layer), x (input value), b (bias), and w (weight) are as follows:

$$f(x) = \frac{1}{1-e^{-x}} \qquad (2\text{-}10)$$

Alias et al. (2019) conducted a study where they compared MLP with 15 different supervised ML approaches to identify the most accurate one for detecting fraudulent transactions. The results showed that MLP generated the highest detection accuracy at 98%. Can et al. (2020) applied MLP and other ML techniques such as DT, RF, and NB. Regarding amount-based profiling, both MLP and classifiers demonstrated substantial improvements. Faridpour and Moradi (2020) present an innovative machine learning model that uses customer profile data to identify cyber fraud in banking transactions. Bank transactional data are utilised and an MLP with an adjustable learning rate is trained to demonstrate transaction authenticity, thus improving the detection process. The proposed model surpasses SVM and LR. The accuracy was 99.90%. Mim et al. (2024) applied MLP and various ML techniques, including LR, RF, XGBoost, SVM, MLP, DT, and KNN, as well as ensemble classifiers, with and without sampling techniques, in the context of credit card cyber fraud detection. Soft voting ensemble learning model was assessed and contrasted with a variety of advanced sampling techniques (including undersampling, oversampling, and hybrid sampling) when applied to unbalanced data to detect credit card cyber fraud. The proposed soft-voting approach exhibits superior performance compared to individual classifiers. It obtained an accuracy of 99.96%, a precision of 98.70, a recall of 96.94%, and f1-score of 87.64%, despite its false negative rate (FNR) of 0.0306.

Kasasbeh et al. (2022) examined the issue of cyber fraud detection in credit card payment systems was examined by Kasasbeh et al. (2022) in the form of a binary classification problem. The proposed solution utilises an MLP ANN-based model to enhance the precision of the detection approach. Analysis of the performance outcomes of these metrics using descriptive statistics. The results indicate that by employing ANN, a high accuracy was achieved across all layers. Specifically, the classification F-measure for cyber fraud cases was 84.76% for ANN with one hidden layer, 85.13% for two hidden layers, and 82.51% for three hidden layers when applied to ANN with one hidden layer. Based on these findings, it can be concluded that ANN are extremely effective in detecting credit card transaction cyber fraud.

A CNN is composed of multiple layers, the outputs of which are used as inputs to the layers that follow (Ketkar et al., 2020; Krichen, 2023). Convolutional Neural Networks (ConvNets) are specifically built to handle data that is structured as several arrays. For example, ConvNets are commonly used to analyse colour images, which consist of three 2D arrays representing pixel intensities in three different colour channels. ConvNets use the properties of natural signals through four fundamental concepts: local connections, shared weights, pooling, and the incorporation of many layers (LeCun et al., 2015). Convolutional networks integrate three architectural ideas to ensure a certain degree of transition, size, and distortion invariance: 1) local receptive fields, 2) shared weights (or duplication of weights), and 3) spatial or temporal subsampling. CNNs are AI systems that accomplish object identification, recognition, and classification in addition to object detection and segmentation in images. These were constructed using multi-layer neural networks. CNN or ConvNet is a well-known architecture for discriminative DL that can be trained automatically from the input object (Taye, 2023).

CNNs, originally developed for image analysis, may be modified, and used to structured data, such as credit card transactions, to enhance cyber fraud detection capabilities. In the present context, CNNs operate by employing convolutional layers to identify spatial patterns, such as atypical spending trends or transaction irregularities. They possess a remarkable ability to discern spatial patterns that might potentially signify instances of fraudulent behaviour, since they have outstanding proficiency in collecting intricate patterns within datasets. CNNs have demonstrated their efficacy in successfully processing multi-channel data, therefore enabling a thorough analysis of transaction information. Due to this attribute, they are particularly well-suited for the identification of complex and geographically dispersed fraudulent patterns in transactions. Figure 2.13 shows the architecture of CNN.

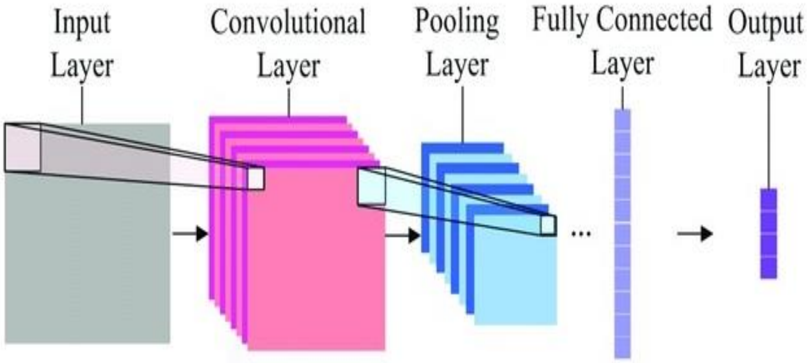


Figure 2.13 *Architecture of the convolutional neural network (CNN).*

Aayushi Agarwal et al. (2021) used DL techniques such as CNN, BILSTM with an Attention layer to classify illegitimate transactions. The CNN-Bi-LSTM-ATTENTION model is highly effective in identifying fraudulent classes. Analysis indicated that the model was adequate and yielded an accuracy of 95%. The results demonstrate that the addition of an attention layer improves the performance of the model, allowing it to accurately discriminate between fraudulent and legitimate transactions. A CNN, NB, DT, and RF hybrid model was deployed by Aswathy and Samuel (2019); these algorithms were used as single models. These were then used as hybrid models using the majority-voting technique. An adaptive boosting algorithm was used to boost the performance of the classifiers.

Most of the information required to identify fraudulent financial transactions is contained in the method proposed by Yamini et al. (2023). Substantial labour savings associated with credit card cyber fraud detection have become feasible due to recent advancements in ML, artificial intelligence, and other critical IT domains. The data are utilised to train the Convolutional Neural Network with an extreme learning machine (CNN-ELM) model after pre-processing and feature retrieval tasks are performed on the input image using principal component analysis. Comparable in accuracy to alternative approaches, such as CNN and ELM, the suggested approach obtains a value of approximately 98.7%. Nalayini et al. (2022) introduced a credit card cyber fraud detector that was optimised for large-scale real-time datasets and utilised a CNN combined with a smart matrix algorithm. The dataset was pre-processed through random sampling to facilitate efficient model training. Normalisation was applied to this pre-processed dataset to obtain the standardised input. Using the ingenious matrix algorithm, the features are sequenced to be selected. Compared to alternative ML approaches, such as K-nearest Neighbour and Naive Bayes, the performance of the three-layered CNN model is superior.

Efficient and effective techniques are required to detect cyber fraud in credit card transactions. Berhane et al. (2023) suggested a hybrid CNN-SVM model for the detection of fraudulent credit card transactions has been suggested by Berhane et al. (2023). Using publicly available real world card transaction dataset, ability of the proposed CNN-SVM was evaluated. By substituting an SVM classifier for the final output layer of the CNN model, the architecture of the hybrid CNN-SVM model was developed. The initial classifier is a support vector machine stacked on top of the fully connected and softmax layers, which was trained using an end-to-end approach. The first classifier consisted of a fully connected layer with softmax. The experimental findings indicate that the hybrid CNN-SVM model achieved classification performances of 91.08%, 90.50%, 90.34%, 90.41, respectively, in terms of accuracy, precision, recall, and F1-score.

DNNs, which provide potent tools for automatically producing high-level abstractions of complicated multimodal data, have recently garnered significant interest from businesses and academics. DNNs learn features independently, resulting in an increasingly accurate learning process. DNNs have been demonstrated to be more efficient and accurate. Four studies used DNN. According to Arya and Sastry G (2020), the proposed model is flexible in terms of data disparity and resistance to hidden transaction patterns. Adaptive optimisation is recommended to improve the prediction of cyber fraud. The results demonstrated its superiority over other current methods.

Habibpour et al. (2023) state that extensive investigative efforts have been devoted to enhancing the precision of point predictions and reducing undesired biases through the development of diverse network architectures and learning models for credit card cyber fraud detection using DNNs. In combination with point estimation, quantifying uncertainty is crucial for preventing suboptimal decisions caused by low confidence, and mitigating model injustice, thereby enabling practitioners to construct reliable systems. In real-world card cyber fraud detection areas, it is crucial to explicitly evaluate the uncertainties associated with DNNs predictions for the following reasons: (a) fraudster strategies are constantly evolving, and as a result, DNNs meet observations that were not created using the same technique as the training distribution; and (b) professional experts inspect very few transactions to update DNNs because of the time-consuming process.

To detect card cyber fraud using transaction data, Three uncertainty quantification (UQ) methodologies were introduced by Habibpour et al. (2023): ensemble, Monte Carlo dropout, and ensemble Monte Carlo dropout. Furthermore, to assess the predictive uncertainty estimates, a number of performance metrics and a UQ confusion matrix were employed. The experimental findings demonstrate that the ensemble approach outperforms the other methods in capturing the uncertainty associated with the generated predictions. Furthermore, the model under consideration demonstrates how the suggested UQ methods enhance the accuracy of point predictions, thereby advancing the cyber fraud prevention process.

The DCNN technique can improve the accuracy of detection when a large volume of data is involved. Chen and Lai (2021) presented existing ML models, including LR, SVM, and RF, as well as autoencoder and other DL models. Results show a detection accuracy of 99% was attained over a 45-s duration. Despite the vast amount of data, this model provides enhanced detection. The DL technique provides high accuracy and rapid patterns for detecting complex and unknown patterns. The 1DCNN, 2DCNN, and DCNN have also been utilised to detect credit card cyber fraud (Cheng et al. (2020); Deepika and Senthil (2019); Nguyen et al. (2020). Karthika and Senthilselvi (2023) designed a 1DCNN with the dual purpose of temporal and spatial feature learning to address the challenges. The CNN foundation model was enhanced by incorporating a dilated convolutional layer (DCL). The imbalance was resolved using undersampling and oversampling techniques. Experiments were conducted on three datasets and compared with an extant CNN model with respect to a variety of parameters. The simulation results demonstrated that the DCNN model, when combined with the sampling technique, achieved an accuracy of 97.39% on a tiny card database. In comparison, CNN achieved an accuracy of 94.44% on the same database.

An RNN, or recurrent neural network, is a structural element that is employed to retain information about previous input sequences. It comprises the links between the internal nodes of a directed graph. This depends on the amount of internal memory required. The architecture of an RNN, a type of deep model, incorporates a feedback-loop structure. "Recurrent" is a term used to describe a process in which the same function is executed for each data input, and the result of the current input is dependent on the computation conducted previously. The dominance of RNN can be attributed to its ability to model sequences by considering the interdependent relationships in the sequence samples. The architecture of the RNN is illustrated in Figure 2.14.

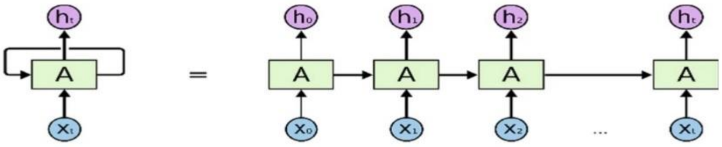


Figure 2.14 *Architecture of Recurrent Neural Network (RNN).*

Bandyopadhyay and Dutta (2020) designed a novel model to identify fraudulent transactions and implement effective cyber fraud-mitigation measures. The stacked-RNN model is suggested and executed by incorporating essential hyperparameter fine-tuning. By adjusting the hyperparameters, it is possible to obtain models with finer particle structures and greater performance. The results indicate that the proposed model is capable of reliably identifying suspect transactions, with a 99.87% success rate. A favourable aspect of the proposed method is its suitability for vast financial datasets. An efficient and error-free mobile transaction system is necessary to alert consumers when deceptive transactions are initiated. Risk factor and threshold analyses were conducted using the proposed method prior to its implementation in real-time transactions. It enables organisations to approve or decline specific transactions.

Forough and Momtazi (2021) proposed a deep RNN-based ensemble model and ANN-based voting approach. The ensemble model leverages a variety of RNN as fundamental classifiers and combines the output using an FFNN as the voting method. This classification system employs several GRU or LSTM networks. The results indicate that the proposed model outperforms the competing models. The proposed model is superior to the existing models in this field. A bidirectional gated recurrent unit (BGRU) was applied by Sadgali et al. (2021). Algorithms such as the GRU, LSTM, BRU, and SMOTE were utilised in this model. BGRU obtained a high accuracy of 97.16%.

LSTM is a helpful technique for predicting cyber fraud because of its historical knowledge and the link between prediction outputs and historical input. The LSTM architecture enables sequence-prediction problems to be learned through long-term reliance. LSTM is a unique form of artificial RNN architecture used to model time-series data in the domain of DL. Long short-term sequence dependencies can be learned, and a transaction label can be predicted given the sequence of prior transactions with LSTM, as opposed to conventional feedforward neural networks, which feature feedback connections between hidden units that correspond to discrete-time steps. The issue of vanishing and exploding gradients, which is evident during the training process of conventional RNNs, has been addressed with the development of

LSTMs. An LSTM unit is composed of a memory cell, where data are stored and modified by three specialised gates: forget gate, input gate, and output gate. Values are retained in the cell for an indefinite period, whereas the three gates control the information flow into and out of the cell. Figure 2.15 illustrates the structure of an LSTM unit (Benchaji et al., 2021).

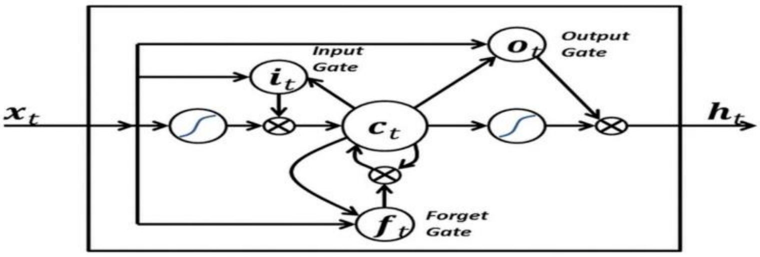


Figure 2.15 *Structure of an LSTM unit.*

Alghofaili et al. (2020) developed a new model to improve both the present detection techniques and detection accuracy considering large amounts of data. Findings demonstrated that LSTM performed perfectly, achieving 99.95% accuracy. Benchaji et al. (2021) recommended a model to record the previous purchasing behaviour of card holders. The results showed that the LSTM model achieved a high level of performance. Owolafe et al. (2021) proposed a new model as a means of mitigating misclassification in cyber fraud detection systems. The application of an LSTM-RNN was implemented to classify financial transactions as fraudulent. The system utilises Principal Component Analysis to extract the desired features from two distinct datasets. Prior to that, the arbitrary assignment method and the min–max scalar algorithm were applied for normalisation. The LSTM-RNN was subsequently trained on the selected pertinent features to classify them. A comparison of the obtained results to previous research revealed that this cyber fraud model achieved both a high rate of accurate classification and a low rate of false alarms. It achieved 99.58 % accuracy in predictions, 99.6 % precision, and 80 % recall. The implementation of this system will empower government agencies and financial institutions engaged in financial transactions to identify instances of fraudulent activity and provide appropriate remedial measures. Fakiha (2023) constructed a forensic detection model for credit card cyber fraud that employed LSTM DNNs to model sequential data. This research aims to determine whether an LSTM-attention algorithm is capable of accurately predicting fraudulent transactions can also identify the most critical transactions in an input sequence. By employing attention mechanisms that enhance the model's performance and the selection of the most pertinent predictive features, uniform manifold approximation, and transaction sequences, the LSTM-attention model achieves its intended level of effectiveness. The findings indicate that forensic credit card cyber fraud detection can be accomplished using LSTM-attention algorithms with a high degree of accuracy and precision. The unique contribution of this study is its effective implementation of an LSTM-attention algorithm for credit card cyber fraud detection, and validation of the model's capability to mitigate fraudulent transactions within banking institutions. Although ML algorithms are commonly used to automatically detect credit card cyber fraud, they fail to account for deceptive conduct or behavioural anomalies that may result in false alerts. Maheshwari et al. (2023) introduced a novel hybrid model with the objective of identifying the occurrence of credit card cyber fraud. Using DL and SMOTE oversampling techniques, this study sought to develop a model able of forecasting credit card cyber fraud. For cyber fraud detection, RNN-LSTM and an attention mechanism have been proposed. It is well known that this model efficiently processes sequential data that contain intricate vector relationships. Comparing the performance of RNN-LSTM to that of ANN, XGBoost, Random Forest, Naive Bayes, and SVM classifiers reveals that our proposed model generates robust results with an accuracy of 99.4%, as demonstrated by the experiments. The model can mitigate global financial losses through the detection and classification of credit card schemes and cyber fraud.

BiLSTM models such as time-series credit card transactions are specifically designed to handle sequential data. In the detection of credit card cyber fraud, BiLSTM models were employed to analyse data in both forward and backward directions. These models utilise LSTM cells equipped with memory units that enable them to record temporal patterns and relationships. Temporal dependencies and patterns that change over time are effectively identified, rendering them a helpful instrument for the detection of time-sensitive fraudulent operations. The bidirectional character of BILSTM models confers an enhanced capacity to discern temporal patterns by incorporating information from both preceding and subsequent contexts. The identification of changing and time-sensitive cyber fraud tendencies in credit card transactions is of utmost importance. Figure 2.16 shows the architecture of the BiLSTM.

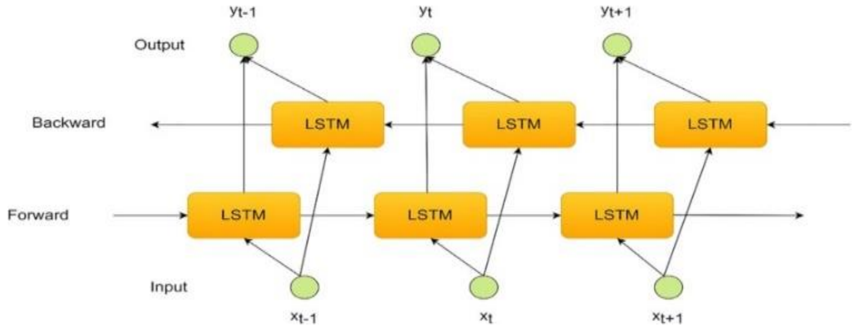


Figure 2.16 *Architecture of Bidirectional Long Short-Term Memory (BiLSTM).*

The issue of imbalanced data is a constraint on the current approaches to detect cyber fraud. To enhance the classification performance of credit card cyber fraud, Narayan and Ganapathisamy (2022) introduced the Hybrid Sampling (HS)- Similarity Attention Layer (SAL)- BiLSTM architecture. The SAL-BiLSTM model is evaluated using two datasets: European data and revolutionary analytics. By hybrid sampling of the minority class and undersampling of the majority class, the SMOTE-ENN decreases data discrepancy. SAL is implemented to quantify the similarity of a data sequence to assign importance to distinctive features and mitigate the issue of overfitting in classification. A benefit of the suggested strategy is that it enhances the representation of the minority class. The proposed SAL method assists in concentrating on the distinctive characteristics of the datasets to enhance classification performance. Forward and backward analyses were performed using the BiLSTM model to identify the pertinent features for classification. The recall value of the proposed HS-SAL-BiLSTM was 99.2%, whereas that of the existing RF-SMOTE-Support Vector Machine (SVM) was 97.7%. A DL-based hybrid approach for detecting fraudulent transactions was applied by Cheon et al. (2021). A Bi-LSTM autoencoder with an isolation forest is incorporated into the new model. This model suggests that fraudulent transactions can be detected at a rate of 87%. The suggested model scored the highest. This model has the potential to be employed as an effective method for detecting cyber fraud.

Deep Belief Networks (DBNs) are generative models that comprise numerous strata of latent stochastic variables. Latent variables with binary values are frequently referred to as concealed units. DBNs consist of a series of Boltzmann Machines interconnected at each layer; each RBM layer can communicate with both the preceding and succeeding layers. DBNs were implemented to process motion capture, video recognition, and image recognition. Operational DBNs: (1) DBNs are trained using greedy learning algorithms. The greedy learning algorithm discovers top-down and generative weights in a layer-by-layer manner. (2) Gibbs sampling is executed on the two uppermost concealed layers using DBNs. A sample was extracted from the RBM, defined by the top two concealed layers at this juncture. (3) A single iteration of ancestral sampling was employed by the DBNs to select a sample from the visible units across the remainder of the model. (4) Using a single bottom-up approach, DBNs discover that the values of the latent variables in each layer can be deduced (Sarumathi & Saraswathy, 2022). Figure 2.17 shows the architecture of Deep Belief Networks (DBNs).

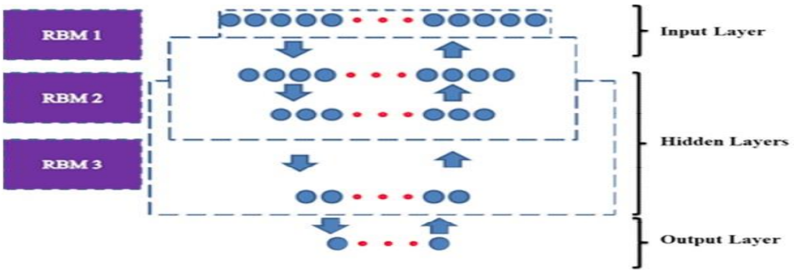


Figure 2.17 *Architecture of Deep Belief Networks (DBNs).*

Fraudsters and identity criminals proliferated in tandem with the number of credit card users. Banks face a significant challenge in determining whether a transaction is a cyber fraud. In a recent publication, Bhowmik et al. (2022) proposed an innovative ensemble learning technique to assess the predictive accuracy of unpaid customer invoices. The model utilised ensemble learning to obtain results after applying four algorithms to the dataset: the NB Classifier Algorithm, LR, DT, and DBN. The determination was made by doing a comparative examination of the outputs of ensemble learning using three algorithms (NB, LR, and DT) on the same dataset. The results underscore the significance of the DBN algorithm and the performance of the proposed ensemble learning model, which incorporates four algorithms, in predicting the default of credit card clients. It also exhibits exceptional precision.

Zhang et al. (2021) proposed a new model utilising DBN and advanced feature engineering based on homogeneity-oriented behaviour analysis (HOBA). The results indicate that the proposed model is effective and capable of identifying cyber fraud. The DBN classifier with HOBA achieves a performance that is superior to that of standard models.

The Restricted Boltzmann machine (RBM) comprises visible and hidden layers linked by symmetrical weights. The neurons in the visible layer correspond to the X inputs, whereas the responses of the neurons H in the hidden layer reflect the eventuality distribution of the inputs. RBMs, which are stochastic neural networks capable of learning from a probability distribution over a given set of inputs, were created by Hinton. Dimensionality reduction, classification, regression, collaborative filtering, feature learning, and topic modelling are all applications of the DL algorithm. RBMs are the fundamental building blocks of Deep Belief Networks (DBNs). The RBMs were partitioned into two levels. Both visible and covert units were seen. Every hidden unit is connected to each visible unit. RBMs lack output nodes and do not have a bias unit coupled to both the visible and concealed units. RBMs consist of a forward pass and reverse pass phase (Sarumathi & Saraswathy, 2022).

A Generative Adversarial Network (GAN) comprises two feed-forward neural networks, a generator and discriminator, competing with each other. GANs, generate new instances of data that are similar to the training data. A generator, which learns to produce fabricated data, and a discriminator, which acquires knowledge from erroneous information, are the two components of GAN. The prevalence of GANs has increased over time. Dark-matter research can be used to enhance astronomical images and simulate gravitational lensing. GANs are employed by video game developers to enhance the low-resolution, two-dimensional textures found in older games by training an image to a resolution of 4 K or higher. GANs facilitate the generation of animated characters and realistic images, the rendering of 3D objects, and photographs of human features. a) Phases of GANs: (1) Through learning, the discriminator becomes capable of differentiating between the generator's fabricated data and the authentic sample data. (2) False data are generated by the generator during the initial training, and the discriminator rapidly acquires the ability to distinguish it. (3) To revise the model, the GAN transmits the results to the discriminator and generator (Sarumathi & Saraswathy, 2022).

In recent years, fraudulent online payments have increased dramatically, in association with the exponential growth of mobile banking and e-commerce. While ML and DL are extensively employed in the domain of credit card cyber fraud detection, the efficacy of conventional binary classification algorithms is limited by the unbalanced nature of typical credit card transaction datasets, which contain significantly less fraudulent data than normal transaction data (Ding et al., 2023). Researchers have oversampled minority class data and employ ensemble learning classification algorithms to avoid this issue. However, oversampling has several drawbacks.

Ding et al. (2023) introduced an innovative approach to enhance the generator component of the Variational Autoencoder Generative Adversarial Network (VAEGAN). Additionally, they proposed a fresh oversampling method that effectively produces diverse and credible minority class data. To train the ensemble learning classification model, minority class cyber fraud data is generated to augment the training set. The experimental results indicate that the oversampling method employing the enhanced VAEGAN is superior in terms of Precision, F1_score, and other metrics when compared to the oversampling methods of GAN, Variational Autoencoder (VAE), and SMOTE, when evaluated on an open credit card dataset. The classification challenge posed by unbalanced data is effectively addressed by the oversampling technique based on the enhanced VAEGAN.

Autoencoders (AEs) are a distinct category of feedforward neural networks characterised by identical inputs and outputs. To solve unsupervised learning issues, autoencoders were developed by Geoffrey Hinton in the 1980s. They are neural networks trained to replicate data from the input to the output layer. Pharmaceutical research, image processing, and popularity forecasting are some applications of autoencoders(Sarumathi & Saraswathy, 2022; Sayan, 2020). The input-output mapping between the encoding and decoding phases was discovered by the AE. The input was mapped to the hidden layer by the encoder, and the decoder reconstructed the input using the hidden layer as the output layer.

Misra et al. (2020) the AE model to cyber fraud detection. In the first step, the two-stage model with an autoencoder converts the transaction characteristics to a lower-dimensional feature vector. Following this, feature vectors are input into a classifier. The proposed model outperforms other models, as evidenced by the results. Wu et al. (2020) employed dual autoencoder generative adversarial networks (DAEGAN) to solve an imbalanced classification problem. The proposed model trains the GAN to duplicate fraudulent transactions for autoencoder training. The proposed model outperformed several other classification algorithms. Owing to extremely skewed class distributions, credit card datasets present unbalanced classification situations. To address this issue, The new model proposed by Tingfei et al. (2020) employs an oversampling technique based on variational automatic coding (VAE) in combination with DL techniques. The results demonstrate that the VAE model outperforms previous oversampling techniques that were based on GAN models.

2.2.1.5 Metaheuristic techniques

Makolo and Adeboye (2021) applied a genetic algorithm and a multivariate normal distribution to an unbalanced dataset to generate a hybrid model. The prediction accuracy was compared with that of the DT, ANN, and SVM. The model yielded a remarkable F-score of 93.5%, whereas ANN is 68.5%, DT is 80.0%, and SVM is 84.2%. Enhanced hybrid system for credit card cyber fraud prediction in Nwogu and Nwachukwu (2019). A genetic algorithm with RF model optimisation (GAORF) was employed. This can assist in resolving the problem of a shortage of transaction data, as well as the problem of inadequate optimisation and convergence of RF algorithms. The model significantly reduces the overall number of misclassifications.

The use of the harmony search algorithm (HAS) with NN to increase cyber fraud detection was described by Daliri (2020). The model uses HAS to optimise the parameters of ANN. The NNHS provides a method based on HAS that successfully predicts the optimal structure for an ANN and identifies the algorithm hidden inside the data. A comparison revealed that the highest accuracy achieved was 86%.

2.3 Chapter summary

This chapter reviews pertinent studies on methodologies for identifying credit card cyber fraud. A variety of factors contribute to the development of cyber fraud that involves credit cards. The situation is exacerbated by the growing technical capabilities of criminals, who exploit security vulnerabilities to obtain confidential information or credit card data from clients. The purpose of the unauthorised use of credit card information is to commit fraudulent activities. The academic literature explores a variety of ML and DL methodologies that have been applied to credit card cyber fraud. Each theory established a foundation for developing credit card cyber fraud models utilising distinct datasets; this was a recurring theme. The development of a comprehensive model is contingent upon the ability to identify the cognitive process that is associated with credit card cyber fraud. This information is acquired through the comprehension of credit card datasets.

Over the past few years, there has been a growing trend in the examination of DL techniques. By employing DL, enhanced precision and optimised performance can be achieved. The application of DL techniques enables the system to respond flexibly to complex data patterns and identify new fraudulent patterns. To detect credit card cyber fraud more effectively, additional research should be conducted on DL techniques. In addition, considering the limitations of each ML/DL technique, it is prudent to consider the integration of ML and DL algorithms to achieve encouraging detection outcomes. Several scholars have proposed integrating DL methods with conventional ML methods to enhance the accuracy of credit card cyber fraud detection from unbalanced datasets. Researchers are encouraged to employ both undersampling and oversampling methods because of the exceedingly skewed nature of the datasets.

The ML and DL techniques have been described. An evaluation was conducted on both ML and DL classifiers used in credit cyber fraud detection. The evaluation included an analysis of their merits and drawbacks, number of identified classes, databases used, and measurement metrics employed. A review of prior research indicates that ML and DL -based cyber fraud recognition algorithms that are designed to identify fraudulent transactions still have room for improvement. Hence, the primary objective of this research is to investigate and formulate novel approaches for identifying prospective instances of credit card cyber fraud through an understanding of advancements made in the fields of cybersecurity and credit card cyber fraud. In addition, a conceptual framework is proposed. In this research, a novel model was developed and evaluated by combining ML and DL techniques. The subsequent sections provide comprehensive explanations of the proposed methods, their outcomes, and a comparison with the current state-of-the-art methods.

CHAPTER 3: RESEARCH METHODOLOGY

3.1 Research philosophy

In science, methodology refers to the various concepts, techniques, and practices that guide empirical research. It should be viewed as comprising the complete research process (i.e., planning and carrying out the research study, drawing conclusions, and distributing the findings (Geoffrey, 2019; Kazdin, 2016). A research methodology is a strategy for resolving an issue - a description of how the research will be undertaken (Gable, 1994; Leedy & Ormrod, 2005). A methodology for research is an operational plan derived from a research design. Additionally, it provides a more detailed description of the methodology used to conduct the research, including the characteristics of the data, the instruments used to collect the data, and the data collection process. The term "research design" refers to the numerous ways in which research can be conducted to address the research question (Geoffrey, 2019). Kazdin (2016) defines research design as "the experimental arrangement or plan used to investigate the research question or hypotheses of interest". Numerous different arrangements exist for evaluating differences in desired characteristics, including those in which experimental manipulation is performed (true experiments) or groups are formed (observational study).

After defining the research objective, we can move on to developing the research strategy and methodology. Numerous research methodologies are valid in the fields of knowledge discovery and information systems. Case studies, field studies, action research, prototyping, and scientific methods such as experimentation are all examples of these methods. Given that this research is aimed at developing robust mechanisms for the knowledge discovery system, these mechanisms or proposed theories must be validated using the traditional scientific method of experimentation. As a result, the research method chosen is a scientific experimental approach combined with action research. The primary characteristics of each approach (scientific method and action research) will be discussed, as well as why they were chosen for this research. In this section, the proposed research methodology for this research is discussed.

3.1.1 Scientific approaches

Scientific approaches are methods that have developed from the scientific tradition. They are characterised by the capacity to repeat experiments, break down complex phenomena into simpler parts, and be proven wrong. These techniques assume that observations of the phenomena being studied are objective and rigorous (Checkland & Holwell, 1998; Geoffrey, 2019; Klein & Lyytinen, 1985). Numerous disciplines share the scientific method. "The research procedures of most academic disciplines follow the dictates of the scientific method. In many instances, only the tools of research are different. The basic procedure of each is the same: to process the data, interpret them, and reach a conclusion based on factual evidence."(Leedy & Ormrod, 2005).

3.1.2 Action research approach

General Action Research (AR) is viewed as a cyclical process(Järvinen, 2005; Susman, 1983). This procedure is partitioned into four distinct phases: plan, act, observe, and reflect. Its objective is to bridge the gap between theory and practice, thereby achieving both practical and research objectives(Järvinen, 2005; Susman, 1983). We chose action research as a primary research approach because it (1) provides a general framework and methodology for conducting research activities in a logical and efficient manner; and (2) action research possesses the strength of evaluation and reflective learning. Evaluating and reflective learning refers to the ability to take a step back and critically analyse an action, decision, or product by focusing on what was done or is being done that incorporates learning to be applied to a new situation. We maintain that rigorous, reflective thinking and thorough documentation of the research process enable effective learning, particularly during the prototype system development phase.

3.2 Research design and approach

The proposed study's research design can be classified as (1) exploratory, (2) observational, (3) experimental, and (4) descriptive. The proposed research methodology, which is derived from the research design, leverages the strengths of both the scientific (empirical) and action research approaches to accomplish the research objectives. It is divided into seven distinct phases. (1) Literature review; (2) Developing a conceptual framework; (3) Developing theoretical models; (4) Choosing a standard dataset; (5) experimental configuration including building a prototype system and carrying out several experiments; (6) Conducting several experiments, evaluating them in the laboratory, and reflecting; (7) interpreting and analysing results, and thesis writing. Figure 3.1 depicts the schema of the research design.

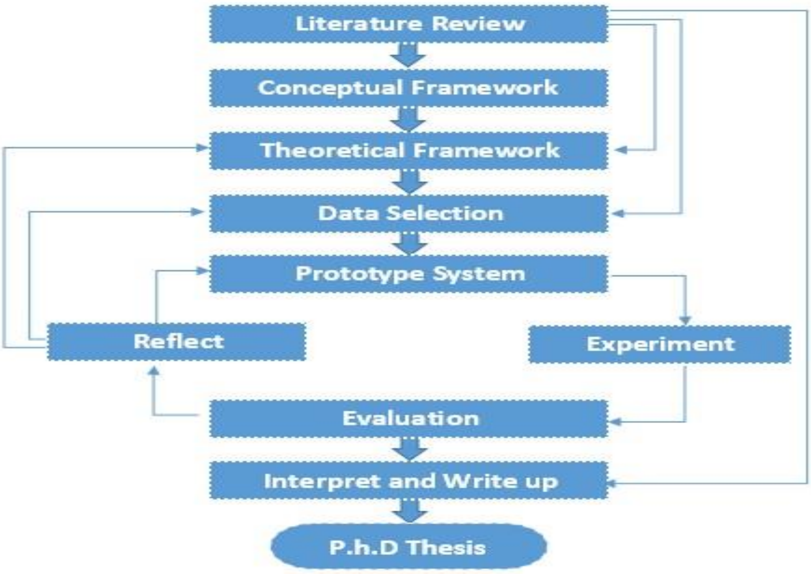


Figure 3.1 *Research design schema.*

3.2.1 Literature review

It is critical for researchers to understand the types of studies that have been conducted in particular areas in order to determine whether their specific research questions have been addressed previously (Geoffrey, 2019). The literature review process entails the researcher examining the available literature to ascertain the status quo, formulating a problem or research enquiry, defending the value of pursuing the established line of enquiry, and comparing the findings and ideas to his or her own. The outcome of this effort is a synthesis of other people's work in a way that demonstrates the exploratory process's success. This phase examines emerging fields for proposed research by examining potential significant issues/problems, relationships, and relevant theories identified through prior research. The primary responsibilities are to conduct critical analysis and evaluation of the literature. Many of the main relevant keywords are Fraud; Cyber fraud; Credit card fraud, Credit fraud; Machine learning; Deep learning; Cybersecurity; Banking industry; Financial transactions; Cybercrime; Supervised learning; Unsupervised learning; Fraud detection; Cyber fraud detection, these keywords will be included in the literature review.

3.2.2 Conceptual framework

Following the study of the literature, an overall research roadmap can be given as a conceptual model for identifying cyber fraud in the banking industry. After determining the issue to be addressed, a preliminary literature review serves to provide a conceptual framework for developing a ML and DL-based cyber fraud detection model. DL techniques such as CNN, LSTM, and BiLSTM are used because of their capacity to minimise overfitting and uncover underlying cyber fraud tendencies, and the capability to manage enormous datasets and the strength in short term data sequence learning. CNN exhibits strong capabilities in extracting patterns from short-term sequences in the data and possesses the ability to effectively process large-scale information. LSTM is capable of learning long-term sequence dependencies and predicting a transaction label based on the sequence of prior transactions by establishing feedback connections between concealed units that are connected with discrete time steps. There are three unique gates that are responsible for updating the information that is stored in a memory cell within the LSTM unit. These gates are the input gate, the forget gate, and the output gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell. BiLSTM is a sequence processing model that consists of two LSTMs: one processing the input in a forward direction, and the other processing it in a backward direction. The conceptual framework for this research can be created similarly to that depicted in Figure 3.2.

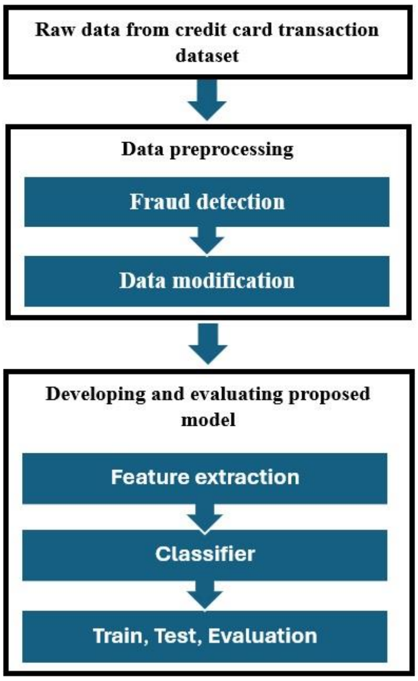


Figure 3.2 *Proposed conceptual framework.*

3.2.3 Theoretical model

This section will develop a new model based on prior research to address the research questions. The goal of this section is to develop and design a novel mechanism based on a literature review to address the challenges associated with detecting cyber fraud in the banking industry. This proposed model needs to be validated to answer all research questions defined in the Introduction section. As mentioned in the literature review section, recently researchers have found that DL techniques such as CNN, LSTM, and BiLSTM are used because of their capacity to minimise overfitting and uncover underlying cyber fraud tendencies, and the capability to manage enormous datasets and the strength in short term data sequence learning (Benchaji et al., 2021; Nguyen et al., 2020).

3.2.4 Data selection

Datasets play a crucial role within the domain of ML. Collecting data may be a challenging endeavour, particularly in domains that relate to finance, such as credit card cyber fraud. In order to evaluate the efficacy of the novel ML and DL models in detecting instances of cyber fraud in credit card transactions, a well-recognised dataset was selected for training and testing purposes. This dataset available at the following link: https://www.kaggle.com/mlg-ulb/creditcardfraud. The dataset was created by the ULB Machine Learning Group. The data set utilised is that of client transactions in a European bank in 2013. In a 24-hour period, the dataset was collected from the real world and comprises 284,807 card transactions. It consists of 31 columns, 30 of which are features and one of which is a target class that ascertains whether or not the transaction is fraudulent. The dataset only comprises numerical (continuous) input variables that have been derived through a feature selection process using PCA. Features are concealed as a result of privacy and confidentiality concerns. The features 'Amount' and 'Time' have not been the only ones to undergo transformation in PCA. The amount of time that has passed between the first entry and each succeeding entry is specified by the 'Time' variable, which stores the number of seconds that have passed. The 'Amount' feature is what ends up determining the final amount of transaction. Variable ‘Class’ is assigned a value of 1 in the event of deception and 0 in the absence of fraud. There are 284,807 transactions in the real-world dataset, with 99.828% of them being legitimate and 0.172 % of them being fraudulent. Fraudulent Transactions is 492 and non-Fraudulent Transactions is 284315. A summary of the research dataset utilised in this research is provided in Table 3.1 below.

Table 3.1 *A summary of the utilised dataset.*

Source	https://www.kaggle.com/mlg-ulb/creditcardfraud
Type	Real-world / Numerical
Size	284,807
Category	Extremely imbalanced
Feature	The majority of the data has been anonymised.
Features	31
Features	Amount: Transaction amount
	Time: Total of time (in seconds) elapsed between two transactions (First and the current one examined).
	Class variable: (0 equals no fraud, 1 equals fraud)
	[V1 … V28]: The output obtained from doing Principal Component Analysis (PCA) for dimensionality reduction on the original raw data is utilised to safeguard the identity of users and secure critical characteristics.

3.2.5 Experimental configuration and results

A prototype modelling system is created to train, test, and assess the improved ML and DL cyber fraud detection model to address research problems four and five. Modelling tests were carried out to confirm the efficacy and efficiency of the suggested models in chapters 4, 5, and 6. The algorithms were run on a PC with an Intel Core i7 processor running at 3.3 GHz and 16 GB of RAM. Python software, originally described by Sanner in 1999, is utilised for model creation and prototyping due to its accessible libraries for DL such Keras (Ketkar, 2017), TensorFlow (Abadi et al., 2016), Scikit-learn (Pedregosa et al., 2011), and Matplotlib(Hunter, 2007). Keras enables quick and efficient prototyping and is compatible with both convolutional and recurrent networks. Matplotlib is a Python package used for 2D charting and statistical analysis of modelling data.

Python version 3.11.4 is employed for the building of different learning models. The decision to utilise the Python programming language was made based on intuition, as Python is widely employed in the field of cybersecurity. The code is executed within the computing environment of Jupyter Notebook. The Jupyter notebook is a JSON document that consists of a structured collection of Input cells for code execution and Output cells for displaying various forms of output such as numerical data, text, graphs, and plots. The Jupyter Notebook facilitates user interaction with code through the web client, allowing for code execution in chunks or as a whole unit.

3.2.6 Evaluation and reflection

Evaluating the outcomes is crucial in research to determine the efficacy and efficiency of the suggested models and address research objectives 4 and 5. Validity pertains to how well the metrics in this research 's dataset accurately capture and quantify the intended concept of predicting credit card cyber fraud. Validity is the degree to which research's approach effectively evaluates the construct it claims to measure. The efficacy of models in detecting instances of credit card fraud in a particular dataset is evaluated in this study through the use of ML and DL techniques. In order to obtain accurate predictions, we implement dataset balancing strategies.

Various measurement systems were utilised in this study to assess the performance of the models. In the process of assessing ML models, it is common to first train the models using a designated set of training data. Subsequently, these models are subjected to testing using a separate set of data, known as testing data, to ascertain their capacity to generalise beyond the training data. Furthermore, to considering the inherent imbalance that exists within the credit card datasets, the major emphasis of our research was on the assessment criteria that are related with the binary classification assignment of credit card cyber fraud. Initially, the algorithms were evaluated by dividing the training and testing datasets using k-fold cross validation, where k=5. Cross-validation is computationally intensive and uses all cases as training and test examples. The method employs a technique that involves training and testing sets. It trains the algorithm K times, each time using a fraction 1/K of the training instances for testing. In practice, the data set D is first chunked into K disjoint subsets (or blocks) of the same size $m \triangleq n/K$. If Tk for the Kth such block, and Dk the training set obtained by removing the elements in Tk from D. The cross-validation estimator is defined as the average of the errors on test block Tk obtained when the training set is derived from Tk: C$V(D)$ = 1 K ∑ 1 m K k=1 ∑ L ($A(Dk$), zi) (Bengio & Grandvalet, 2003).

In order to evaluate the efficacy of various ML algorithms, assessment metrics such as confusion matrix, accuracy, precision, recall, F1-score, and AUC are implemented. Confusion matrix was implemented to illustrate a variety of metrics that evaluate the trade-off between specificity and selectivity, with the objective of reducing the time required for both Type I and Type II errors. Confusion matrix is widely regarded as a straightforward and effective method for assessing outcomes. It enables the comparison of expected and observed data to represent the number of accurately classified data instances, employing four distinctly different values:

True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). In the event that both the observed and anticipated values are positive, such as in the case of credit card cyber fraud, is referred to as TP. A scenario TN is one in which the observed and predicted statistics both suggest negativity, such as the absence of credit card cyber fraud. FP denotes a situation in which the anticipated value is positive, despite the fact that the observed value is negative. A false negative (FN) is the opposite of a false positive (FP).

Accuracy, which is the reciprocal of the error rate, is a widely employed performance measure for evaluating algorithms in classification tasks. However, it may not be the most appropriate statistic for unbalanced datasets. By accurately identifying such occurrences, the recall metric demonstrates that the classifier is effective in reliably recognising actual incidents of credit card fraud. Recall is the percentage of instances within a specific category that the model is capable of accurately identifying. There is a positive correlation between recall and the number of occurrences recovered from the minority group. The efficacy of the classifier in accurately identifying genuine instances of credit card cyber fraud is exemplified by the measure of recall. The concept of precision refers to a quantitative measurement that assesses the effectiveness of a hypothesis within the context of a hypothetical situation in which the anticipated result is positive.

The F1-score is a statistic that is frequently represented as the harmonic mean and includes both recall and accuracy. The F1-score is a widely employed metric for evaluating information retrieval systems, including ML models (Rijsbergen, 1979). The Area Under the Curve of the Receiver Operating Characteristic (AUC-ROC) is a widely employed assessment tool in situations involving unbalanced data (Hanley & McNeil, 1982; Metz, 1978). The training examples may be evaluated using the F1-score metric, which combines precision and recall. Furthermore, a curve can be constructed to demonstrate the correlation between the false positive rate and the recall. The model's efficacy enhances as the area under the curve (AUC) value approaches 1. Table 3.2 provides a thorough summary of the performance metrics.

Table 3.2 *Performance indicators.*

Metrics	Description	Equation	Range
Accuracy (A)	Examine the number of TPs	$A = \frac{TN + TP}{TN + FN + TP + FP}$	[0-1]
Recall	The proportion of TP to a TP and FN	$R = \frac{TP}{TP + FN}$	[0-1]
Precision	The proportion of TP to a TP and FP	$P = \frac{TP}{TP + FP}$	[0-1]
F1- Score	The Combination of precision and recall	$F_1 = 2\frac{P * R}{P + R}$	[0-1]
AUC	The area between two coordinates that are bounded by the function and the x-axis.	$AUC = \int_a^b f(x)dx$	[0-1]

3.2.7 Interpretation and write up

The results of the literature review, experimental results, and model development were jointly analysed, interpreted, and reported. The research results were described in publications that were suitable for the research.

3.3 Chapter Summary

This chapter provides a discussion of the principles and forms of research methodology employed in the present PhD thesis. This section provides an explanation of the construction of the research design framework, including the stages involved and the components utilised. This research introduces and clarifies the used databases, configuration setups, and evaluation measures utilised in the suggested models. The objective of this chapter is to present a comprehensive summary of the material covered in Chapters 4, 5, and 6, which will be further discussed in subsequent chapters. The subsequent chapters provide a description and analysis of the suggested models for detecting cyber fraud in credit card transactions.

CHAPTER 4: NOVEL HYBRID STACKING ML MODEL

This Chapter describes the successful development of a hybrid ML approach that integrates DT, RF, SVM, XGBoost, CatBoost, and LR with an ensemble learning technique. This approach addresses the current challenges of detecting cyber fraud in credit card transactions by combining numerous base models. Stacking is an ensemble learning technique that combines multiple base models to improve predictive performance. In this implementation, base models include diverse algorithms such as DT, RF, SVM, XGBoost, CatBoost, and LR. The meta-model (final estimator) makes the ultimate prediction by aggregating predictions from base models. As a consequence of the stacking classifier's ability to leverage on the benefits of individual models while simultaneously resolving the shortcomings of those models, it has the potential to achieve superior performance in comparison to any single model. The research applies the credit card dataset, to train the algorithm and subsequently, detect the cyber fraud. To attain this, implements a correlation-based feature selection technique to identify and extract the most relevant features for predicting a target variable. To improve the computational efficiency of the algorithm, the full dataset was reduced to only the most significant input features by applying the PCA, and the outputs of the selected features were then transferred to the new model that combine ML techniques DT, RF, SVM, XGBoost, CatBoost, and LR using stacking ensemble techniques. The research's novelty is the development of a novel ML model that can extract and select the most prominent features and classify transactions as either cyber fraud or non-fraud. The novel model was subsequently conditioned to further tuned the outputs of the ML techniques to classify transactions. The contributions of this research work will result in:

1. Efficient and effective algorithms are designed to extract the most important features by using correlation-based, Random Forest, XGBoost, and Permutation Feature Importance for feature selection technique and the PCA dimension reduction method.
2. A novel hybrid ML model that automatically classify transaction to fraud and non-fraud is created.
3. The novel hybrid ML with stacking model consists of both the feature extraction and the classification algorithm in a single workflow that that can easily identify credit card fraud.

Figure 4.1 illustrates a block-diagram of the proposed modelling framework that was developed in this chapter. In essence, it is composed of three primary components that are designed to enhance the algorithm's overall effectiveness. The pre-processing stage was where the original dataset was transferred. This stage applied procedures to remove any data cleaning and missing values, as well as normalizing techniques, to adjust the dataset before it was incorporated into the feature extraction phase and the model training stage. The fine-tuned pre-trained was subsequently employed to extract these features in a new proposed framework for feature selection extraction. This was later output into the PCA stage, aimed to reduce the dimensionality of the extracted features from the dataset. The following subsection provided an explanation of the novel hybrid stacking ML model.

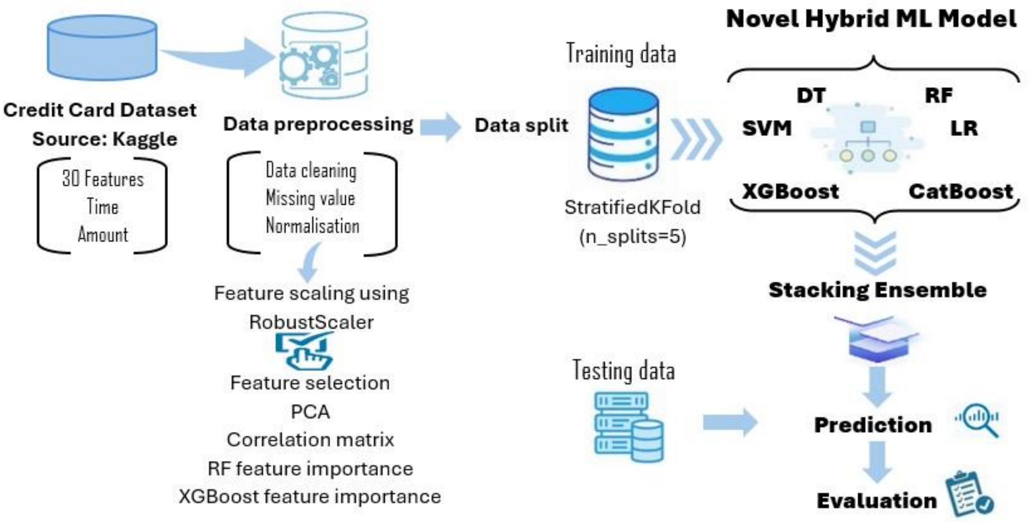


Figure 4.1 *The novel hybrid stacking ML model.*

The fundamental principle underlying a hybrid classifier is the integration of several models in order to enhance the overall stability and predictive capability of the composite model. The selection of six techniques, namely DT, RF, SVM, XGBoost, CatBoost, and LR, is based on their shown improved prediction accuracy in binary classification issues. The data is subjected to pre-processing through the normalisation process. Furthermore, feature selection is implemented to mitigate the detrimental effects of the curse of high dimensionality. Undesirable features have been deactivated. Next, the ML classifiers are created and undergo training. Subsequently, the categorisation of credit card cyber fraud is performed using a meta-model, namely LR , which is selected as the ultimate estimator to combine the predictions of the base models.

4.1 Data processing

To facilitate the development of the cyber fraud detection model, a dataset containing credit card transaction information was obtained and loaded into the analysis environment. The dataset, referred to as 'creditcardfull.csv,' contains transaction records and relevant features. The initial exploration of the dataset allows us to gain insights into the structure of the data and the available features. This step is essential for understanding the nature of the data and determining the necessary pre-processing and feature engineering steps required for model development.

Credit card transactions totalling 284,807 were conducted in the real world over the course of 24 hours. There are 31 columns in this dataset, 30 columns representing attributes and one column representing the target class, which shows whether a transaction is fraudulent or genuine. The dataset comprises 28 variables, denoted as $\{V1, V2, . . . , V28\}$, which have been converted from the original set of variables using principal component analysis (PCA). Dataset file is formatted in Comma-Separated Values (CSV) format. The CSV file can be read using the panda’s module in Python.

Table 4.1 presented below displays the first five rows of the dataset, together with a subset of values for the features up to V9. The graphic displays numerical values for all the features, with PCA values being withheld for confidentiality reasons. The dataset includes a time characteristic, denoted as "time," which is followed by variables V1 to V28. The variable "Amount" is then included, and the dataset concludes with the variable "Class."

Table 4.1 *Top 5 rows of the dataset.*

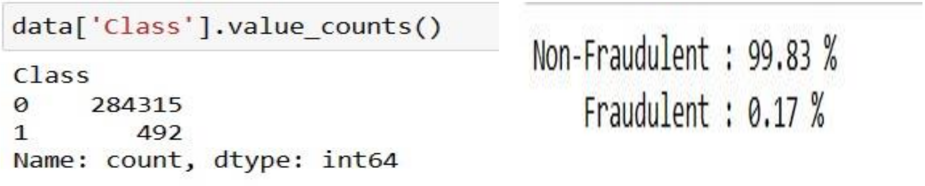
Time	V1	V2	V3	V4	V5	V6	V7	V8	V9
0.0	-1.35907	-0.072781	2.536347	1.378155	-0.338321	0.462388	0.239599	0.098698	0.363787
0.0	1.191857	0.266151	0.166480	0.448154	0.060018	-0.082361	-0.078803	0.085102	-0.255425
1.0	-1.358354	-1.340163	1.773209	0.379780	-0.503198	1.800499	0.791461	0.247676	-1.514654
3.0	-0.966272	-0.185226	1.792993	-0.863291	-0.010309	1.247203	0.237609	0.377436	-1.387024
4.0	-1.158233	0.877737	1.548718	0.403034	-0.407193	0.095921	0.592941	-0.270533	0.817739

Upon examining the dataset, the subsequent task was determining the distribution of the data and visualising its statistical characteristics. Table 4.2 illustrates the statistical measures of mean, maximum, and minimum values that may be derived through the utilisation of the describe () function.

Table 4.2 *Statistical measures for dataset.*

	Time	V1	V2	V3	V4	V5	V6	V7	V8
count	**284807.00000**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**	**2.848070e+05**
mean	**94813.859575**	**1.168375e+05**	**3.416908e+00**	**-1.379537e+15**	**2.074095e+05**	**9.604606e+00**	**1.487371e+15**	**-5.556476e+16**	**1.213418e-15**
std	**47488.145955**	**1.958569e+00**	**6.161530e+00**	**1.516255e+00**	**1.418569e+00**	**1.380247e+00**	**1.332271e+00**	**2.661650e+00**	**1.356254e+01**
min	**0.000000**	**-6.540751e+01**	**-7.271735e+01**	**-4.832559e+01**	**-1.171353e+01**	**-1.137433e+01**	**-2.616105e+01**	**-4.357742e+01**	**-1.343047e+01**
25%	**54201.500000**	**-9.203734e+00**	**-9.584998e+00**	**8.903484e+01**	**0.848401e+01**	**-9.915977e+00**	**-7.829569e+00**	**-5.540750e+01**	**-2.086279e+01**
50%	**84692.000000**	**1.810882e+00**	**6.548556e-02**	**6.849463e+00**	**4.433585e+00**	**-5.433583e+00**	**-2.741871e+00**	**-2.225804e+00**	**-2.355890e+01**
75%	**139532.00000**	**3.315428e+00**	**8.307232e+00**	**1.027196e+00**	**7.433413e+01**	**3.905486e+00**	**1.119624e+00**	**3.986548e+00**	**3.272530e+01**
max	**172792.00000**	**2.454930e+05**	**2.205773e+05**	**9.382558e+01**	**1.687534e+01**	**4.804617e+00**	**7.330163e+00**	**1.205859e+05**	**2.000721e+01**

Unbalanced data refers to categorisation difficulties characterised by disparate quantities of cases across distinct classes. The presence of imbalanced data is a prevalent occurrence in datasets of a generic nature. Cyber fraud detection systems often encounter an amount of highly imbalanced data, where most credit card transactions are genuine but just a small fraction is fraudulent. The Plotly Python graphing package was utilised to build a bar graph. The class variable was assigned to the X-axis, transactions were assigned to the Y-axis. In order to assess the presence of data imbalance in relation to the goal variable, namely the class, let us proceed with an examination of the data. Figure 4.2 shows the number of non-fraud and fraud transactions . Figure 4.3 shows the percentage of fraudulent vs non- fraudulent transactions. The imbalanced distribution of classes is shown in a bar chart, as depicted in Figure 4.4.



```
data['Class'].value_counts()
```

```
Class
0    284315
1       492
Name: count, dtype: int64
```

```
Non-Fraudulent : 99.83 %
    Fraudulent : 0.17 %
```

Figure 4.2 *The number of non-fraud and fraud transactions.*

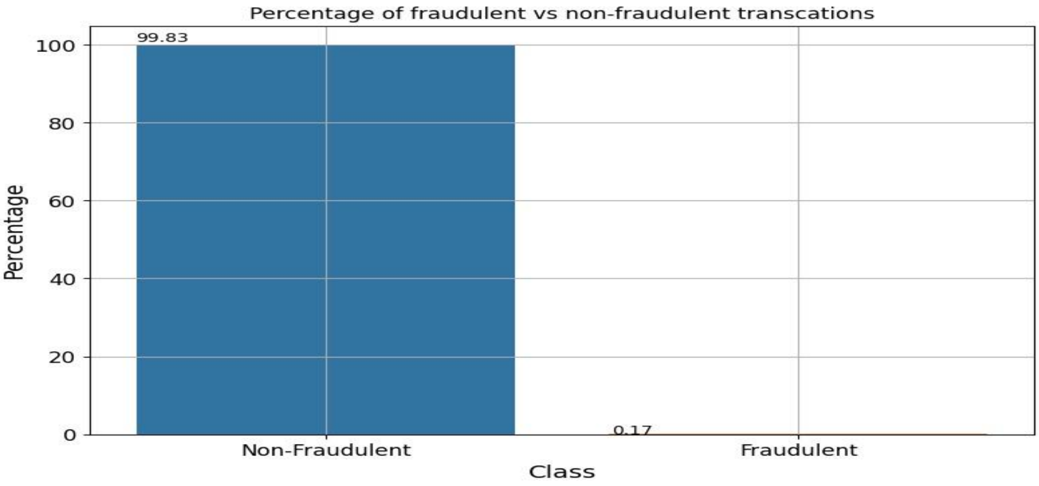


Figure 4.3 *The percentage of fraudulent vs non- fraudulent transactions.*

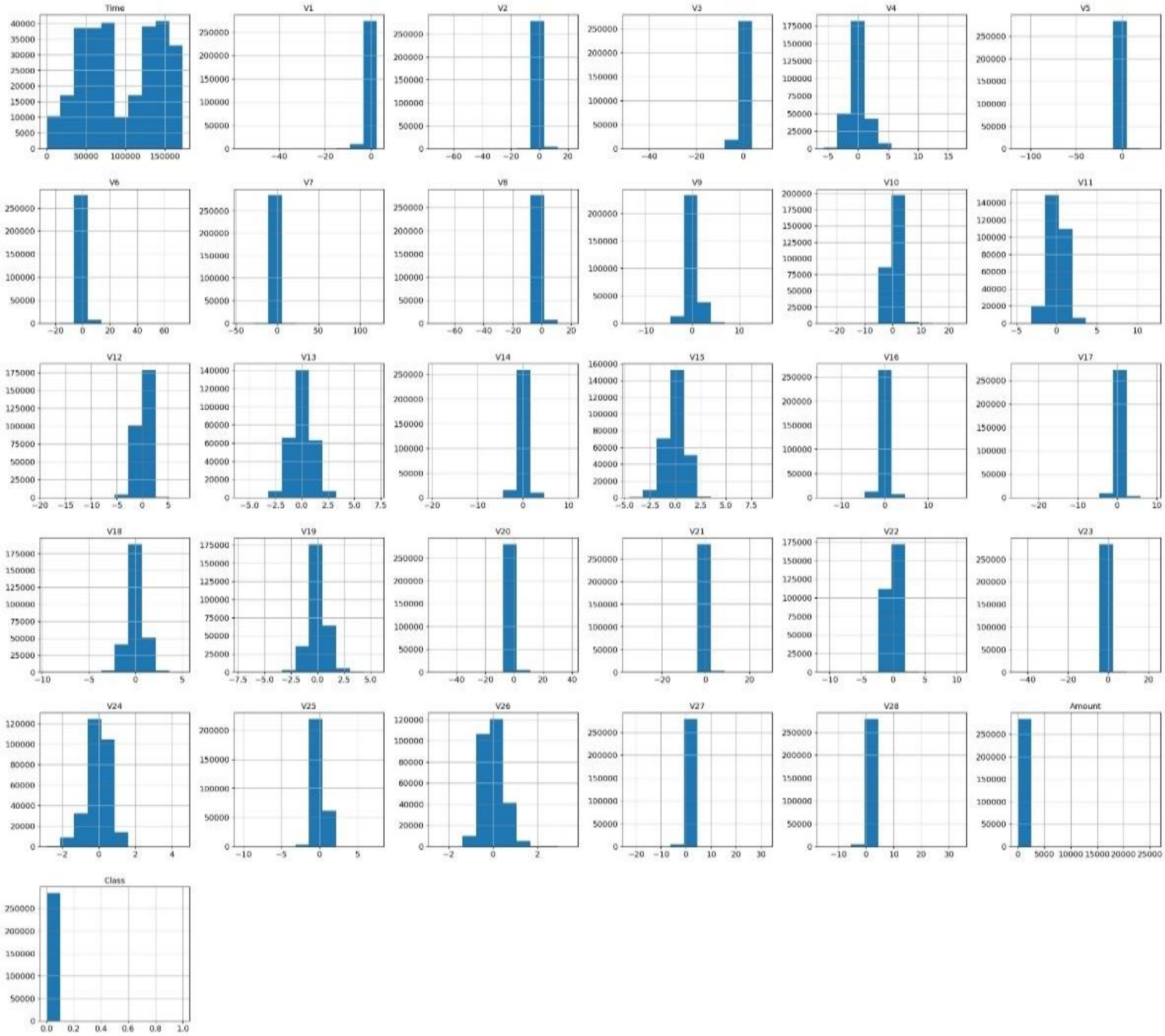


Figure 4.4 *The imbalanced distribution of classes.*

The dataset has very high-class imbalance. Only 492 records are there among 284315 records which are labelled as fraudulent transaction. To understand the data's characteristics, detect potential issues, and make informed decisions throughout the data science and ML workflow. Visualising the distribution of all variables in our dataset is a fundamental and exploratory step in data analysis. Figure 4.5 shows the plotting all the variable in displot to visualise the distribution. Observation we can see most of the feature’s distributions are overlapping for both the fraudulent and legitimate transactions.

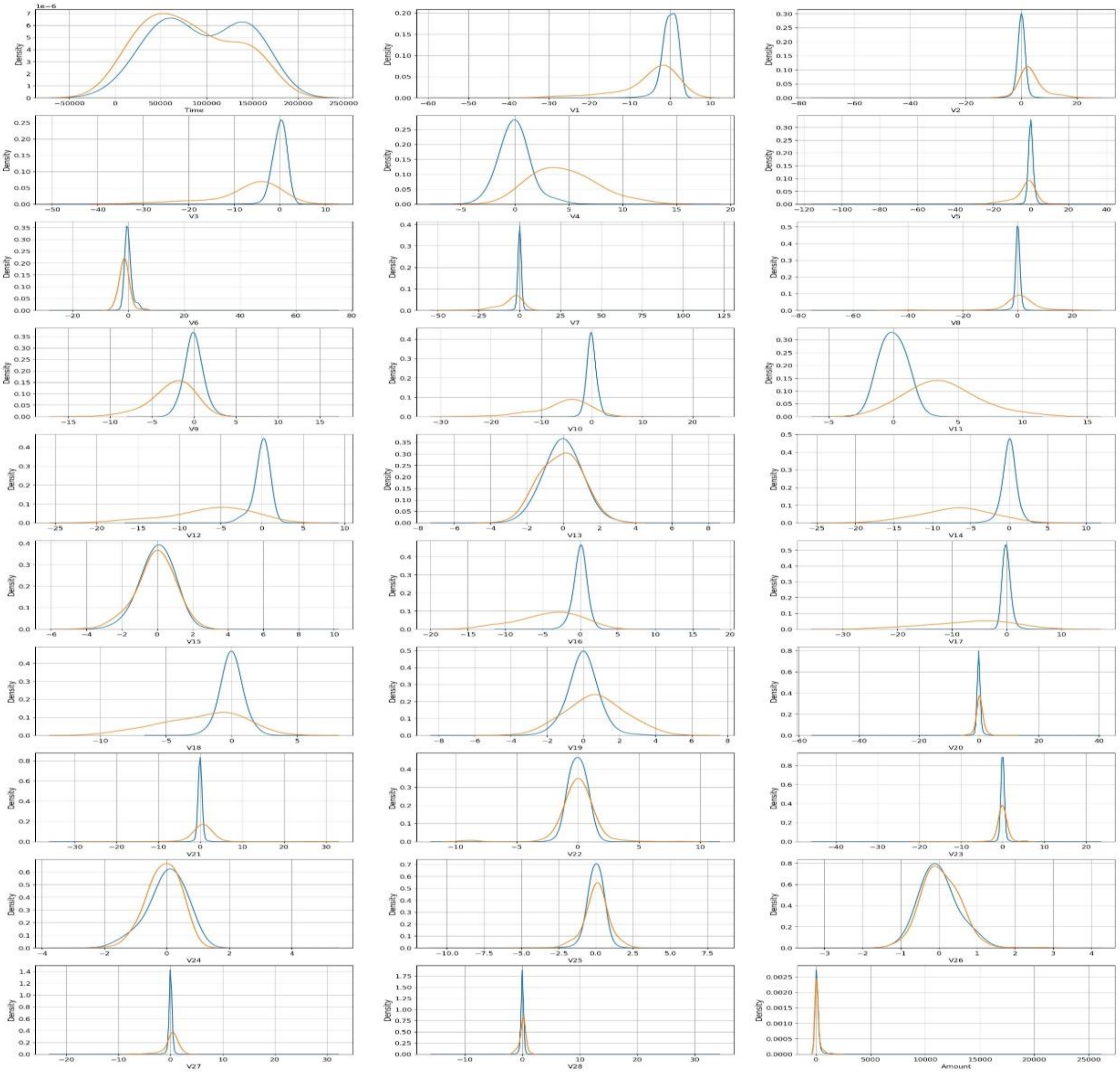


Figure 4.5 *Distribution of all variables.*

Certain features demonstrate a significant level of selectivity in their distribution for the two values of Class. Specifically, V4 and V11 have distinctly separate distributions for Class values 0 and 1. V12, V14, and V18 are partly separated. The profiles of V1, V2, V3, and V10 are quite distinct. V25, V26, and V28 exhibit comparable profiles for the two Class values.

In generally, the features distribution for genuine transactions (values of Class = 0) is centred on 0, occasionally with a lengthy queue at one of the extremities. Concurrently, the fraudulent transactions (values of Class = 1) exhibit a skewed distribution.

Creating boxplots to visualise the distribution of features concerning the 'Class' variable, representing fraud and non-fraud cases. By employing seaborn's boxplot function, each feature is compared across both classes, facilitating the identification of potential

variations or patterns between fraudulent and legitimate transactions. This method offers a clear and concise overview of feature distributions, aiding in the detection of outliers or distinctive characteristics associated with fraudulent activities. Through such visualisations, analysts can gain insights into the discriminative power of individual features and their relevance in distinguishing fraudulent transactions from legitimate ones. Figure 4.6 shows the subplots for visualising features for each cases fraud and non-fraud.

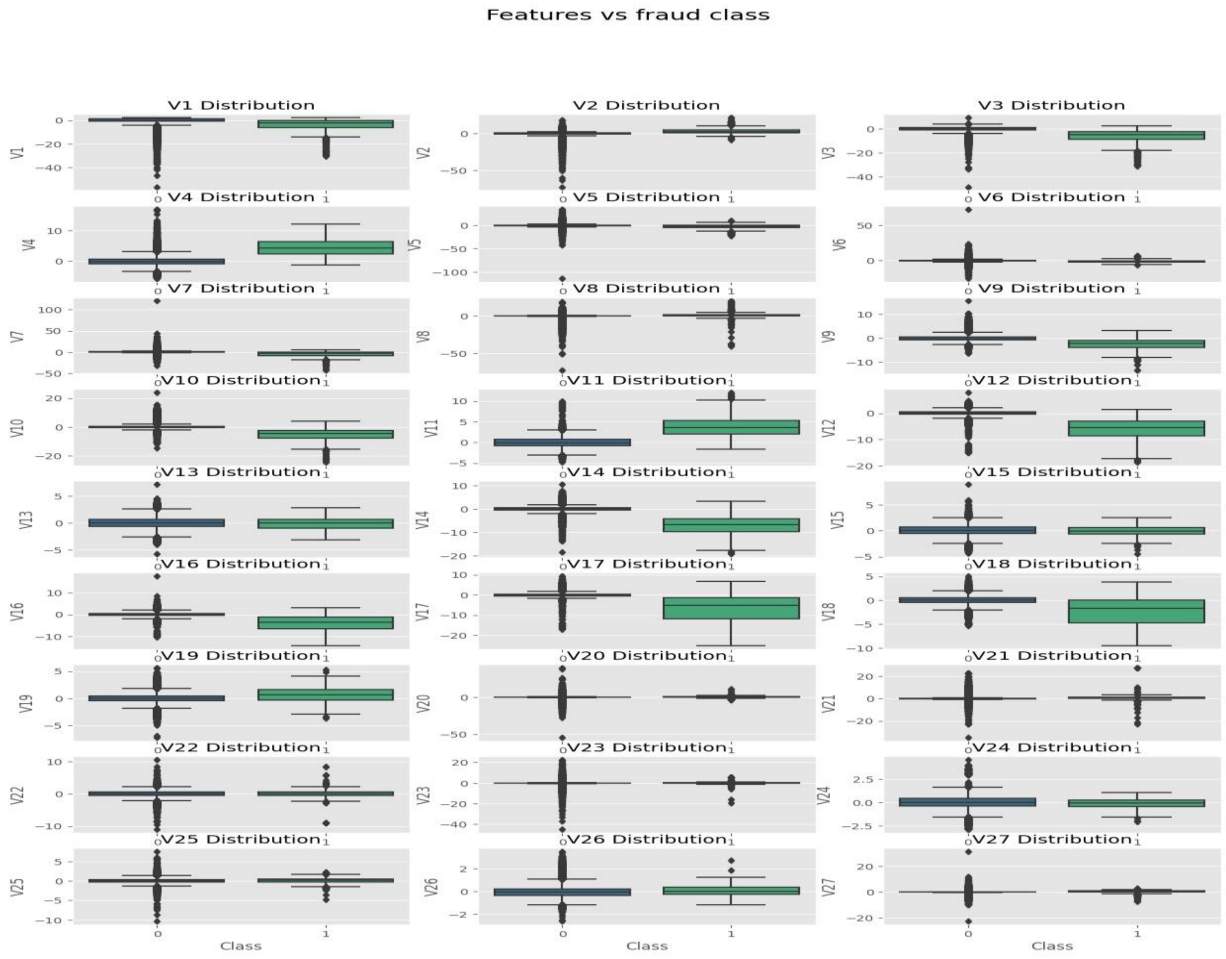


Figure 4.6 *The subplots for visualising features.*

Because of the dataset's extreme imbalance, with real transactions accounting for 99.827% and cyber fraud transactions for just 0.173%, training the model using raw data is unlikely to yield the desired outcomes, despite potentially high evaluation metrics. Figure 4.7 shows the Plot of credit card dataset.

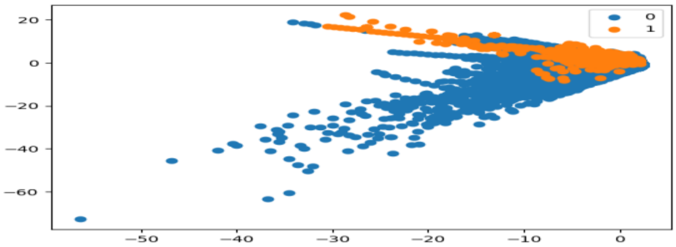


Figure 4.7 *The Plot of credit card dataset.*

It is imperative to pre-process the data before executing an ML algorithm. This is because the predictors are generated with unique requirements by various models, and the prediction output may be influenced by the data training. The primary objectives of data pre-processing are the cleansing and transformation of data to a state that exhibits reduced bias, identification and handling of missing values, and increased variability. The dataset is composed of numerical values that are acquired through the Principal Component Analysis (PCA) process. Nevertheless, the original characteristics have not been released because of the confidentiality concern. A total of 30 features were created, with 28 of them being derived by principal component analysis. PCA is a widely used method for reducing the dimensionality of data. It involves transforming a high-dimensional dataset into a lower-dimensional representation by identifying a smaller collection of feature variables that capture the most important information from the original variables. 'Amount' and 'Time' are the only attributes that have not been converted into principal components. The Pre-processing tasks have been accomplished by utilising the Python data manipulation package pandas and the ML module sci-kit learn. The sequential process is visually depicted in Figure 4.8.

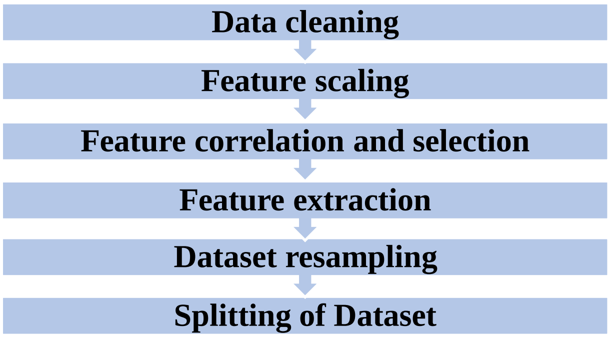


Figure 4.8 *The Data Pre-processing steps.*

4.1.1 Data cleaning

In the Python environment, the credit card dataset was imported by employing the appropriate import command. Subsequently, a comprehensive data cleansing procedure was implemented. In the course of data cleansing, two primary activities are frequently implemented. The initial task entails the elimination of null values and absent values from the dataset. The second responsibility is the management of outliers, which are data points that deviate considerably from the majority of the dataset. In total, the dataset contains 284,807 transactions. No null values were present in the dataset.

4.1.2 Feature scaling

For the purpose of normalising the range of independent variables within a dataset, this is an additional stage of the data pre-processing technique. Depending on the scaling technique used, it is centred around 0 or within the range of 0 to 1. Some ML algorithms may be overlooked or skewed by the extreme values of input variables that are pertinent to the additional input variables. The efficacy of a variety of ML algorithms and models can be enhanced by ensuring that all features have similar scales, which is implemented through feature scaling. Robust scaling is a good choice when data contains outliers. It scales features based on their interquartile range (IQR), making it robust to outliers. Using the Robust Scaler technique, which is also referred to as robust standardisation, we have implemented feature scaling. The 25th, 75th, and median 50th percentiles can be calculated to accomplish scaling. After the median of each variable is subtracted, the resulting values are divided by the interquartile range (IQR), which is the difference between the 75th and 25th percentiles.

This process performs scaling using the RobustScaler on the features in both the training and test datasets (X_train and X_test). The features in the training data (X_train) will be scaled using the RobustScaler fitted on the training data. This means that the mean and interquartile range (IQR) of the features in the training data will be used for scaling. The scaled values will replace the original features in the training data. This is a common Pre-processing step to ensure that the feature has the same scale in both training and test data when working with ML models.

4.1.3 Feature correlation and selection

Feature selection is a pivotal aspect of ML and data analysis, aiming to identify the most relevant subset of features from a dataset. Its primary goal is to enhance model performance, reduce computational complexity, and improve interpretability. By selecting informative features, feature selection addresses the challenge of the curse of dimensionality, where high-dimensional datasets pose difficulties for ML algorithms due to increased complexity and the risk of overfitting. One significant benefit of feature selection is its ability to improve model interpretability. By identifying key factors driving predictive performance, it enables stakeholders to gain insights and make informed decisions based on the underlying data patterns. Moreover, feature selection methodologies, including filter methods, wrapper methods, and embedded methods, offer different approaches to selecting the most relevant features for a given problem.

Filter methods assess feature relevance independently of the model, using statistical metrics such as correlation coefficients. In contrast, wrapper methods evaluate feature subsets iteratively through model training, employing strategies like forward selection or backward elimination. Model training is integrated with feature selection in embedded methods, which select features based on their contribution to model performance. The benefits of feature selection extend beyond model performance improvements. It also enhances model interpretability by identifying influential attributes and simplifies model architectures, leading to improved scalability and maintainability. However, feature selection is not without its challenges. The curse of dimensionality, feature interaction, redundancy, overfitting, and the need for domain expertise are some of the challenges that need to be addressed.

Despite these challenges, feature selection significantly impacts various aspects of data analysis, including exploratory data analysis (EDA), model development, and predictive modelling. It informs EDA by identifying key features for visualisation and analysis, guides model development by improving model interpretability, and enhances predictive accuracy, enabling data-driven decision-making. Feature selection remains a critical component of ML and data analysis, enabling organisations to derive actionable insights and drive business value from data-driven initiatives. As organisations continue to leverage advanced analytics, feature selection remains essential for data-driven innovation and maintaining a competitive advantage in the digital era.

Prior to commencing the implementation phase of a predictive analysis, it is advisable to first engage in data visualisation. The data can be visually represented by analysing the correlation between variables, which is indicative of the interdependent relationship between the variables. In general, feature variables that have a greater correlation with the response variable tend to exert a more pronounced influence during the training phase. The correlation matrix depicted in Figure 4.9 provides an analysis of the pairwise correlation among the variables. The provided correlation matrix indicates that there is no significant link seen among the major components V1 to V28. Upon further examination, it is evident that the response variable 'Class' exhibits both positive and negative correlations with the main components.

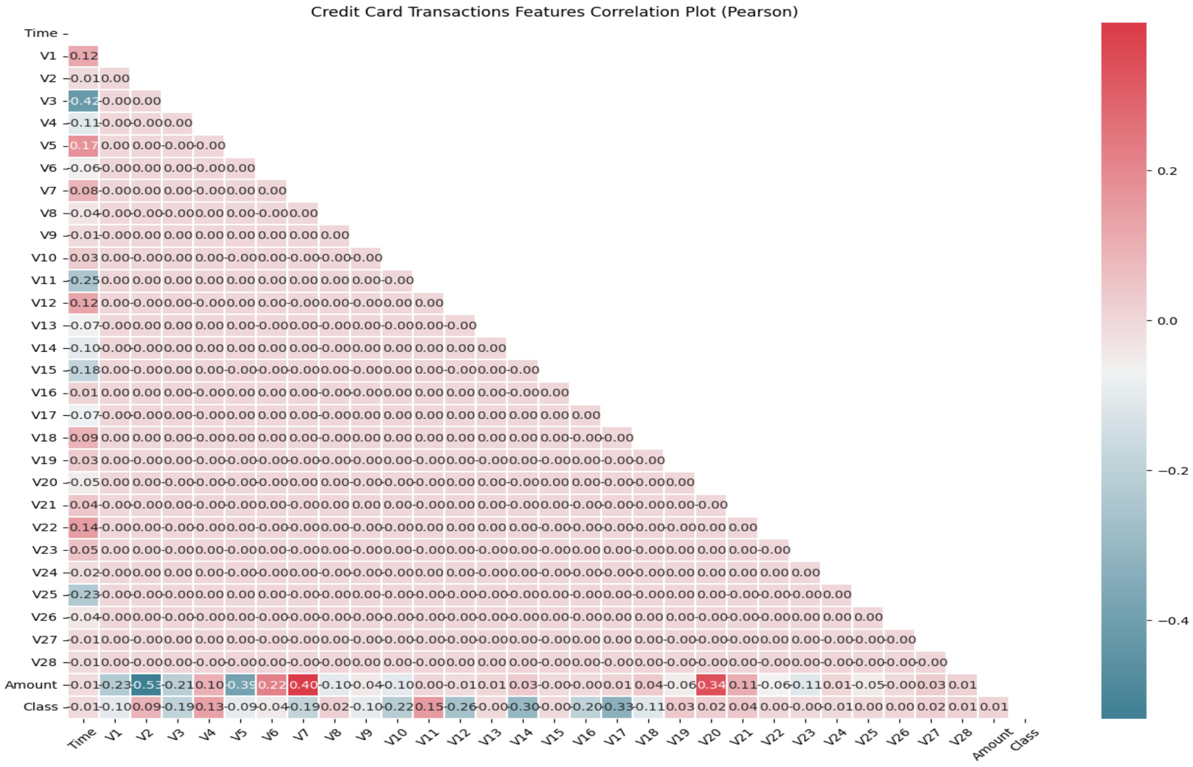


Figure 4.9 *Feature Correlation Graph.*

To conducts feature importance analysis and visualises the correlation matrix of features, specifically in the context of credit card transactions. Initially, it computes the feature importance by calculating the Pearson correlation coefficient between each feature and the target variable 'Class'. The resulting correlation values are sorted in descending order to identify the most influential features in predicting credit card cyber fraud. Subsequently, a heatmap of the correlation matrix is generated using Seaborn's heatmap function. This heatmap illustrates the pairwise correlations between all features, employing a diverging colour palette to distinguish positive and negative correlations. Additionally, a mask is applied to conceal the upper triangle of the matrix, eliminating redundant information. By visualising the correlation matrix, analysts can discern the relationships between different features and their predictive significance, aiding in feature selection and model interpretation.

Furthermore, Figure 4.10 presents the top 17 features alongside their importance scores, derived from the absolute correlation values with the target variable 'Class'. This insight enables a focused understanding of which attributes contribute most significantly to predicting credit card cyber fraud. Analysing the correlation matrix and feature importance facilitates the identification of informative features for modelling purposes, potentially leading to more efficient and interpretable models. By prioritising the most relevant attributes, this approach can enhance model performance and streamline computational resources. Ultimately, this feature analysis aids in constructing more accurate and effective predictive models. There are certain features that do not contribute to the modelling process, but rather increase the model's complexity. Consequently, we implemented a feature value evaluation to eliminate inefficient features.

An evaluation of the importance of each feature in the dataset is essential to get substantial values for all included characteristics. From the provided information, it was feasible to ascertain the extent of each component's contribution to the model. The significance of each feature in the model's development was demonstrated by the increasing relevance scores of each feature. The systematic classification of features was facilitated by the calculation of this importance score for each feature. Since Random Forest, XGBoost and permutation methods are often performed well for cyber fraud detection tasks, utilising them, we produced the feature importance graphs depicted in Figures 4.11, 4.12 and 4.13 below for dataset. Our comprehension of the data was enhanced by the interpretation and analysis of significant features. V17, V12, and V14 were among the most significant and beneficial features of the dataset when Random Forest was employed. The dataset's most significant and beneficial features were V14, V10, and V12 when XGBoost was employed. Upon conducting permutation, V14, V17, and V4 were among the most significant and advantageous features of the dataset.

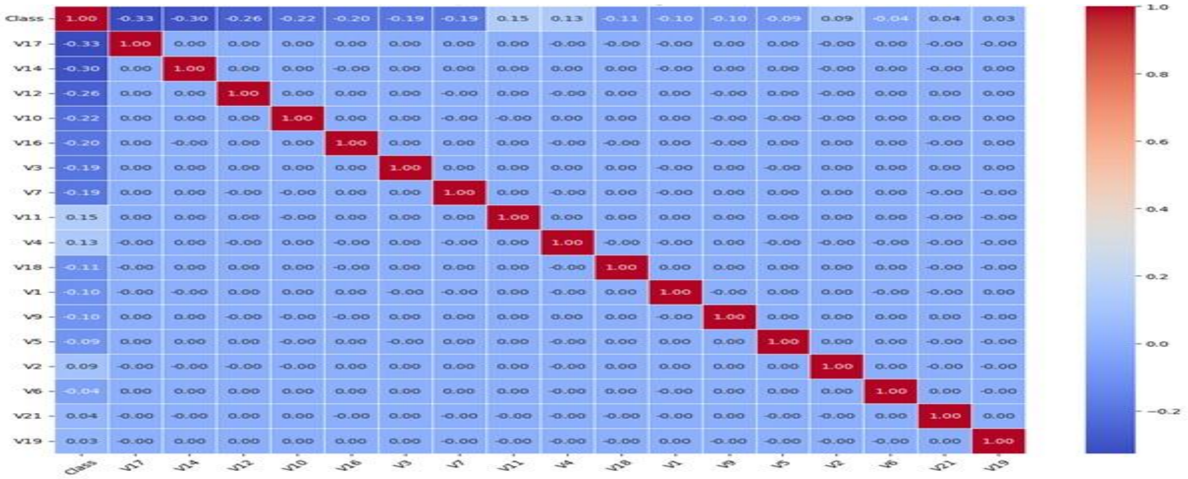


Figure 4.10 *Top 17 features alongside their importance scores.*

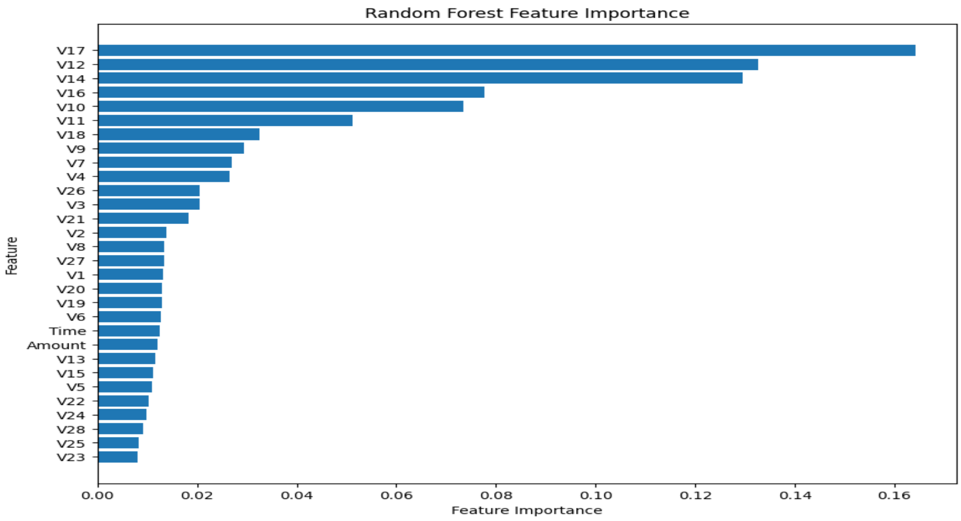


Figure 4.11 *RF Feature Importance.*

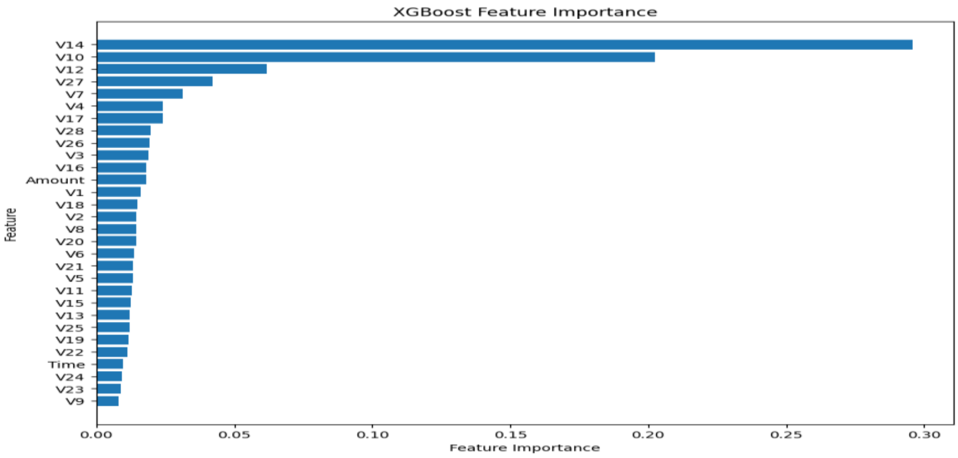


Figure 4.12 *XGBoost Feature Importance.*

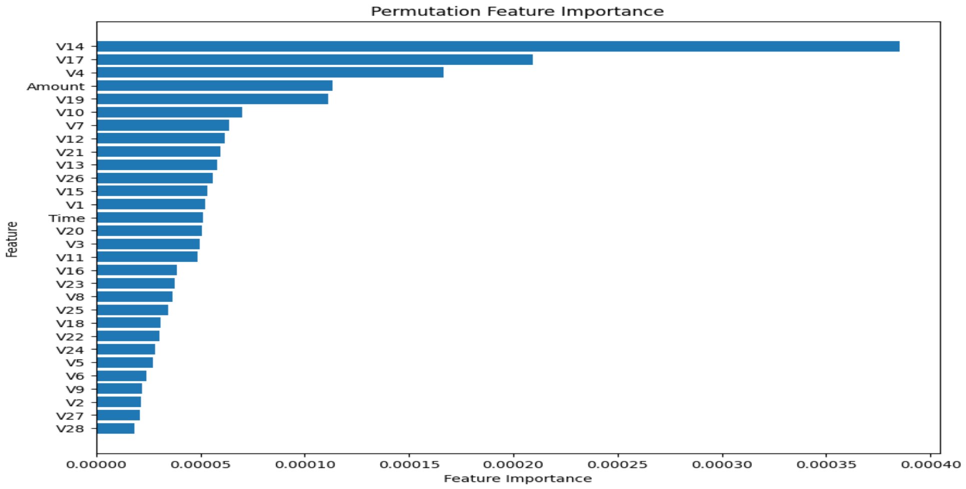


Figure 4.13 *Permutation Feature Importance.*

The model's features and identifiers were also chosen based on their relevance. It was beneficial to classify both independent and dependent variables and confirm all of the variables for the model. Utilising four distinct feature importance techniques, including the correlation matrix, Random Forest (RF), XGBoost, and permutation analysis, we evaluate the significance of features. Following the analysis with each method, we compare the results and aggregate the top 17 important features. The features are V17, V14, V12, V10, V16, V3, V7, V11, V4, V18, V1, V9, V5, V2, V6, V21, and V19. These selected features are then designated for utilisation in subsequent stages of the process.

4.1.4 Feature extraction

Feature extraction is a critical pre-processing step in ML that transforms raw data into a set of informative features, thereby improving model performance, efficiency, and interpretability. Technique like PCA is used for dimensionality reduction. Feature extraction helps reduce noise, prevent overfitting, and lower computational costs by selecting and transforming the most relevant data. Ultimately, this step enhances model accuracy and generalisation by simplifying the input data. We applied PCA as a dimensionality reduction algorithm on our dataset to produce robust and discriminative features for detecting fraudulent transactions. Figure 4.14 illustrates the PCA technique on the dataset. The data were reduced to three dimensions using default parameters, with plots coloured according to transaction labels: red for normal transactions and blue for fraudulent ones. PCA showed acceptable discrimination. Therefore, we selected PCA as the reduction algorithm for extracting embedding features for the training and testing phases.

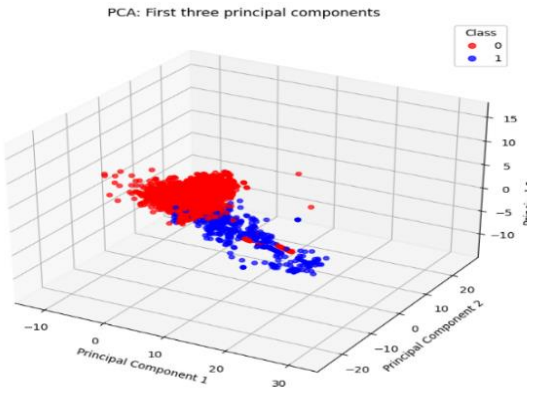


Figure 4.14 *Principal Component Analysis (PCA).*

4.2 Data splitting

The initial stage of the development of ML models is the splitting of data into training and testing sets. The process involves the division of a dataset into two distinct subsets, the training set and the testing set.

The training set comprises a significant amount of data and is utilised to instruct the ML model. The training set serves as the primary source of information for the model to acquire knowledge of patterns, correlations, and underlying structures inherent in the data. Fundamentally, this collection serves as the underlying basis upon which the model's prognostic capacities are constructed. Conversely, the testing set, including a smaller subset of the data, fulfils a distinct objective. The purpose of this evaluation is to determine the model's ability to apply its acquired knowledge to unfamiliar data. The testing set serves as a reference point, allowing for the assessment of the model's efficacy and its capacity to provide precise predictions on novel, unobserved data.

Splitting of data into distinct training and testing sets is a fundamental technique in the field of ML. This method serves two crucial purposes: mitigating the risk of overfitting and verifying the performance of the model in real-world situations. In addition, the process of randomising the data prior to splitting is of utmost importance to mitigate any potential biases or patterns inherent in the data that might potentially distort the outcomes. This procedure establishes the foundation for the training, assessment, and enhancement of models, ultimately resulting in the development of more efficient ML models. In each experimental iteration, the complete dataset is partitioned into a training set comprising 80% of the data and a test set including the remaining 20%. The training set was utilised for resampling, hyperparameter tuning, and model training, while the test set was employed to evaluate the performance of the trained model.

4.3 The novel hybrid ML model

The credit card cyber fraud detection model we have developed integrates advanced ML techniques, aiming to provide a strong and all-encompassing approach to detecting and mitigating fraudulent credit card transactions. The utilisation of several algorithms in this hybrid model enhances accuracy and flexibility, hence assuring its effectiveness in the dynamic area of financial cyber fraud.

The preliminary data pre-processing consists of separating features from the objective variable ('Class') and dividing the data into training and testing sets in an 80-20 ratio. After removing missing values and standardising features using StandardScaler, resulting in X_train and X_test, the imbalanced dataset undergoes division for ML algorithms (RF, DT, LR, KNN, SVM, XGBoost, and CatBoost). Evaluation metrics assess algorithmic performance. Cross-validation with 5 folds on training data presents scores, Next, predict test set labels and calculate classification metrics (precision, recall, F1 score, and AUC). A heatmap and ROC curve provide insights into classification accuracy and model effectiveness, contributing to the research's evaluation. The research investigates a range of algorithms, employing GridSearchCV to optimise them prior to assessment using critical metrics. The evaluative framework goes beyond AUC diagrams, addressing AUC limitations in severely imbalanced datasets, offering a comprehensive understanding of each algorithm's efficacy in cyber fraud detection.

4.3.1 Machine learning techniques

A. Decision tree (DT)

The Decision Tree (DT) is a supervised ML algorithm used for classification tasks. It is trained on a dataset consisting of features (X) and a target variable (y), with the objective of predicting the target variable based on the input features. In this specific implementation, the

DT classifier is configured with the following hyperparameters; 'max_depth': 70, 'min_samples_split': 8, 'min_samples_leaf': 10, and 'criterion': "entropy".

The 'max_depth' parameter controls the maximum depth of the decision tree, limiting its complexity and preventing overfitting. 'Min_samples_split' specifies the minimum number of samples required to split an internal node, while 'min_samples_leaf' sets the minimum number of samples required to be at a leaf node. These parameters help regulate the size of the tree and improve its generalisation capability. Additionally, the 'criterion' parameter determines the function used to measure the quality of a split. In this case, "entropy" is chosen, which computes the information gain based on the entropy.

In summary, In order to produce a tree-like structure, the Decision Tree model employs recursive partitioning. Each interior node in this structure is a representation of a feature, each branch is a representation of a decision that is based on that feature, and each leaf node is a representation of a class designation. By tuning hyperparameters and evaluating performance metrics, this model provides insights into the dataset's underlying patterns and facilitates accurate classification.

B. Random Forest (RF)

The Random Forest (RF) model is an effective ensemble learning technique that is employed for classification tasks. During the training process, it generates multiple decision trees and outputs the mode of the classes (classification) or the mean prediction (regression) of individual trees. Here's a detailed analysis of the model: The dataset is initially split into features (X) and the target variable (y). Subsequently, the data is further partitioned into training and testing sets using an 80-20 split ratio to facilitate model evaluation. The Random Forest Classifier is initialised with specified hyperparameters, including: 'n_estimators': This parameter determines the number of trees in the forest. In the provided code, n_estimators are set to 100. Having a larger number of trees typically leads to better performance as it reduces overfitting and improves the stability of the model. However, increasing this number indefinitely can also increase computational costs. The maximum depth of each forest decision tree is controlled by the max_depth parameter. Deeper trees can capture more complicated data linkages. Set max_depth to 10 to restrict tree complexity and prevent overfitting. 'min_samples_split': This parameter sets the minimum sample count needed to divide an internal node during tree-building. In the code, min_samples_split is set to 2, indicating that a node will only be split if it contains at least 2 samples. Setting this parameter too low can lead to overfitting, while setting it too high can result in underfitting. A value of 2 is commonly used as a starting point and can be updated based on the complexity. 'min_samples_leaf': The min_samples_leaf parameter specifies the minimum number of samples required to be at a leaf node. It prevents the tree from splitting nodes that contain too few samples, which can lead to overfitting. In the provided code, min_samples_leaf is set to 1, meaning that each leaf node must contain at least 1 sample. Like min_samples_split, this parameter helps control the tree's complexity and generalisation ability. 'random_state': This parameter sets the random seed for reproducibility. By fixing the random_state to 42, the same sequence of random numbers will be generated each time the model is trained, ensuring consistent results across different runs. This is particularly important for research and development purposes, as it allows for the comparison of results and facilitates debugging.

Overall, these hyperparameters were selected to achieve a balance between the generalisation ability and model complexity, with the objective of constructing a Random Forest classifier that can effectively capture data patterns while preventing overfitting. Adjustments to these parameters can be made based on the specific characteristics of the dataset and the desired trade-offs between bias and variance.

C. Support Vector Machine (SVM)

The Support Vector Machine (SVM) is a widely used supervised learning algorithm that is employed for classification tasks. The dataset is initially split into features (X) and the target variable (y). Subsequently, the data is further partitioned into training and testing sets using an 80-20 split ratio to facilitate model evaluation. The Support Vector Classifier (SVC) is initialised with specified hyperparameters, including Kernel Function: SVMs use a kernel function to map the input data into a higher-dimensional space where the classes are more separable. In the provided code, the radial basis function (RBF) kernel, denoted by ‘rbf’, is used. Support Vectors: Support vectors are the data points that lie closest to the decision boundary (hyperplane) and influence the position and orientation of the hyperplane. These are the critical elements that determine the structure of the decision boundary.

The margin is the distance between the decision boundary and the nearest data point from each class. SVM aims to maximise this margin, as it represents the confidence of the classifier in its predictions and helps in reducing overfitting. The regularisation parameter, denoted by ‘C’, controls the trade-off between maximising the margin and minimising the classification error. In the provided code, ‘C’ is set to 1.0, representing a balanced approach between the two objectives. The gamma parameter determines the influence of each training example in the kernel function. In the code, ‘gamma’ is set to ‘scale’, which automatically calculates gamma as 1 / (n_features * X.var ()), where X is the training data. This setting ensures that the scale of the kernel coefficient is appropriate for the input data. SVMs make predictions by evaluating the decision function, which assigns a class label to new data points based on their position with respect to the decision boundary. In binary classification, the decision function computes the signed distance of a data point to the hyperplane. Positive distances correspond to one class, while negative distances correspond to the other class.

In summary, the SVM model leverages the concept of maximising the margin to separate different classes in the feature space, making it effective for binary classification tasks.

D. XGBoost

The XGBoost is a scalable and efficient implementation of gradient boosting machines. Gradient boosting is an ensemble learning technique that builds multiple decision trees sequentially, where each subsequent tree corrects the errors of its predecessors. The XGBoost classifier is initialised with the objective function set to 'binary: logistic,' indicating binary classification. The evaluation metric is specified as 'logloss,' which measures the logarithmic loss between the predicted probabilities and the actual labels. Logloss provides a metric for the predictive efficacy of a model by penalising erroneous estimates. The learning rate, which is represented by 'eta' is responsible for determining the step size at each iteration while also minimising the loss function. The model is more resistant to overfitting when the learning rate is reduced; however, it necessitates a greater number of iterations to achieve convergence. In this case, 'eta' is set to 0.1, striking a balance between training speed and model performance. The 'max_depth' parameter determines the maximum depth of each decision tree in the ensemble. In order to restrict the complexity of particular trees, the value of 'max_depth' is specified to 6. Subsample Ratio (subsample) and Column Subsampling (colsample_bytree) parameters control the subsampling of training instances and features when constructing each tree, respectively. Subsampling introduces randomness, reducing overfitting and improving model performance. The 'subsample' and 'colsample_bytree' are set to 0.8, indicating that 80% of the training instances and features are sampled during tree construction. To ensure reproducibility, a random seed of 42 is set. This fixes the randomness in the model's training process, resulting in consistent results across different runs.

E. CatBoost

CatBoost classifier is a robust gradient boosting algorithm renowned for its efficiency in handling categorical features. Unlike many other ML models, CatBoost does not require pre-processing steps like one-hot encoding or label encoding, as it automatically handles categorical variables internally. This characteristic significantly simplifies the workflow and reduces the risk of introducing errors during data pre-processing. One of the standout features of CatBoost is its ability to handle categorical data seamlessly. By employing a specialised algorithm for processing categorical variables, CatBoost effectively captures the inherent structures and relationships present in such data types. This capability is particularly advantageous in real-world scenarios where datasets frequently contain a mix of numerical and categorical features, as it eliminates the need for manual feature engineering.

The CatBoost classifier is initialised with default parameters, leveraging its powerful default settings such as the 'Logloss' objective function and accuracy. However, one of the significant advantages of CatBoost lies in its flexibility, as it enables fine-tuning through a range of hyperparameters. These hyperparameters, including 'iterations', 'learning_rate', 'depth', and 'l2_leaf_reg', play pivotal roles in shaping the behaviour and performance of the model. The 'iterations' parameter regulates the quantity of boosting iterations or trees that are constructed within the training process. Increasing the iterations can enhance the model's capacity to capture complex patterns in the data, but it also risks overfitting if not carefully regulated. The 'learning_rate' parameter, often referred to as the step size or shrinkage factor, governs the magnitude of updates made to the model's weights during each iteration. Lower learning rates produce smoother convergence but may require more iterations for optimal performance. The 'depth' parameter determines the maximum depth of each tree in the ensemble. Controlling the tree depth helps strike a balance between complexity and generalisation. The 'l2_leaf_reg' parameter introduces L2 regularisation, also known as weight decay, to penalise large parameter values and prevent overfitting. By adjusting the regularisation strength, users can control the model's tendency to fit noise in the data, thus improving its ability to generalise to unseen examples.

These hyperparameters collectively offer users granular control over the model's complexity, regularisation, and learning dynamics, empowering them to fine-tune the algorithm's behaviour to suit specific dataset characteristics and optimisation objectives. This flexibility, combined with CatBoost's inherent ability to handle categorical features seamlessly, makes it a preferred choice for a wide range of ML tasks, from classification to regression and ranking problems.

F. Logistic Regression (LR)

The Logistic Regression (LR) model employed in the provided code is a fundamental, yet powerful classification algorithm widely used in various ML applications. Logistic regression is particularly suited for binary classification tasks, where the goal is to predict the probability of an observation belonging to a particular class.

The LR classifier is initialised with specific hyperparameters tailored to its optimisation process. The choice of solver in logistic regression plays a crucial role in optimising the model's parameters to fit the training data effectively. In this implementation, the 'liblinear' solver is utilised due to its efficiency and versatility, particularly suited for small to medium-sized datasets. The 'liblinear' solver is well-known for its ability to handle both L1 and L2 regularisation penalties, providing flexibility in controlling model complexity and addressing multicollinearity issues in the dataset. The 'L1' regularisation penalty, also known as Lasso regularisation, encourages sparsity in the model by penalising the absolute magnitude of the coefficients. It effectively performs feature selection by shrinking less important features' coefficients to zero, thus simplifying the model and improving its interpretability. On the other hand, the 'L2' regularisation penalty, also called Ridge regularisation, penalises the squared magnitude of the coefficients, mitigating their excessive growth and diminishing the model's susceptibility to data outliers. Furthermore, fixing the 'random_state' parameter to 42 ensures reproducibility across different runs of the model. By setting a specific random seed, the randomness involved in the model's initialisation, including the shuffling of data during training and the random initialisation of coefficients, is controlled. This reproducibility is crucial for debugging, validation, and comparison purposes, as it guarantees consistent results and facilitates the identification of any discrepancies or inconsistencies in model's behaviour.

In summary, the LR model offers a robust and interpretable approach to binary classification tasks, providing insights into the relationship between input features and the target variable while delivering competitive performance in terms of predictive power.

To evaluate the model’s performance, 5-fold cross-validation is conducted, generating strong estimates of the area under the Receiver Operating Characteristic (ROC) curve (AUC-ROC) across different data subsets. After cross-validation, the model is trained on the training set, and predictions are made on the test set. The probabilities for the positive class are computed, and the AUC-ROC score is calculated. An AUC value of 0.5 indicates random guessing and 1.0 indicates perfect discrimination on the ROC curve, which compares true positive rate (sensitivity) to false positive rate (1-specificity). Furthermore, performance metrics are computed to evaluate the model's predictive accuracy and balance between precision and recall. A confusion matrix is generated to summarise the model's performance in terms of true positive, true negative, false positive, and false negative predictions. Additionally, a classification report provides detailed metrics for each class, including precision, recall, and F1-score.

Algorithm 4-1: Novel hybrid stacking ML Model

```
Procedure StackingClassifierTraining (X, y, test_size, random_state)
          Pre-process (X, y)
          Split dataset into (X_train, X_test, y_train, y_test)
          Normalize (X_train, X_test)
          Model ← Create stacking classifier with RandomForestClassifier
                  as meta-classifier()
                  ('Random Forest', rf_classifier)
                  ('SVM', svm_classifier)
                  ('Logistic Regression', lr_classifier)
                  ('Decision Tree', dt_classifier)
                  ('XGBoost', xgb_classifier)
                  ('CatBoost', catboost_classifier)
                Cross-validation: StratifiedKFold with cv_folds
 for k ← 0 to n-1 do
            Train StackingClassifier (X_train, y_train)
            Predict y_pred on (X_test)
            Evaluate Model:
                Accuracy ← accuracy_score(y_test, y_pred)
                Precision ← precision_score(y_test, y_pred)
                Recall ← recall_score(y_test, y_pred)
                F1 Score ← f1_score(y_test, y_pred)
                ROC AUC Score ← roc_auc_score(y_test, y_pred)
       Print Evaluation Metrics
end for
End Procedure
```

4.4 Experimental result

Initially, we assess the performance of ML algorithms individually and in the absence of employing ensemble techniques. The outcomes derived from this evaluation are comprehensively depicted in Table 4.3, illustrating a comparative analysis of the algorithms. Figure 4.15 shows the performance of ML algorithms without ensemble techniques. Figures 4.16,4.17,4.18,4.19 ,4.20, and 4.21 show Confusion Matrix for ML techniques.

Table 4.3: *Algorithms performance.*

ML technique	Accuracy	Precision	Recall	F1 score	AUC
DT	99.93%	89.89%	81.63%	85.56%	90.80%
RF	99.96%	97.40%	76.53%	85.71%	97.25%
SVM	99.94%	97.02%	66.33%	78.79%	95.13%
XGBoost	99.95%	95.00%	77.55%	85.39%	97.83%
CatBoost	99.96%	97.44%	77.55%	86.36%	98.37%
LR	99.92%	88.06%	60.20%	71.52%	97.01%

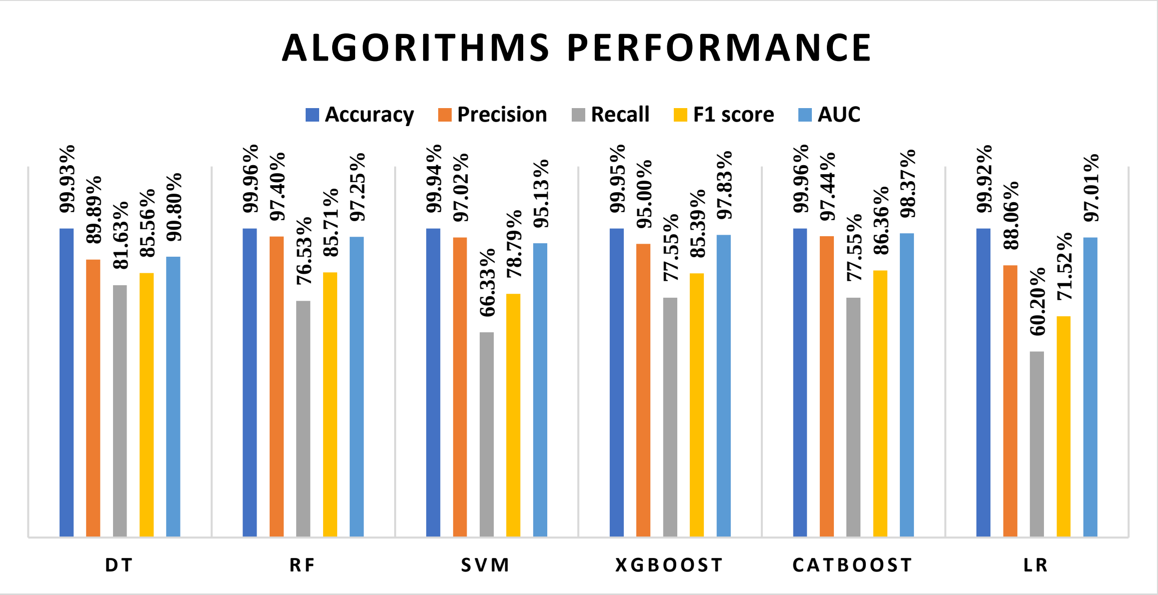


Figure 4.15 *Algorithms performance before applying ensemble techniques.*

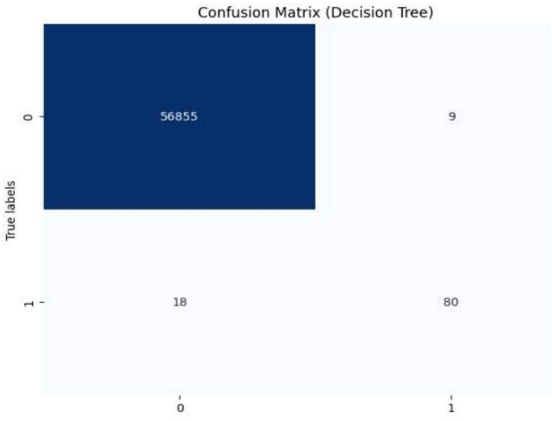


Figure 4.16 *Confusion matrix- DT*

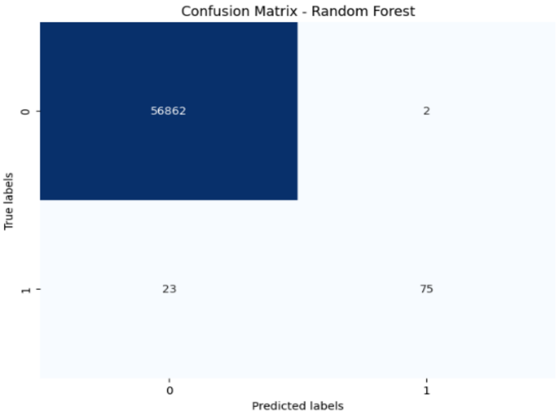


Figure 4.17 *Confusion matrix- RF*

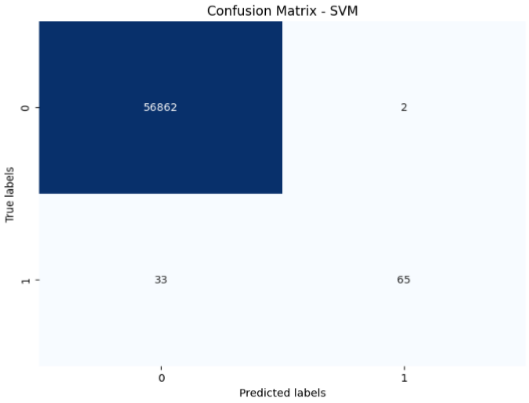


Figure 4.18 *Confusion matrix -SVM*

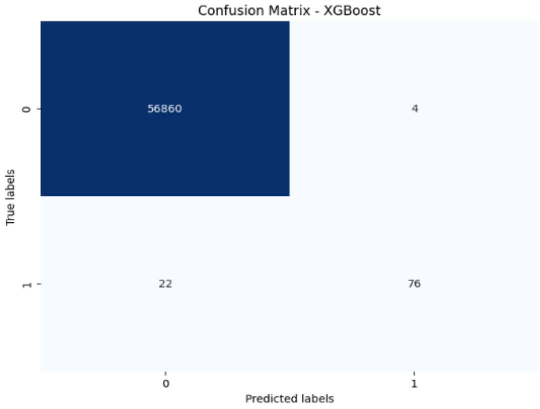


Figure 4.19 *Confusion matrix -XGBoost*

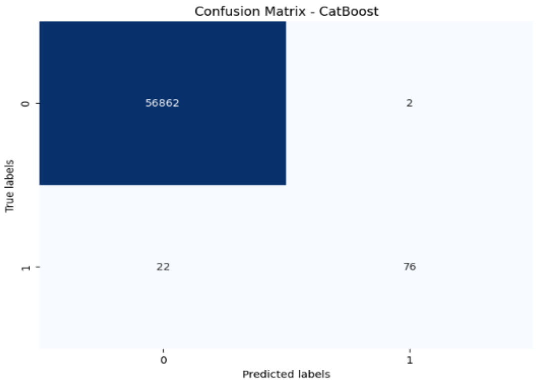


Figure 4.20 *Confusion matrix -CatBoost*

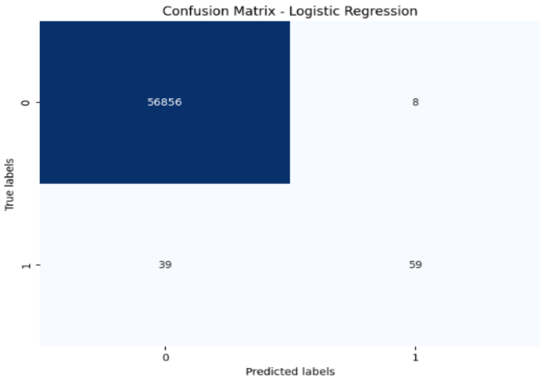


Figure 4.21 *Confusion matrix -LR*

4.4.1 Comparison between ML Techniques before ensemble

The DT model's capacity to effectively categorise most occurrences while sustaining a low false positive rate is shown by its elevated accuracy and precision (99.93% and 89.89%, respectively). This suggests that the model is proficient in distinguishing between positive and negative instances, thereby reducing the potential of misclassifications. However, the model's relatively low recall (81.63%) suggests that it may encounter difficulty in identifying all positive instances, potentially leading to false negatives. In other words, there might be instances of the positive class that the model fails to capture. This limitation is further reflected in the F1 score (85.56%). Despite these drawbacks, the model's moderate AUC score (90.80%) suggests that it still exhibits reasonable discriminative ability between positive and negative instances. Overall, while the Decision Tree model demonstrates strong classification performance in terms of accuracy and precision, There is potential for enhancement in its capacity to accurately record all positive instances without sacrificing precision.

The RF model demonstrates exceptional accuracy and precision, achieving 99.96% and 97.40%, respectively. This high precision score indicates a low false positive rate. However, the relatively lower recall of 76.53% suggests that the model may miss some positive instances. Despite this, the model's F1 score remains decent at 85.71%, striking a balance between precision and recall. Moreover, the high AUC score of 97.25% signifies excellent discriminative ability, Additional evidence supporting the model's exceptional ability to differentiate between positive and negative instances. Overall, while the RF model excels in accuracy, precision, and AUC score, there is potential for enhancement in recall to capture more positive instances effectively.

The SVM model showcases commendable accuracy and precision, achieving 99.94% and 97.02%, respectively, which reflects its ability to correctly classify the majority of instances and maintain a low false positive rate. However, the recall is relatively lower at 66.33%, indicating that it may overlook some positive instances, leading to false negatives. Despite this, the F1 score and AUC, measuring 78.79% and 95.13%, respectively, demonstrate moderate performance overall. Additionally, the AUC score suggests that the model exhibits reasonable discriminative ability, although further optimisation could lead to better performance in distinguishing between positive and negative instances. Overall, while the SVM model shows strengths in accuracy and precision, there are opportunities for enhancement to achieve better recall and overall performance.

XGBoost demonstrates exceptional performance across various metrics, boasting a high accuracy of 99.95% and precision of 95.00%. This suggests that the model effectively classifies most instances while maintaining a low false positive rate, indicative of its robustness in prediction. Notably, XGBoost achieves a balanced trade-off between recall and precision, signifying its ability to effectively identify positive instances without significantly compromising precision. The F1 score is also high at 85.39%, highlighting the model's capability in achieving harmonious performance across these key metrics. Additionally, the AUC score of 97.83% reflects the model's excellent discriminative ability, indicating its effectiveness in distinguishing between positive and negative instances. Overall, XGBoost demonstrates outstanding performance across multiple evaluation metrics, making it a powerful choice for classification tasks, particularly when a balance between precision, recall, and discriminative ability is essential.

CatBoost stands out with exceptional performance across multiple evaluation metrics, showcasing its effectiveness in classification tasks. With a remarkable accuracy of 99.96%, CatBoost ensures accurate predictions on most instances, reflecting its ability to correctly classify data points. Additionally, the model achieves an impressive precision of 97.46%, indicating a low false positive rate and high confidence in its positive predictions. This suggests that when CatBoost predicts a positive instance, it is highly likely to be correct. Furthermore, the model demonstrates a commendable recall of 77.55%, signifying its ability to capture a significant portion of positive instances in the dataset. One of CatBoost's notable strengths lies in its balanced performance across various metrics, as evidenced by its high F1 score of 86.35%. The F1 score is an appropriate metric for evaluating the overall efficacy of a model in binary classification tasks, as it encompasses both precision and recall values. Moreover, CatBoost exhibits strong discriminative ability, as evidenced by its high AUC score of 98.27%. The model's ability to differentiate between positive and negative instances is indicated by the AUC score, with higher values indicating superior performance in this regard. Overall, CatBoost demonstrates outstanding performance across accuracy, precision, recall, F1 score, and AUC, making it a compelling choice for classification tasks where high predictive performance and strong discriminative ability are essential. Its ability to maintain high precision while capturing a substantial portion of positive instances underscores its utility in real-world applications where precision and recall are equally important.

The LR model achieves commendable accuracy, indicating its capability to correctly classify the majority of instances, with an accuracy of 99.92%. However, upon closer inspection, its precision and recall metrics reveal areas for improvement. While the model achieves a precision of 88.06%, indicating a relatively low false positive rate, its recall is notably lower at 60.20%. This suggests that the model may struggle to capture a significant portion of positive instances, leading to a higher number of false negatives. Consequently, there is a risk of misclassifying positive instances as negative. The F1 score is moderate at 71.52%, indicating that the model's performance is slightly compromised due to the imbalance between precision and recall. While the model's AUC score is relatively high at 97.01%, indicating good discriminative ability, the moderate F1 score suggests that there is room for enhancement in correctly recognising positive cases. Overall, the Logistic Regression model demonstrates high accuracy but falls short in terms of precision and recall, particularly in correctly identifying positive instances. While its AUC score suggests reasonable discriminative ability, the model's performing might be improved by addressing the imbalance between precision and recall, potentially through feature engineering or parameter tuning.

Overall, all models exhibit high accuracy, but their performance varies in terms of precision, recall, F1 score, and AUC. CatBoost and XGBoost generally perform the best across all metrics, followed closely by Random Forest. Decision Tree and SVM show slightly lower performance, while Logistic Regression performs the poorest in terms of recall and precision.

When comparing models, it is advisable to prioritise the F1-score due to the dataset's imbalance, while avoiding the use of accuracy and AUC for several reasons. The accuracy metric evaluates the number of correctly predicted records. However, in an imbalanced dataset, even if all records are predicted as the majority class (0 in this case), the accuracy score could still surpass 99%. On the other hand, AUC treats positive and negative cases equally, making it suitable for balanced datasets but less effective for imbalanced ones. In contrast, the F1-Score offers a more suitable alternative. It represents the harmonic mean of precision and recall, assigning greater weight to lower values. Consequently, a high F1-score is only achieved when both precision and recall are high, making it more sensitive to true positive cases. Given that positive instances are rare in imbalanced datasets, F1-Score becomes a more reliable metric for comparing multiple models, ensuring a balanced evaluation that prioritises the correct identification of positive cases.

In assessing the effectiveness of the ML models utilised, we will prioritise the F1-Score as our primary evaluation metric. The F1-Score provides a balanced measure by considering both the precision and recall of the models. It is particularly useful in scenarios where there is an imbalance in the class distribution, as it gives equal weight to both false positives and false negatives. By focusing on the harmonic mean of precision and recall, the F1-Score emphasises the model's ability to accurately identify positive instances while minimising false positives. This makes it a suitable metric for comparing multiple models, especially in cases where positive cases are rare, as it provides a comprehensive assessment of a model's performance across different aspects of classification accuracy.

When comparing the performance of various ML algorithms based on their F1 Scores, CatBoost emerges as the top performer with the highest F1 Score of 87.01%. This indicates that CatBoost achieves an excellent balance between precision and recall, effectively distinguishing between positive and negative instances while minimising both false positives and false negatives. Following closely behind, Random Forest (RF) demonstrates a commendable F1 Score of 85.71%, showcasing its ability to accurately classify instances while maintaining a low rate of misclassification. Similarly, XGBoost exhibits a robust F1 Score of 85.39%, reflecting its capacity to achieve a balance between precision and recall in classification tasks. However, the Decision Tree (DT) algorithm, while still performing well with an F1 Score of 85.56%, shows slight room for improvement in accurately identifying positive instances. Conversely, Logistic Regression (LR) and Support Vector Machine (SVM) models lag with lower F1 Scores of 71.52% and 78.79%, respectively, suggesting potential challenges in effectively classifying positive instances. In summary, CatBoost and RF stand out as top performers in terms of F1 Score, highlighting their strong performance in classification tasks. Figure 4.22 shows the F1 score for ML techniques.

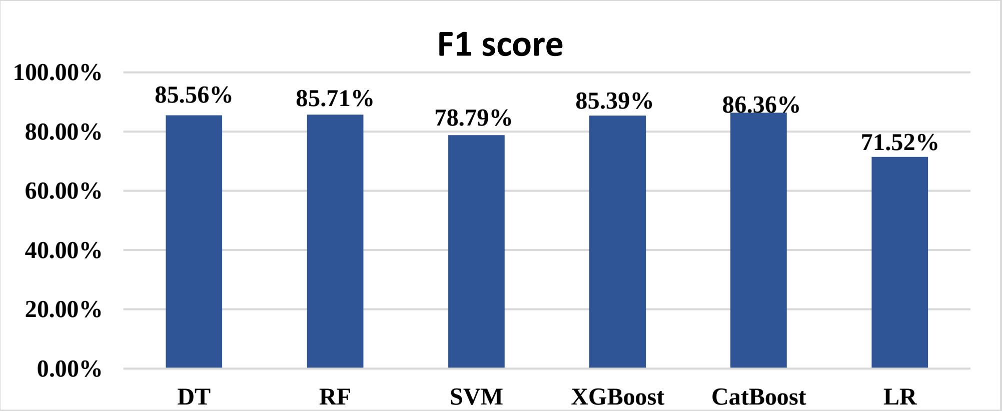


Figure 4.22 *F1 score for ML techniques.*

4.4.2 Ensemble Techniques

Ensemble techniques in ML are essential strategies employed to enhance the predictive power of models by leveraging strengths of multiple individual models. These methods stem from the understanding that while individual models may have limitations or biases, combining their predictions can lead to more accurate and robust outcomes. By aggregating the predictions of diverse models, ensemble methods capitalise on the collective intelligence of the ensemble to produce results that are often superior to those of any single model operating in isolation. The concept underlying ensemble methods is akin to the "wisdom of the crowd," where the combined judgment of a group of individuals tends to be more accurate than that of any single member of the group. This implies that the combined conclusion of multiple models is typically more reliable than the conclusion of any single model in the domain of machine learning. Ensemble methods achieve this by considering different perspectives, learning patterns, and capturing various aspects of the data, leading to a more comprehensive understanding of the underlying relationships.

Ensemble methods find applications across a wide range of tasks in both classification and regression domains. In classification tasks, ensembles can effectively handle complex decision boundaries, improve the classification of minority classes, and mitigate the impact of noisy or conflicting data points. In regression tasks, ensembles can provide more robust predictions by reducing the influence of outliers and improving the generalisation performance of the models. Ensemble techniques provide a number of benefits, one of the most important of which being their flexibility and suppleness to a variety of datasets and issue scenarios. They can accommodate various types of base models, ranging from simple ones like decision trees to more complex ones like neural networks. Ensemble techniques can also be tailored to address specific challenges such as overfitting, bias, or imbalanced data distributions. Moreover, ensemble methods offer a practical approach to model combination, allowing for seamless integration of multiple models into a cohesive framework. This integration can take different forms, including averaging predictions, combining probabilities, or training meta-learners to learn optimal combinations of base model outputs. Overall, ensemble techniques represent a powerful arsenal in the ML toolkit, enabling practitioners to harness the collective intelligence of diverse models for improved predictive performance. As ML continues to advance and datasets grow in complexity, ensemble methods are likely to remain indispensable for achieving state-of-the-art results across various applications.

1- Stacking is a more sophisticated ensemble technique that employs a meta-learner to integrate predictions from numerous base models. Instead of simply averaging or voting, stacking trains a meta-learner on the predictions of base models to learn how to best combine their outputs. This approach can capture complex relationships between models and potentially outperform individual models. However, stacking requires more computational resources and data for training compared to other ensemble methods.

2- Bagging entails the independent training of multiple base learners on random subsets of the training data, which are typically sampled with replacement. Each base learner learns from a different perspective of the data, and their predictions are then combined through averaging (for regression) or voting (for classification). Bagging helps to reduce variance and overfitting, particularly in high-variance models such as decision trees.

3- Voting is a simpler ensemble method that combines predictions from multiple models using a straightforward strategy such as majority voting (for classification) or averaging (for regression). It is easy to implement and works well when the base models are diverse and complementary. Voting can be further categorised into hard voting, where the final prediction is based on the majority vote of the base models, and soft voting, where predictions are weighted based on confidence scores.

4- The Random Subspace involves randomly selecting subsets of features (columns) from the original dataset for training each base model. Instead of sampling instances (rows) as in traditional bagging, this method focuses on sampling features, effectively creating diverse subsets of the feature space. The Random Subspace encourages each model to specialise in learning from distinct aspects of the data. During prediction, each base model operates within its randomly selected feature subspace, and the ensemble combines their predictions to make the final decision. This method is effective in reducing overfitting, improving generalisation, and increasing model diversity.

The Stacking ensemble method is tailored for classification tasks, aiming to boost predictive performance by amalgamating predictions from multiple base classifiers through a meta-classifier. This technique, renowned for its potency in leveraging the diverse strengths of individual models, strives to achieve more accurate and robust predictions. The research utilised a diverse set of base classifiers, including RF, SVM, LR, DT , XGBoost, and CatBoost.

The selection of these classifiers was based on their proved efficacy in managing a variety of data types and the complexity of their modelling. The dataset was pre-processed by splitting it into features (X) and the target variable (y) using selected features. Subsequently, the data was split into training and testing sets with a ratio of 80:20, ensuring that the model was trained on a sufficient amount of data while still having a separate set for evaluation. Standardisation was applied to normalise the data, which is a common pre-processing step recommended for many ML algorithms to ensure that all features contribute equally to the model's performance. Moreover, a stratified k-fold cross-validation approach with 5 splits was employed to ensure robust model evaluation and prevent overfitting, thus enhancing the generalisation ability of the model.

The base classifiers were configured with various hyperparameters to capture different aspects of the data and improve the diversity of the ensemble. Random Forest was chosen as the final meta-classifier for the stacking ensemble due to its proven effectiveness in handling complex datasets and mitigating overfitting. To achieve superior performance, the stacking ensemble capitalises on the strengths of individual models by combining the predictions of multiple base classifiers. After fitting the stacking classifier on the training data, predictions were made on the test set. To evaluate the efficacy of the stacking model in identifying fraudulent transactions, performance evaluation metrics were calculated, including accuracy, precision, recall, F1 score, and ROC AUC score. These metrics provide a comprehensive assessment of the model's predictive performance.

The results revealed that the stacking ensemble outperformed individual base classifiers in terms of overall predictive performance. The stacking ensemble effectively classified both fraudulent and non-fraudulent transactions, as evidenced by its elevated accuracy, precision, recall, and F1 score. Additionally, the ROC AUC score suggested that the model exhibited strong discriminatory power between the two classes, further confirming its effectiveness in cyber fraud detection. Additionally, the confusion matrix offered comprehensive analysis of the model's performance across various classes. The classification report offered a comprehensive summary of precision, recall, and F1 score, providing valuable information for model evaluation and comparison.

The findings provide evidence that stacking is an excellent strategy for ensemble learning, showing the potential of this technique to improve the resilience and reliability of predictive models in applications that are used in the real world. Further research could explore additional ensemble configurations and feature engineering techniques to further improve model performance and address the evolving challenges of cyber fraud detection in financial transactions. Table 4.4 shows the results after applying Stacking ensemble method. Figure 4.23 shows the F1 score for ML techniques after Stacking ensemble.

Table 4.4: *Results after applying stacking ensemble.*

ML technique	Accuracy	Precision	Recall	F1 score	AUC
DT	99.93%	89.89%	81.63%	85.56%	90.80%
RF	99.96%	97.40%	76.53%	85.71%	97.25%
SVM	99.94%	97.02%	66.33%	78.79%	95.13%
XGBoost	99.95%	95.00%	77.55%	85.39%	97.83%
CatBoost	99.96%	97.44%	77.55%	86.36%	98.37%
LR	99.92%	88.06%	60.20%	71.52%	97.01%
Stacking Hybrid ML model	99.96%	98.73%	79.59%	88.14%	89.80%

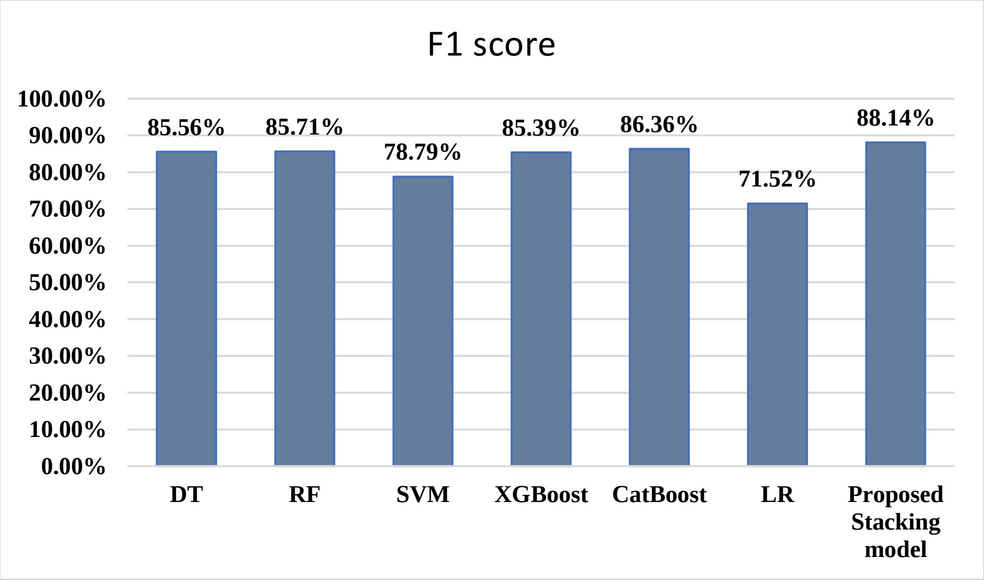


Figure 4.23 *F1 score for ML techniques after Stacking ensemble.*

The results presented in the table offer a nuanced comparison between the Stacking Hybrid ML model and other individual ML algorithms in the context of credit card transaction cyber fraud detection, with a particular emphasis on the F1 score. The F1 score provides insight into a model's overall effectiveness in identifying fraudulent transactions accurately while minimising false positives.

Comparing the F1 scores of the Stacking Hybrid ML model with those of individual algorithms reveals a notable disparity. While algorithms like RF, SVM, LR, XGBoost, and

CatBoost demonstrate respectable F1 scores, they ultimately fall short of matching the performance achieved by the Stacking ensemble model. In particular, the Stacking ensemble model exhibits a significantly higher F1 score compared to all individual algorithms, underscoring the effectiveness of the ensemble approach in bolstering predictive accuracy and robustness. For instance, the Stacking ensemble model achieved an F1 score of 88.14%, surpassing the F1 scores of individual algorithms such as DT (85.56%), RF (85.71%), SVM (78.79%), LR (71.52%), XGBoost (85.39%), and CatBoost (86.36%). This substantial difference highlights the superiority of the Stacking ensemble model in achieving a balanced trade-off between precision and recall, thereby enhancing its capability in accurately detecting fraudulent transactions.

The choice of Stacking as the ensemble technique for the novel Hybrid model stems from its inherent advantages over other ensemble methods. Stacking allows for the combination of diverse base classifiers, each capturing unique aspects of the data, leading to a more comprehensive understanding of the underlying patterns. By leveraging the collective intelligence of multiple models, Stacking synthesises a robust predictive model that is less prone to individual model biases and overfitting. Moreover, Stacking fosters a collaborative synergy among constituent algorithms, enabling them to complement each other's strengths and mitigate weaknesses. This collaborative nature empowers the ensemble model to navigate the complex landscape of credit card transaction cyber fraud detection with heightened precision and accuracy. Additionally, Stacking facilitates continuous learning and adaptation, allowing the model to evolve in response to emerging cyber fraud patterns and evolving threats, thereby enhancing its resilience in real-world scenarios. In summary, the empirical findings underscore the pivotal role of the Stacking Hybrid ML model in advancing the detection of cyber fraud. Through its superior performance, as evidenced by the higher F1 score compared to all individual algorithms, the ensemble approach offers a potent tool for enhancing predictive accuracy and resilience in cyber fraud detection systems, thereby addressing the evolving challenges of financial cyber fraud detection.

4.4.3 Comparison with other ensemble techniques

In addition to the Stacking Hybrid ML model, this research delves into the exploration of several other ensemble techniques, aiming to thoroughly evaluate their comparative performance. These alternative methodologies comprise a diverse array, including the application of Voting, Subspace Random ensemble, and bagging ensemble approaches. Each of these techniques offers unique strategies for combining multiple base classifiers to enhance predictive accuracy. The empirical findings derived from these experiments are systematically documented in Table 4.5, thereby facilitating a comprehensive examination and analysis within the intricate landscape of credit card transaction cyber fraud detection. Figure 4.24 shows comparison with other ensemble techniques.

Table 4.5: *Algorithms performance on different ensemble techniques.*

ML technique	Accuracy	Precision	Recall	F1 score	AUC
Stacking ML model	99.96%	98.73%	79.59%	88.14%	89.80%
Voting ensemble	99.95%	97.37%	75.51%	85.06%	97.45%
Subspace Random ensemble	99.95%	95.95%	72.45%	82.56%	86.22%
Bagging ensemble	99.95%	97.30%	73.47%	83.72%	86.73%

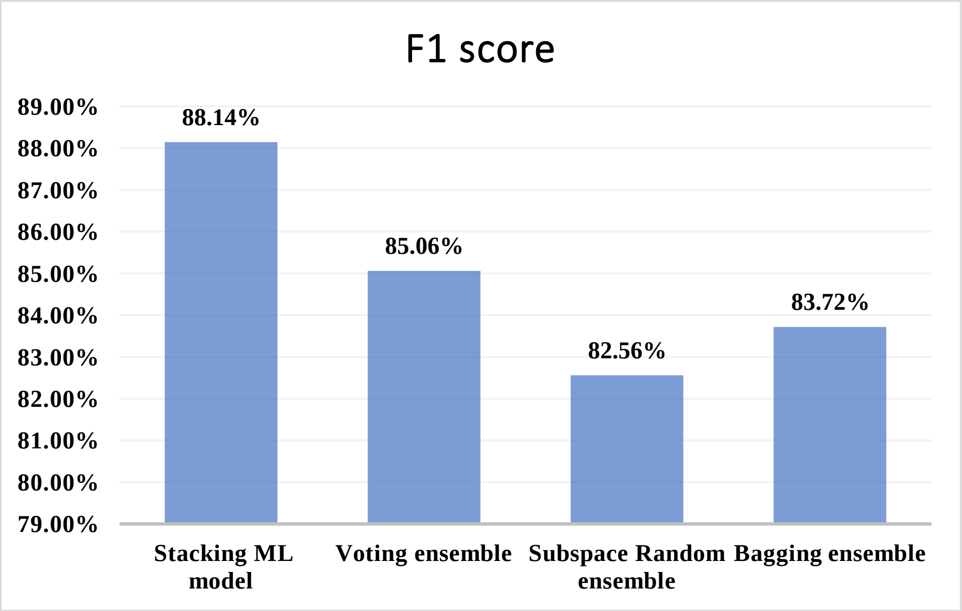


Figure 4.24 *Comparison with other ensemble techniques.*

When comparing the ensemble techniques based on F1 score, the Stacking Hybrid ML model emerges as the top performer. With an F1 score of 88.14%, the Stacking model achieves the highest balance between minimising false positives and false negatives among all the techniques evaluated. This indicates its effectiveness in accurately classifying both fraudulent and non-fraudulent transactions while maintaining a high level of precision and recall simultaneously.

Following the Stacking model, the Voting ensemble demonstrates a respectable F1 score of 85.06%. However, it falls short of matching the performance of the Stacking model, suggesting that the Stacking approach leverages the diversity of base classifiers more effectively to achieve better overall predictive performance. The Subspace Random ensemble and Bagging ensemble techniques exhibit F1 scores of 82.56% and 83.72%, respectively. While these scores indicate reasonable performance, they are notably lower compared to both the Stacking model and the Voting ensemble. This suggests that the Stacking and Voting approaches may offer superior predictive power and robustness in capturing fraudulent transactions compared to the Subspace Random and Bagging ensembles.

In summary, based on F1 score comparison, The Stacking Hybrid ML model is the most effective technique for detecting credit card cyber fraud among the ensemble methods that have been evaluated, with the Voting ensemble following closely behind. These findings underscore the significance of leveraging ensemble techniques, particularly stacking, to improve the efficacy and accuracy of cyber fraud detection systems.

4.5 Chapter summary

The chapter elaborated on the successful new hybrid design of an advanced ML model tailored specifically for credit card transaction cyber fraud detection. It utilised a diverse set of ML algorithms, including RF, SVM, LR, DT, XGBoost, and CatBoost. These algorithms were selected based on their diverse nature and proven effectiveness in handling various types of data and modelling complexities. The chapter meticulously detailed the intricacies of data pre-processing and feature engineering, including data cleaning, encoding categorical variables, handling missing values, and conducting exploratory data analysis (EDA). Insights gleaned from EDA informed critical decisions regarding feature scaling, ensuring that numerical features were standardised to optimise model performance. This comprehensive pre-processing pipeline laid the foundation for accurate and reliable cyber fraud detection models. The subsequent section of the chapter delved into model selection and evaluation, subjecting six distinct ML algorithms to rigorous assessment across various performance metrics.

CatBoost emerged as the top performer with the highest F1 Score of 86.36%, effectively distinguishing between positive and negative instances while minimising false positives and false negatives. Following closely behind, Random Forest demonstrated a commendable F1 Score of 85.71%, showcasing its ability to accurately classify instances while maintaining a low rate of misclassification. The chapter emphasised the significance of prioritising the F1 score, particularly in datasets characterised by class imbalances, for a nuanced comparison of model effectiveness.

The Stacking Hybrid ML model emerged as particularly effective for combining predictions from multiple base classifiers using a meta-classifier, resulting in superior predictive performance. When comparing with individual ML performance, the stacking ensemble model exhibited a significantly higher F1 score compared to all individual algorithms, With an F1 score of 88.14%, Among all the techniques that have been assessed, the Stacking model obtains the most optimal balance between minimising false positives and false negatives. This indicates its effectiveness in accurately classifying both fraudulent and non-fraudulent transactions while maintaining a high level of precision and recall simultaneously, underscoring its effectiveness in bolstering predictive accuracy and robustness. The choice of stacking as the ensemble technique stemmed from its inherent advantages over other ensemble methods.

Experimental findings underscored the stacking ensemble model's capability to outperform other ensemble techniques, showcasing an F1 score of 88.14%. The Voting ensemble demonstrated respectable performance but fell short of matching the stacking model's performance, achieving an F1 score of 85.06%. Subspace Random and Bagging ensemble techniques exhibited reasonable performance but were notably lower compared to stacking and voting approaches, with F1 scores of 82.56% and 83.72%, respectively.

In summary, Among the individual algorithms and ensemble methods that were assessed, the stacking hybrid ML model was the most effective method for detecting credit card transaction cyber fraud. The results emphasised the need of using ensemble approaches, especially stacking, to enhance the efficiency and accuracy of cyber fraud detection systems.

CHAPTER 5: NOVEL HYBRID CNN-BILSTM MODEL

Although, the effectiveness of the Stacking Hybrid ML model in Chapter 4 was outstanding, In recent years, there has been an increasing emphasis on the examination of DL techniques. By employing DL, one can achieve enhanced precision and optimised performance. The application of DL techniques enables the system to respond flexibly to complex data patterns and to identify new fraudulent patterns. To detect credit card cyber fraud more effectively, therefore, it is recommended that additional research be conducted on DL techniques. In addition, considering the limitations of ML techniques, it is prudent to contemplate the DL algorithms to achieve encouraging detection outcomes. Several scholarly articles have proposed DL methods to advance the accuracy from unbalanced datasets. Researchers are also encouraged to employ both undersampling and oversampling methods due to the exceedingly skewed nature of the datasets. The novel hybrid DL model, developed in this chapter, consists of developing DL techniques CNN-BiLSTM. Figure 5.1 shows the novel hybrid CNN-BiLSTM model.

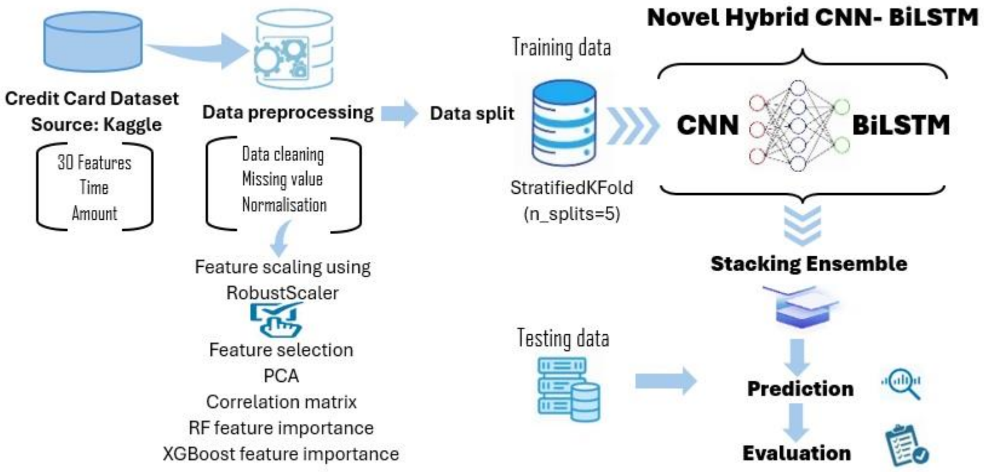


Figure 5.1 *Novel Hybrid CNN-BiLSTM model.*

The novel hybrid DL model was trained and tested in real-world dataset. The research work's novelty and contributions were as follows:

1. A new approach including DL techniques to classify transaction to fraud and nonfraud.
2. The whole novel hybrid DL model framework as known Hybrid CNN-BiLSTM was trained and tested in a real-world dataset and the obtained results indicates the novel hybrid CNN-BiLSTM model has high performance.

The following is a concise explanation of the novel hybrid DL model classifier.

5.1 CNN model

The simplified CNN model developed to efficiently classify transactions as either fraudulent or non-fraudulent based on selected features. This model leverages the power of convolutional layers to capture complex patterns in the data while maintaining a streamlined architecture to enhance performance and reduce computational overhead.

The CNN model starts with an input layer that receives the normalised feature set. Normalisation is crucial for ensuring that every feature has an equal contribution to the model, hence enhancing the convergence and stability of the training process. The input features are reshaped to be compatible with the Conv1D layers, setting the stage for the convolutional operations.

The first layer in the CNN architecture is a Conv1D layer with 32 filters and a kernel size of 3. This layer uses the ReLU (Rectified Linear Unit) activation function, which introduces non-linearity into the model and helps in learning complex patterns by allowing the network to perform non-linear transformations. The convolution operation in this layer scans the input sequence to detect local patterns, such as correlations between features that may indicate fraudulent behaviour. Following this, a MaxPooling1D layer with a pool size of 2 is used to reduce the spatial dimensions of the output. This pooling operation helps in down-sampling the feature maps, reducing parameters and computational load while preserving the most critical features detected by the convolutional layer.

The second convolutional block consists of another Conv1D layer, this time with 64 filters and a kernel size of 3, again using the ReLU activation function. By increasing the number of filters, this layer can capture more detailed and diverse patterns in the data. Another MaxPooling1D layer follows this convolutional layer, further reducing the dimensionality and focusing on the most prominent features.

A Flatten layer is used to prepare the data for the fully linked layers, converting the 3D tensor output from the previous layer into a 1D vector. Flattening process is crucial for transitioning from the convolutional to the dense part. First dense layer consists of 64 neurons and utilises the ReLU activation function. This layer serves to combine the features extracted by the convolutional layers and learns higher-level representations of the data. A Dropout layer with a dropout rate of 0.5 is included after the Dense layer to prevent overfitting. During the training process, Dropout determines a random percentage of the input units to be zero, which helps the model generalise better by preventing it from becoming too reliant on any specific set of features. Dense layers with solitary neurons and sigmoid activation functions are the model's final layers. This layer generates a probability score ranging from 0 to 1, which denotes the probability of a fraudulent transaction. The sigmoid function is particularly suited for binary classification tasks like this, as it maps the input to a range between 0 and 1, making it easy to interpret the output as a probability. Table 5.1 shows the CNN model structure.

Table 5.1 *CNN model structure.*

Layer (Type)	Output Shape	Param #
conv1d_56 (Conv1D)	(None, 15, 32)	128
max_pooling1d_56 (MaxPooling1D)	(None, 7, 32)	0
conv1d_57 (Conv1D)	(None, 5, 64)	6208
max_pooling1d_57 (MaxPooling1D)	(None, 2, 64)	0
flatten_28 (Flatten)	(None, 128)	0
dense_56 (Dense)	(None, 64)	8256
dropout_28 (Dropout)	(None, 64)	0
dense_57 (Dense)	(None, 1)	65
Total params: 14,657 Trainable params: 14,657 Non-trainable params: 0		

The choice of CNN layers in this model is motivated by their ability to efficiently capture and process spatial hierarchies in the data. Convolutional layers are adept at detecting local patterns and correlations, which are crucial for identifying fraudulent transaction patterns that might not be evident from a simple linear model. The use of MaxPooling layers ensures that the most important features are retained while reducing the computational complexity, making the model more efficient. The combination of convolutional and dense layers allows the model to first extract and then interpret complex features, leading to more accurate classification. By leveraging these CNN layers, the model balances complexity and performance, making it both powerful in detecting fraudulent transactions and practical for deployment in real-time systems. This streamlined architecture ensures that the model remains computationally efficient while providing high accuracy and robustness in cyber fraud detection.

The dataset is first pre-processed by selecting relevant features and splitting it into training and testing sets. The StandardScaler is then employed to normalise the features, ensuring that they have a mean of 0 and a standard deviation of 1. The normalised data is reshaped to meet the input requirements of the Conv1D layer, which converts each sample into a 3D array with dimensions that are appropriate for the CNN.

The model is compiled using the Adam optimiser, binary cross-entropy loss, and accuracy as a metric. Adam (Adaptive Moment Estimation) is chosen for its efficient and effective training capabilities, leading to faster convergence and better performance compared to traditional stochastic gradient descent (SGD). The model is then wrapped in a KerasClassifier for integration with scikit-learn's cross-validation functions. The training process involves using Stratified K-Fold cross-validation with 5 splits to ensure that the model is assessed robustly on various subsets. Stratified K-Fold maintains the class distribution in each fold, providing a reliable estimate of the model's performance and reducing the risk of overfitting to a particular subset of the data.

During training, the model's performance is monitored on both the training and validation sets using the accuracy metric. The learning curves for accuracy and loss are plotted to visualise the model's learning process over 20 epochs, with a ReduceLROnPlateau callback applied to adjust the learning rate dynamically. This callback helps the model to converge more smoothly and avoid getting stuck in local minima by reducing the learning rate when the performance on the validation set plateaus. Following the training phase, the model is assessed on the test set. Predictions are made, and several performance metrics are calculated. Overall, the CNN model's structured approach in training and evaluation ensures a robust and effective system for detecting fraudulent credit card transactions. By leveraging techniques such as Adam optimisation, cross-validation, and various performance metrics, the model is fine-tuned to provide accurate and reliable predictions, resolving unbalanced data and feature extraction and pattern recognition issues.

5.2 BiLSTM model

The simplified Bidirectional Long Short-Term Memory (BiLSTM) model is developed to accurately classify credit card transactions as either fraudulent or non-fraudulent. This model leverages the temporal dependencies and sequence learning capabilities of LSTM networks, augmented by the bidirectional architecture to capture patterns from both past and future contexts within the sequence data.

The model starts with pre-processing steps that include splitting the dataset into features (X) and target labels (y), followed by a train-test split to create training and testing datasets. The features are then normalised using StandardScaler. This normalisation step is crucial for improving the convergence speed and stability of the model during training. After normalisation, the data is reshaped to fit the input requirements of the LSTM layers. Specifically, each sample is transformed into a 3D array of shape (samples, timesteps, features), where samples is the number of transactions, timesteps is the number of features per transaction, and features is set to 1 to indicate a single time-step input per feature.

The first layer in the BiLSTM model is a Bidirectional LSTM (BiLSTM) layer with 32 units. The bidirectional nature of this layer means it consists of two LSTMs: one processing the input sequence forward and the other backward. This configuration allows the model to capture dependencies and patterns in both directions, optimising its capacity to comprehend complex temporal relationships within the data. The LSTM units use the tanh activation function, which helps in squashing the outputs to be between -1 and 1, providing a non-linear transformation that captures intricate patterns within the sequence data.

A Dropout layer with a rate of 0.3 is implemented subsequent to the BiLSTM layer. Dropout randomly zeros a percentage of input units during training. This prevents the model from becoming too dependent on any collection of neurons, thereby preventing overfitting. This enhances the model's generalisation capability. Next, the model includes a Dense layer with 16 units and a ReLU (Rectified Linear Unit) activation function. This layer serves to combine the features extracted by the BiLSTM layer, performing higher-level abstractions. The ReLU activation is chosen for its efficiency and ability to mitigate the vanishing gradient problem, making the training process faster and more effective.

The last Dense layer has one neuron with sigmoid activation. This layer generates a probability score between 0 and 1, which denotes the probability of a fraudulent transaction. The sigmoid activation is particularly suited for binary classification tasks, as it maps the input to a range between 0 and 1, facilitating probabilistic interpretation. Table 5.2 shows the BiLSTM model structure.

Table 5.2 *BiLSTM model structure.*

Layer (Type)	Output Shape	Param #
Bidirectional (Bidirectional)	(None, 64)	8704
Dropout (Dropout)	(None, 64)	0
Dense (Dense)	(None, 16)	1040
dense_1 (Dense)	(None, 1)	17
Total params: 9,761 Trainable params: 9,761 Non-trainable params: 0		

The choice of BiLSTM layers in this model is driven by their superior ability to learn long-term dependencies in sequential data. LSTMs are designed to remember information over long periods, which is critical for tasks involving sequences where past and future contexts are relevant. The bidirectional architecture enhances this capability by processing the sequence in both forward and backward directions, thereby capturing a more comprehensive range of patterns and dependencies. By incorporating Dropout layers, the model mitigates the risk of overfitting, which is essential given the imbalance in the dataset where fraudulent transactions are much rarer than non-fraudulent ones. The dense layers further abstract the features learned by the BiLSTM, culminating in a robust binary classification through the sigmoid-activated output layer.

This BiLSTM model balances complexity and computational efficiency while maintaining high accuracy and robustness in detecting fraudulent transactions, making it well-suited for real-time detection applications. Model's design ensures that it can effectively generalise from the training data to accurately identify cyber fraud in new, unseen transactions.

The dataset is first pre-processed by selecting relevant features and splitting it into training and testing sets. The features are then normalised using the StandardScaler. The normalised data is reshaped to fit the input requirements of the LSTM layer, where each sample is transformed into a 3D array with dimensions suitable for the BiLSTM model. The model is compiled using the Adam optimiser, binary cross-entropy loss, and accuracy as a metric. The model is then wrapped in a KerasClassifier for integration with scikit-learn's cross-validation functions. The training process involves using Stratified K-Fold cross-validation with 5 splits to ensure the model is evaluated robustly on different subsets of the data.

During training, the model's performance is monitored on both the training and validation sets using the accuracy metric. The learning curves for accuracy and loss are plotted to visualise the model's learning process over 20 epochs, with a ReduceLROnPlateau callback applied to adjust the learning rate dynamically. This callback helps the model to converge more smoothly and avoid getting stuck in local minima by reducing learning when validation set performance plateaus. Once the training phase is complete, the model is assessed using the test set. Predictions are made, and several performance metrics are calculated. Overall, the BiLSTM model's structured approach in training and evaluation ensures a robust and effective system for detecting fraudulent credit card transactions.

5.3 Novel hybrid CNN-BiLSTM model

The novel hybrid DL model leverages a sophisticated architecture that combines Convolutional Neural Networks (CNN) and Bidirectional Long Short-Term Memory (BiLSTM) networks. This model is meticulously designed to capture both local patterns and long-term dependencies in sequential data, optimising performance for binary classification tasks. The model architecture is comprised of numerous layers, each of which makes a distinct contribution during the feature extraction and classification process:

Conv1D Layer: The architecture begins with a 1-dimensional convolutional layer equipped with 64 filters and a kernel size of 3. This layer utilises the ReLU activation function to introduce non-linearity and efficiently capture local patterns in the input sequence. The input shape for this layer is defined as (number of features, 1). By setting the number of filters to 64, the model aims to learn distinct feature maps from the raw input.

MaxPooling1D Layer: The MaxPooling layer is implemented to downscale the feature maps, with a pool size of 2. This layer reduces the spatial dimensions of the feature maps by half, retaining the most significant features while mitigating the computational load and helping to prevent overfitting.

Bidirectional LSTM Layer: Following the convolutional layers, the model employs a Bidirectional LSTM (BiLSTM) layer with 64 units. Unlike traditional LSTMs, BiLSTMs process the input sequence in both forward and backward directions, enabling the model to capture dependencies from both past and future contexts.

Dropout Layer: To prevent overfitting, a Dropout layer with a dropout rate of 0.5 is added after the BiLSTM layer. This technique randomly drops 50% of the neurons during each training iteration, which helps in promoting model generalisation by preventing co-adaptation of neurons.

Dense Layer: After the dropout mechanism, there is a Dense layer that is fully linked and has 32 neurons. This layer is activated by ReLU. This layer integrates the features extracted by the previous layers, allowing the model to learn complex feature interactions.

Output Layer: The final Dense layer consists of a single neuron with a sigmoid activation function. This layer generates a probability score between 0 and 1, which denotes the probability of the positive class. The sigmoid activation is particularly suited for binary classification tasks, as it effectively maps the input to a possibility distribution.

Convolutional Layers (Conv1D) is essential for detecting local patterns and hierarchical feature representations within the data. These layers excel at capturing short-term dependencies and significant local features, which form the foundational building blocks for further processing. MaxPooling Reduces the spatial dimensions, retaining the most salient features while reducing computational complexity and mitigating overfitting. LSTMs are inherently powerful for modelling long-term dependencies in sequential data. By employing a bidirectional approach, the model gains a comprehensive understanding of the sequence, capturing context from both past and future states. This bidirectional processing is particularly advantageous for sequences where the order of elements is critical. The dense layers enable the integration and transformation of features extracted by previous layers. The inclusion of dropout layers mitigates overfitting by introducing stochasticity during training, promoting better generalisation to unseen data. Table 5.3 shows the novel hybrid CNN- BiLSTM model structure.

Table 5.3 *Novel hybrid CNN- BiLSTM model structure.*

Layer (Type)	Output Shape	Param #
conv1d_1 (Conv1D)	(None, 15, 64)	256
max_pooling1d_1 (MaxPooling1D)	(None, 7, 64)	0
bidirectional_1 (Bidirectional)	(None, 128)	66048
dropout_1 (Dropout)	(None, 128)	0
dense_2 (Dense)	(None, 32)	4128
dense_3 (Dense)	(None, 1)	33
Total params: 70,465 Trainable params: 70,465 Non-trainable params: 0		

The Adam optimiser is chosen for its adaptive learning rate capabilities and efficient handling of sparse gradients. The binary cross-entropy loss function is appropriate for binary classification tasks, as it measures the difference between the predicted probabilities and the actual class labels. This loss function is particularly effective for models outputting probabilities, ensuring that the model's predictions are calibrated accurately. ReLU activation function is employed in the hidden layers due to its ability to introduce non-linearity. ReLU is computationally efficient and helps in mitigating the disappearing gradient issue, allowing the model to learn more effectively. Callbacks are functions that can be applied at certain stages of the training process, such as at the end of each epoch. In the novel hybrid CNN- BiLSTM model, we define two callbacks: ReduceLROnPlateau and EarlyStopping. ReduceLROnPlateau reduces the learning rate when a metric has stopped improving, and EarlyStopping stops training when a monitored metric has stopped improving. These callbacks help in improving training efficiency. Cross-validation using cross_val_score was used to evaluate the model's performance on several training data subsets. Finally, we fit the model on the entire training set (X_train_normalized, y_train) using the fit method. During the training process, we provide the number of epochs, batch size, validation split, and callbacks that will be utilised. This step trains the model on the training data and evaluates its performance on the validation set.

The sequence of layers in this model is meticulously designed to maximise the strengths of different neural network components and to ensure efficient and effective learning. Starting with Conv1D layers is crucial for extracting local patterns. These layers are adept at identifying short-term dependencies and significant local features, which serve as the foundational building blocks for subsequent processing stages. Introducing MaxPooling early reduces the dimensionality of the feature maps, retaining essential features and reducing computational complexity. Following the convolutional layers with BiLSTM layers allows the model to capture long-term dependencies and contextual information from both past and future states. The bidirectional nature of these layers enhances the model's ability to understand the sequential context of the data. The Dense layers integrate and transform the features extracted by the previous layers. These layers, combined with Dropout, reduce the dimensionality and focus on the most relevant aspects for classification, ensuring robust and accurate predictions. The sigmoid activation function is well-suited for binary classification, as it effectively maps the input to a possibility score between 0 and 1.

The combination of Conv1D and BiLSTM layers ensures that the model captures both local patterns and long-term dependencies, providing a comprehensive feature extraction and understanding of the data. MaxPooling layers enhance the stability and efficiency of the model, ensuring faster and more reliable training. The inclusion of Dropout layers mitigates overfitting, promoting better generalisation to unseen data. This novel hybrid CNN- BiLSTM model combines the strengths of CNN and BiLSTM architectures to effectively capture and prioritise important features in sequential data. This design leads to improved performance in binary classification tasks by leveraging local pattern recognition and long-term dependency modelling. The careful selection and ordering of layers, along with robust algorithmic components, make this model a powerful tool for tackling complex binary classification challenges.

Algorithm 5.1: Novel hybrid CNN-BiLSTM model

```
Procedure CNN_BiLSTM (input, n, j, batch)
    Pre-process (input)
    Split data into (X_train, X_test, y_train, y_test)
for k ← 0 to n-1 do
        Model ← create_CNN_BiLSTM_model()
            Function create_CNN_BiLSTM_model()
            Input Layer: Conv1D
            MaxPooling1D
           Bidirectional LSTM
           Dropout
          Dense
          Output Layer: Dense
           Compile Model
           End Function
        for epoch ← 0 to j-1 do
            Train Model (X_train, y_train, batch_size, epochs)
        end for
        Evaluate Model (X_test, y_test)
end for
End Procedure
```

The novel hybrid CNN- BiLSTM model is a DL architecture designed for detecting cyber fraud in credit card, featuring a combination of convolutional and recurrent layers. It begins with a Conv1D layer with 64 filters and a kernel size of 3, which captures local patterns in the input data. This is followed by a MaxPooling1D layer that reduces the spatial dimensions. The model then incorporates a Bidirectional LSTM layer with 64 units in each direction, allowing it to capture long-term dependencies and contextual information from both past and future sequences. A Dropout layer with a 50% dropout rate is included to prevent overfitting.

Finally, Two Dense layers are incorporated into the model: the first with 32 neurons and ReLU activation for integrating features, and the second with one neuron and sigmoid activation to produce positive class probability.

We trained and evaluated the novel hybrid DL model using a comprehensive dataset containing credit card transactions. The dataset was pre-processed to normalise the features and split into training and testing sets to assess the model's generalisation ability. During training, we employed cross-validation techniques and monitored performance metrics. Additionally, we utilised callbacks such as early stopping and learning rate reduction to prevent overfitting and enhance convergence.

5.4 Experimental result

The findings and results section of this thesis presents a detailed examination of three advanced DL models applied to detect cyber fraud. Each model, namely the CNN model, BiLSTM model, and the novel hybrid CNN-BiLSTM model, is meticulously evaluated to assess its efficacy in distinguishing between fraudulent and legitimate transactions. We assess the performance of CNN, BiLSTM, and CNN-BiLSTM models.

5.4.1 CNN model

The evaluation of the CNN model reveals a comprehensive assessment of its performance across various metrics. Comparing the performance metrics of the CNN model across different configurations of epoch size and batch size provides valuable insights into the model's behaviour under varying training conditions.

In the scenario with an epoch size of 20 and a batch size of 64, the model demonstrates a loss value of 0.002668, indicating minimal errors during the training process. A total of 77 true positives (TP) have been obtained, with only 11 false positives (FP), indicating a high precision of 87.50%, which represents the model's ability to minimise false alarms. However, the recall rate of 78.57% suggests that the model may miss some actual fraudulent transactions. The F1 score, a harmonic mean of precision and recall, is 82.80%. The AUC is 89.28%, suggesting the model's strong ability to discriminate between fraudulent and legitimate transactions. Similarly, the area under the precision-recall curve (PRC) is 68.79% indicating the model's effectiveness in identifying fraudulent cases.

In the scenario with an epoch size of 50 and a batch size of 128, the model exhibits a slightly lower loss value of 0.001816, suggesting improved training efficiency compared to the previous configuration. The number of true positives increases to 76, while the false positives increase marginally to 7. Consequently, the precision decreases slightly to 91.57%, indicating a slightly higher rate of false alarms compared to the previous configuration. However, the recall rate improves to 77.55%, indicating a better ability to capture actual fraudulent transactions. The F1 score remains relatively stable at 83.98%, reflecting a similar balance between precision and recall as observed in the previous configuration. The AUC improves slightly to 0.8877, indicating a slightly better discriminatory ability compared to the previous configuration. Similarly, the PRC also shows a slight improvement to 71.05%, indicating a slightly better performance in identifying fraudulent cases.

Overall, the comparison highlights the sensitivity of the CNN model's performance to variations in training parameters such as epoch size and batch size. While both configurations exhibit strong performance in detecting fraudulent transactions, the choice between them may depend on considerations such as computational resources and the specific requirements of the application. Table 5.4 shows the result of CNN model using several Epochs. Figure 5.2 shows the training and validation accuracy. Figure 5.3 training and validation loss. Figure 5.4 shows Confusion matrix for CNN model.

Table 5.4 *Results of CNN model using several Epochs.*

Matrix	**Epoch size 20, batch size 64**	**Epoch size 50, batch size 128**
Loss	0.002668	0.001816
TP	77	76
FP	11	7
TN	56853	56857
FN	21	22
Accuracy	99.94%	99.95%
Precision	87.50%	91.57%
Recall	78.57%	77.55%
Cross-Validation/Mean Accuracy	99.93%	99.94%
F1 score	82.80%	83.98%
AUC	89.28%	88.77%
PRC	68.79%	71.05%
Total fraudulent transaction	98	98

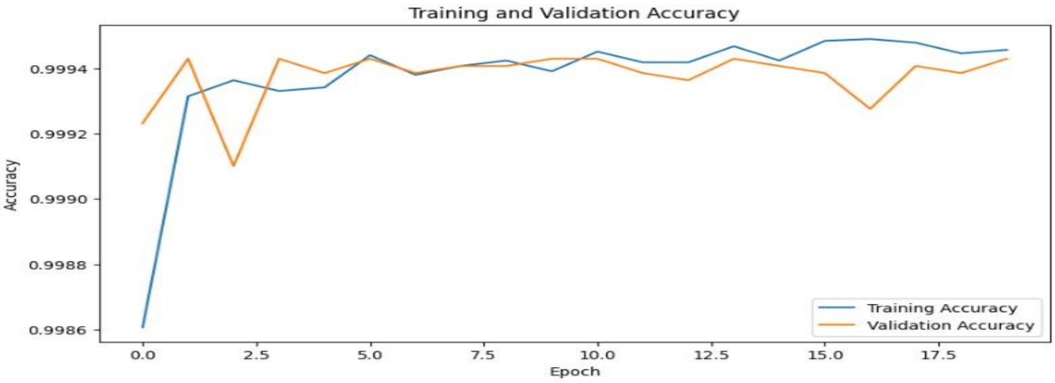


Figure 5.2 *Training and validation accuracy (CNN).*

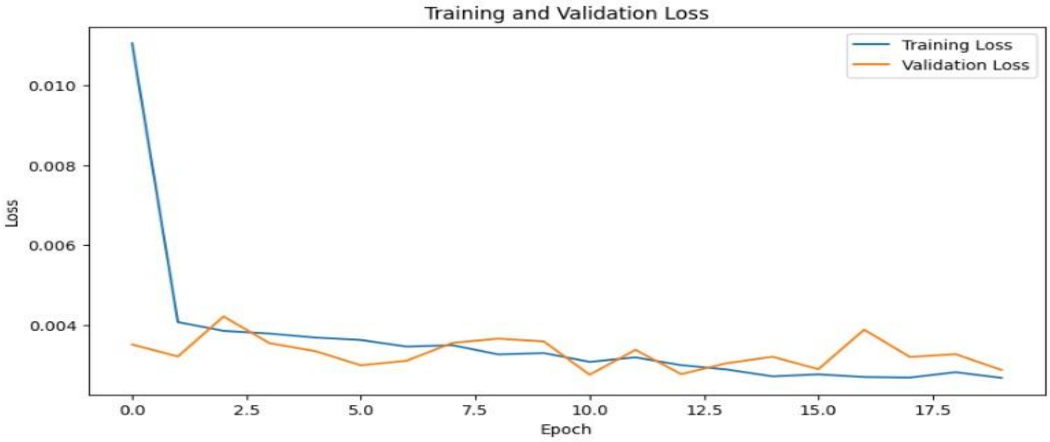


Figure 5.3 *Training and validation loss (CNN).*

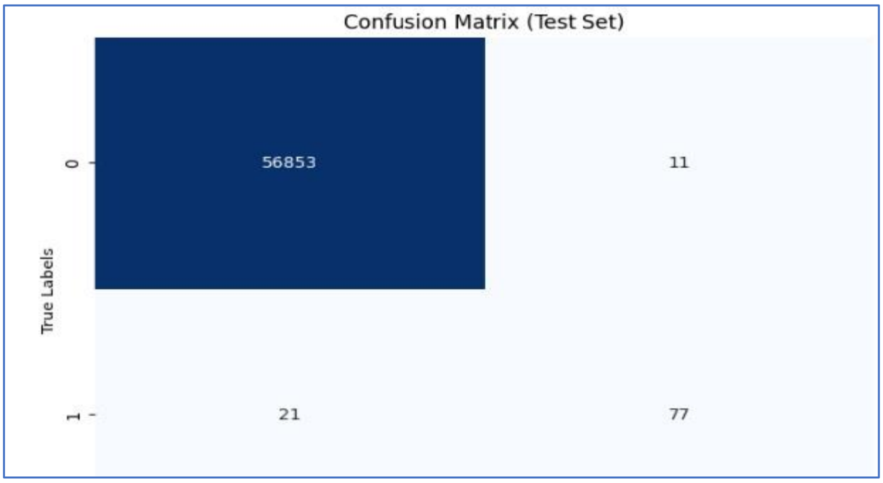


Figure 5.4 *Confusion matrix of CNN model.*

CNN demonstrates consistent and robust performance across different configurations of epoch size and batch size. In both scenarios, the CNN achieves exceptionally high accuracy, with values exceeding 99.9%, indicating its effectiveness in detecting fraud.

The CNN maintains high precision values, which suggest a low rate of false positives. The precision of the data increases slightly as the epoch and group sizes increase, both configurations still exhibit precision values above 91.5%, highlighting the CNN's ability to correctly identify fraudulent transactions without falsely flagging legitimate ones.

The model's recall rate, which assesses its capacity to accurately identify every relevant instances of fraudulent transactions, remains consistently high in both configurations, with values surpassing 78.5%. This indicates that the CNN effectively captures the majority of fraudulent cases, minimising the number of false negatives.

The F1 score remains high in both scenarios, exceeding 83.8%. The results of this imply that the performance is balanced between precision and recall, with the CNN achieving high accuracy while effectively identifying fraudulent transactions.

In terms of the area under the receiver operating characteristic curve (AUC), the CNN demonstrates strong discriminatory ability. This implies the CNN can accurately detect fraudulent from non-fraudulent transactions. Similarly, the area under the precision-recall curve (PRC) also indicates strong performance, with values exceeding 71.04%. This suggests that the CNN performs well in identifying fraudulent cases, particularly in scenarios where the dataset may be imbalanced.

The final loss value on the test set is 0.002668 and 0.001816 which indicates a low average error between the predicted and actual class labels. The mean accuracy of the model during cross-validation is 99.94%. This indicates that, on average, the model correctly classifies 99.94% of the samples across different folds of the dataset. Such high accuracy is impressive, especially for a simplified model, and suggests that the model is substantially successful at distinguishing between fraudulent and non-fraudulent transactions. Overall, CNN demonstrates consistent and reliable performance in detecting fraudulent transactions, making it a robust choice for cyber fraud detection applications.

5.4.2 BiLSTM model

Comparing the performance metrics of the BiLSTM model across different configurations of epoch size and batch size provides insights into its behaviour under varying training conditions.

In the scenario with an epoch size of 20 and a batch size of 64, the model demonstrates a loss value of 0.00295, indicating minimal errors during the training process. A total of 76 true positives (TP) have been obtained, with 11 false positives (FP), resulting in a precision of 87.36%. The recall rate is 77.55%, and the F1 score is 82.16%, suggesting a competent balance between precision and recall. The AUC is 88.77%, indicating the model's strong ability to discriminate between fraudulent and legitimate transactions. However, the area under the precision-recall curve (PRC) is 67.78%, suggesting a relatively lower performance in identifying fraudulent cases compared to the receiver operating characteristic curve (ROC).

In the scenario with an epoch size of 50 and a batch size of 128, the model exhibits a slightly lower loss value of 0.002386, suggesting improved training efficiency compared to the previous configuration. The number of true positives increases to 82, with 11 false positives, resulting in a precision of 88.17%. The recall rate improves to 83.67%, and the F1 score increases to 85.86%, indicating a better balance between precision and recall compared to the previous configuration. The AUC improves to 91.83%, indicating a better discriminatory ability compared to the previous configuration. Similarly, the PRC also shows improvement to 73.80%, indicating a better performance in identifying fraudulent cases. While both configurations exhibit strong performance in detecting fraudulent transactions, the choice between them may depend on considerations such as computational resources and the specific requirements of the application. Table 5.5 shows the result of BiLSTM model using several Epochs. Figure 5.5 shows the training and validation accuracy. Figure 5.6 training and validation loss. Figure 5.7 shows Confusion matrix for BiLSTM model.

Table 5.5 *Results of BiLSTM Model using several Epochs.*

Matrix	**Epoch size 20, batch size 64**	**Epoch size 50, batch size 128**
Loss	0.00295	0.002386
TP	76	82
FP	11	11
TN	56853	56853
FN	22	16
Accuracy	99.94%	99.95%
Precision	87.36%	88.17%
Recall	77.55%	83.67%
Cross-Validation/Mean Accuracy	99.94%	99.94%
F1 score	82.16%	85.86%
AUC	88.77%	91.83%
PRC	67.78%	73.80%
Total fraudulent transaction	98	98

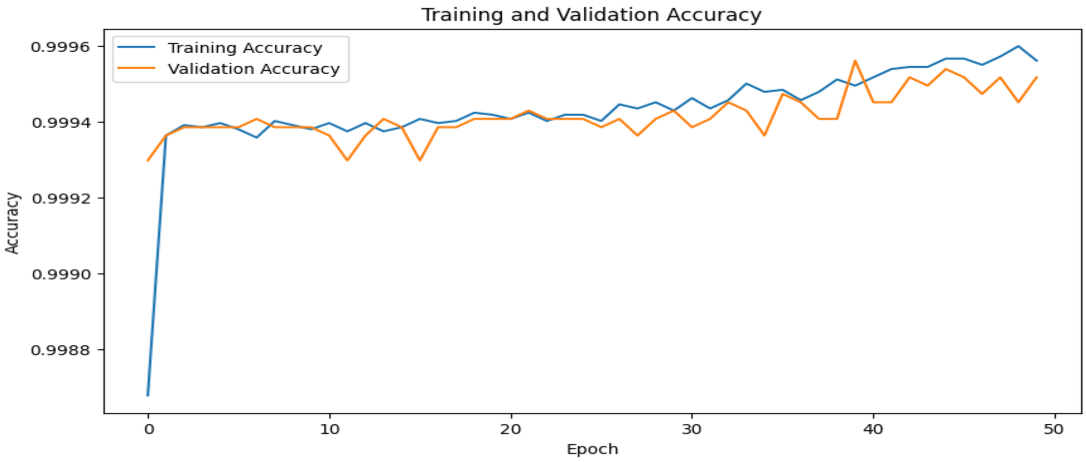


Figure 5.5 *Training and validation accuracy(BiLSTM).*

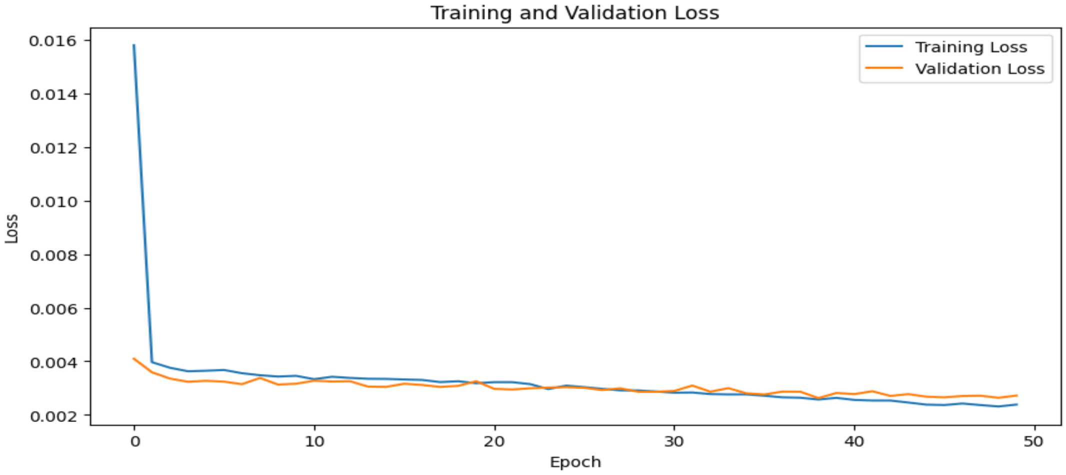


Figure 5.6 *Training and validation loss(BiLSTM).*

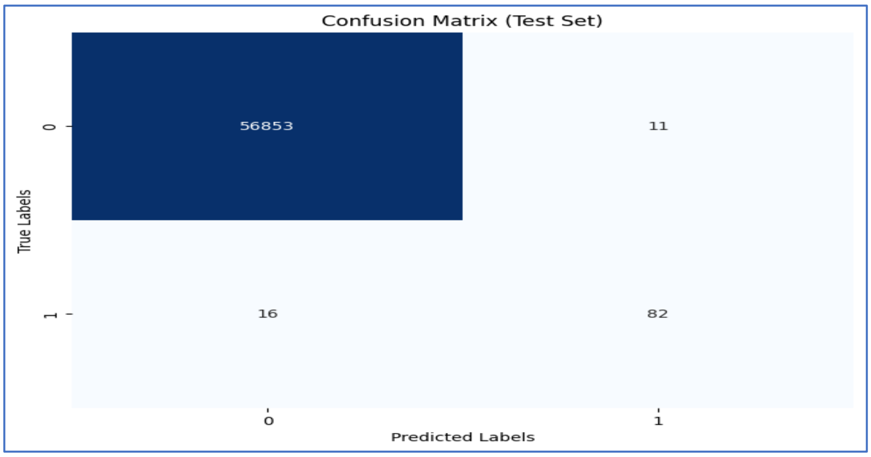


Figure 5.7 *Confusion matrix of BiLSTM model.*

The BiLSTM model demonstrates strong performance in detecting fraudulent transactions, with consistent results across different configurations of epoch size and batch size. In terms of accuracy, the BiLSTM model achieves high values, exceeding 99.9% in both configurations. This indicates its ability to correctly classify transactions as fraudulent or non-fraudulent with a high degree of precision. The precision of the BiLSTM model remains consistently high, with values exceeding 87% in both scenarios. This suggests a low rate of false positives.

Recall rate of the BiLSTM model also remains consistently high across both configurations, exceeding 77.5%. This suggests that the model is capable of accurately capturing the majority of fraudulent cases, minimising the number of false negatives. The F1 score remains high in both scenarios, exceeding 82%. The BiLSTM model achieving high accuracy while effectively identifying fraudulent transactions.

The AUC further confirms the BiLSTM model's strong discriminatory ability, with values exceeding 0.887 in both configurations. This suggests that the model can effectively distinguish between fraudulent and non-fraudulent transactions with high accuracy. Similarly, the PRC indicates strong performance, with values exceeding 67.70% in the first configuration and 73.80% in the second configuration. This suggests that the BiLSTM model performs well in identifying fraudulent cases, particularly in circumstances where the dataset may be imbalanced. The final loss value on the test set is 0.002386, which suggests a low average error between the predicted and actual class labels. The mean accuracy of the model during cross-validation is 99.94%. This indicates that, on average, the model correctly classifies 99.94% of the samples across different folds of the dataset. This level of accuracy is noteworthy, particularly for a simplified model, and it implies that the model is highly effective in distinguishing between fraud and non-fraud. The standard deviation of 0.0001 shows that the accuracy scores across different folds are very close to the mean accuracy. This low standard deviation implies that the model's performance is consistent and stable across various subsets of the data. Consistency in performance is crucial for ensuring the model's reliability when applied to new, unseen data. Overall, the BiLSTM model demonstrates consistent and reliable performance in detecting fraudulent transactions, making it a robust choice for cyber fraud detection applications.

5.4.3 Novel hybrid CNN-BiLSTM

The model demonstrated exceptional performance across various metrics, showcasing its efficacy in the detection of fraudulent transactions. During the training phase, the loss decreased consistently from an initial value of 0.0074 to 0.0021, while the accuracy remained impressively high, stabilising at around 99.95%. This trend indicates that the model effectively learned the distinguishing features of fraudulent transactions without significant overfitting. The model's capacity to generalise effectively to unseen data was further justified by the validation loss, which exhibited a downward trend. The validation accuracy stabilised at approximately 99.94%, demonstrating a high level of consistency between the training and validation sets.

On the test set, the model's accuracy was observed to be 99.96%. Precision and recall, crucial metrics for evaluating the model's effectiveness in cyber fraud detection, were reported at 93.02% and 81.63%, respectively. The high precision suggests that when the model predicts a transaction as fraudulent, it is correct 93.02% of the time. However, the recall, although strong, is lower than precision, indicating that the model successfully identifies 81.63% of actual fraudulent transactions. This balance between precision and recall is reflected in the F1 score of 86.96%, which underscores a well-maintained equilibrium between the two metrics.

The ROC AUC score of 97.28% serves to emphasise the model's capacity to differentiate between fraudulent and non-fraudulent transactions. This high score indicates excellent discrimination capability. Similarly, the Precision-Recall AUC score of 85.80% provides additional assurance of the model's effectiveness, especially in handling the imbalanced nature of the dataset. These metrics collectively emphasise the model's robustness in distinguishing between the two classes.

The confusion matrix reveals that the model identified 80 out of 98 fraudulent transactions correctly (True Positives) but missed 18 (False Negatives). Additionally, the model accurately classified 56,858 non-fraudulent transactions (True Negatives) while incorrectly flagging 6 non-fraudulent transactions as fraudulent (False Positives). Table 5.6 shows the result of novel hybrid CNN-BiLSTM model. Figure 5.8. Training and validation accuracy(CNN-BiLSTM). Figure 5.9 Training and validation loss (CNN-BiLSTM). Figure 5.10 shows the confusion matrix of CNN-BiLSTM model.

Table 5.6 *Results of Novel Hybrid CNN-BiLSTM Model.*

Matrix	CNN-BiLSTM model(Epoch 100, Batch size 64)
Loss	0.002147
TP	80
FP	6
TN	56858
FN	18
Accuracy	99.96%
Precision	93.02%
Recall	81.63%
F1 score	86.96%
AUC	97.28%
PRC	85.80%
Total fraudulent transaction	98

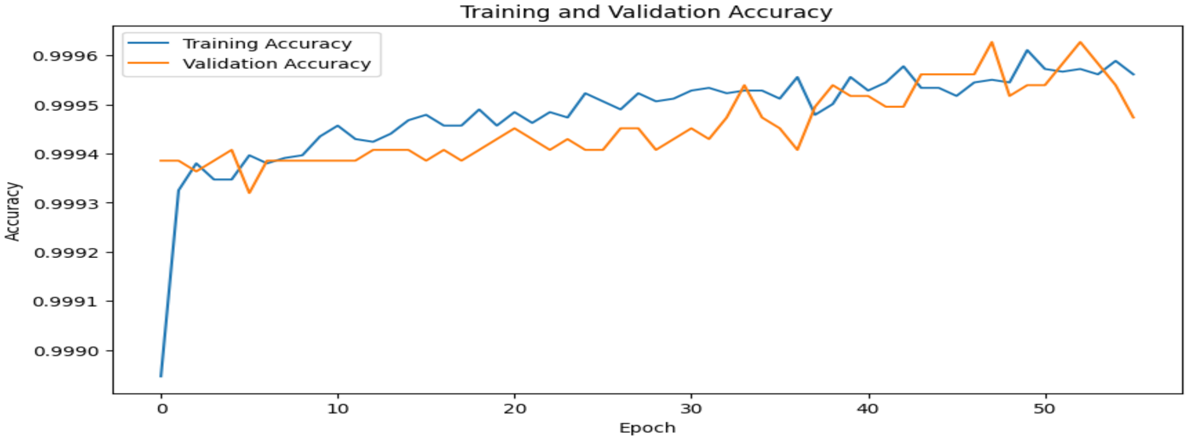


Figure 5.8 *Training and validation accuracy(CNN-BiLSTM).*

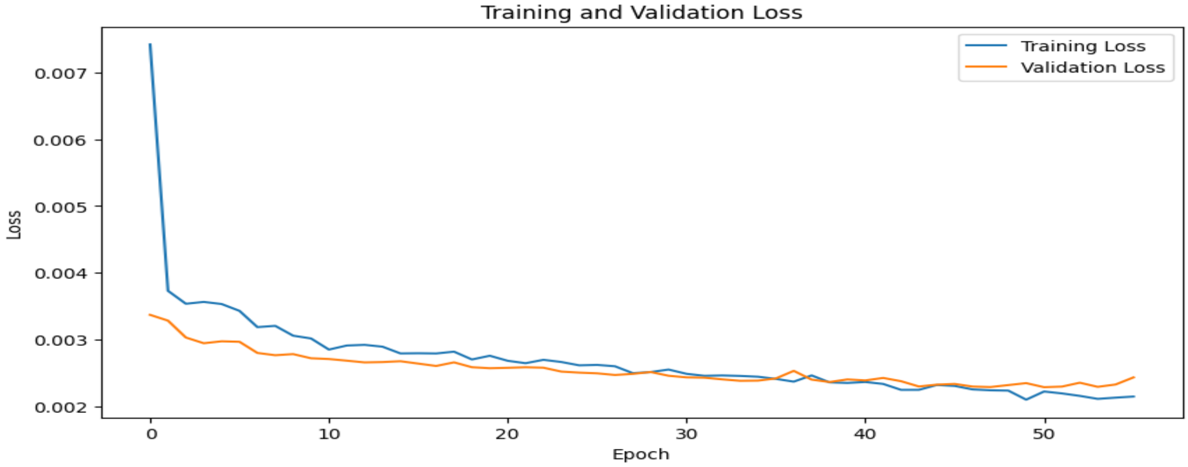


Figure 5.9 *Training and validation loss(CNN-BiLSTM).*

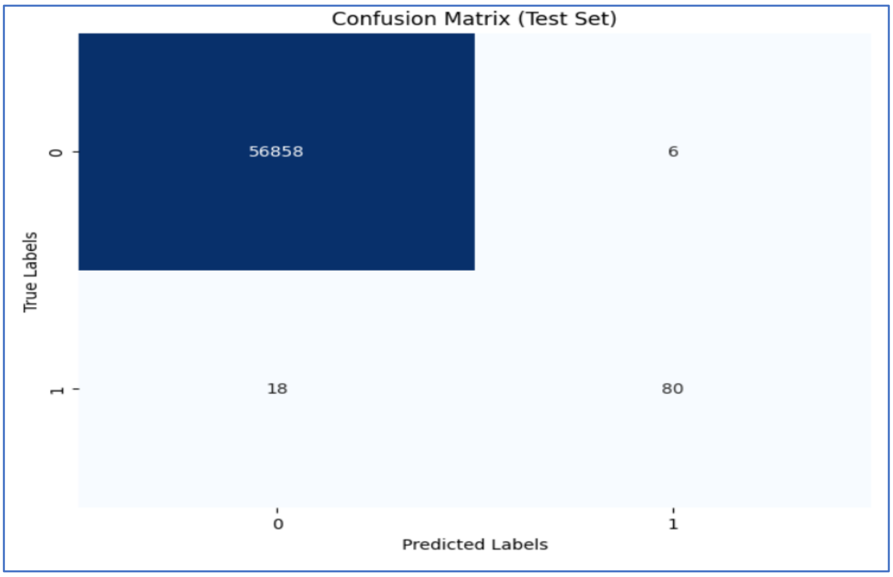


Figure 5.10 *Confusion matrix of CNN-BiLSTM model.*

5.4.4 Comparison novel hybrid CNN-BiLSTM model with CNN and BiLSTM

The performance of the three models—CNN, BiLSTM, and CNN-BiLSTM—was evaluated using several key metrics: loss, true positives (TP), false positives (FP), true negatives (TN), false negatives (FN), accuracy, precision, recall, F1 score, AUC, and PRC. Each model exhibited unique strengths and weaknesses, providing insights into their efficacy in cyber fraud detection. The CNN model achieved the lowest loss value of 0.001816, indicating its high efficiency in minimising errors during the training process. In comparison, the BiLSTM model had the highest loss at 0.002386, suggesting it faced more difficulty in learning from the training data. The hybrid CNN-BiLSTM model had a loss of 0.002147, which was slightly higher than the CNN but lower than the BiLSTM, reflecting a balanced learning process that integrates both convolutional and sequential processing capabilities.

In terms of detecting fraudulent transactions (True Positives), the BiLSTM model performed the best with 82 correctly identified fraudulent transactions, followed by the CNN-BiLSTM model with 80, and the CNN model with 76. Conversely, the CNN-BiLSTM model had the lowest number of false positives (6), making it the most precise in avoiding incorrect cyber fraud classifications. The CNN model had 7 false positives, while the BiLSTM model had the highest with 11. Regarding true negatives and false negatives, all models performed similarly, with slight variations that align with their precision and recall metrics.

The CNN-BiLSTM model exhibited the highest accuracy at 99.96%, showcasing its exceptional capacity to accurately classify both illegitimate and non-fraudulent transactions. The BiLSTM model followed closely with an accuracy of 99.953%, and the CNN model was slightly lower at 99.949%. These high accuracy values reflect the overall effectiveness of all three models in handling the classification task.

Precision was highest in the CNN-BiLSTM model at 93.02%. The CNN model also showed strong precision at 91.57%, while the BiLSTM model lagged at 88.17%. Recall, on the other hand, was highest for the BiLSTM model at 83.67%, suggesting it was more effective in identifying actual fraudulent transactions. The CNN-BiLSTM model had a recall of 81.63%, and the CNN model had the lowest recall. These differences highlight a trade-off between precision and recall among the models.

The F1 score, which balances precision and recall, was highest for the CNN-BiLSTM model at 86.96%, indicating its superior overall performance in cyber fraud detection. The BiLSTM model followed with an F1 score of 85.86%, showing its robustness despite lower precision. The CNN model, while strong, had the lowest F1 score at 83.98%, reflecting its relatively lower recall. The CNN-BiLSTM model also excelled in AUC (97.28%) and PRC (85.80%). The BiLSTM model achieved an AUC of 91.83% and a PRC of 73.80%, indicating good but lesser performance compared to the CNN-BiLSTM. The CNN model had the lowest AUC and PRC , which still indicate solid performance but less effectiveness in dealing with class imbalance and prediction robustness.

Focusing on the F1 score, the CNN-BiLSTM model stands out as the best performer with an F1 score of 86.96%. This indicates its balanced approach in maintaining high precision and recall, making it highly effective for cyber fraud detection tasks where both metrics are crucial. The BiLSTM model, with an F1 score of 85.86%, also performed well, particularly excelling in recall. The CNN model, while strong in precision, had a lower F1 score of 83.98%, highlighting its relative weakness in recall. Therefore, the CNN-BiLSTM model is the most robust and balanced choice among the three, providing the best performance in detecting fraudulent transactions accurately and consistently. Table 5.7 shows Comparison of models performance. Figure 5.11 shows comparison of models performance.

Table 5.7 *Comparasion of models performance.*

Matrix	CNN	BiLSTM	CNN-BiLSTM model
Loss	0.001816	0.002386	0.002147
TP	76	82	80
FP	7	11	6
TN	56857	56853	56858
FN	22	16	18
Accuracy	99.949%	99.953%	99.958%
Precision	91.57%	88.17%	93.02%
Recall	77.55%	83.67%	81.63%
F1 score	83.98%	85.86%	86.96%
AUC	88.77%	91.83%	97.28%
PRC	71.05%	73.80%	85.80%
Total fraudulent transaction	98	98	98

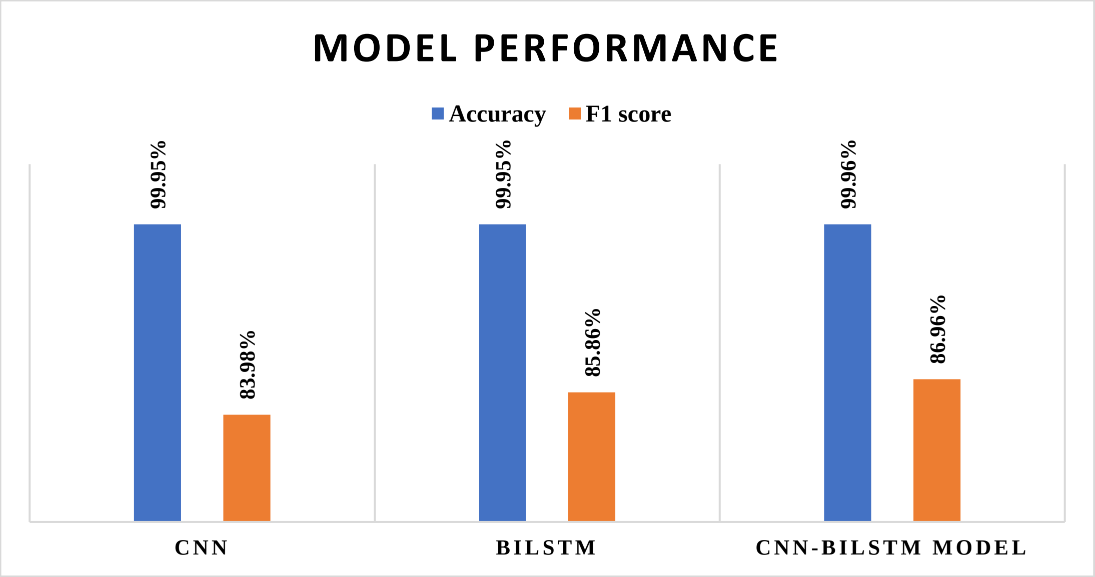


Figure 5.11 *Comparasion of models performance.*

In conclusion, the comparative analysis of the CNN, BiLSTM, and CNN-BiLSTM models reveals that the CNN-BiLSTM hybrid model offers the most robust performance for cyber fraud detection, achieving the highest F1 score of 86.96%, superior precision, and a well-balanced recall. This model also exhibited exceptional overall metrics, such as the highest accuracy, AUC, and PRC, which suggest its ability to accurately differentiate between fraudulent and non-fraudulent transactions. The CNN-BiLSTM model is the optimal choice for effective cyber fraud detection due to its balanced and superior performance, which surpasses that of the BiLSTM model in recall and the CNN model in loss minimisation.

5.5 Chapter Summary

This chapter presented a comparative analysis of three DL models—CNN, BiLSTM, and CNN-BiLSTM—tailored for cyber fraud detection. Despite the remarkable effectiveness of the novel hybrid DL models, there has been a noticeable shift towards employing DL techniques in recent years, driven by their enhanced precision and optimised performance. DL enables systems to adapt to complex data patterns and identify new fraudulent behaviours, surpassing traditional ML methods. As ML techniques have limitations, further research into DL algorithms is recommended for better cyber fraud detection outcomes, especially considering the skewed nature of credit card cyber fraud datasets.

The CNN-BiLSTM hybrid model, developed and tested on a real-world dataset, stands out due to its integration of convolutional layers and bidirectional LSTM networks, capturing both local and sequential data patterns. The CNN component efficiently captures complex data patterns, while the BiLSTM component handles temporal dependencies effectively. The model architecture consists of layers designed to optimise feature extraction and classification, including Conv1D layers, MaxPooling, Bidirectional LSTM, and dense layers, all contributing to a balanced and robust performance in cyber fraud detection.

Evaluation metrics such as loss, true positives, false positives, accuracy, precision, recall, F1 score, AUC, and PRC were used to assess the performance of the models. The CNN-BiLSTM model demonstrated superior performance, achieving the highest accuracy (99.96%), precision (93.02%), and F1 score (86.96%). Its AUC (97.28%) and PRC (85.80%) values further highlight its robust ability to distinguish between fraudulent and non-fraudulent transactions. While the BiLSTM model showed strengths in recall and the CNN model excelled in minimising loss, the CNN-BiLSTM model's balanced approach ensures its superior effectiveness in cyber fraud detection tasks, making it the optimal choice for real-world applications.

CHAPTER 6: NOVEL HYBRID STACKING ML+DL MODEL

Although, the effectiveness of the CNN-BiLSTM hybrid model in Chapter 5 was outstanding and despite the model's strengths, there are areas for improvement. The F1 score while good, indicates that there are still fraudulent transactions that go undetected. Improving F1score is critical in cyber fraud detection scenarios to minimise the risk of overlooking fraudulent activities. This suggests that although the model demonstrates outstanding results on the training data, it is necessary to guarantee that it can successfully apply its knowledge to new data. To address these concerns, considering the limitations of each ML/DL technique and in order to detect credit card cyber fraud more effectively and achieve encouraging detection outcomes, it is recommended to contemplate the integration of ML and DL algorithms in order to achieve encouraging detection outcomes. Several scholarly articles have proposed integrating DL methods with conventional ML methods in order to improve the accuracy of credit card cyber fraud detection from unbalanced datasets. Researchers are also encouraged to employ both undersampling and oversampling methods due to the exceedingly skewed nature of the datasets. The techniques of ML and DL were described. An evaluation was conducted on both ML and DL classifiers utilised in credit cyber fraud detection. The evaluation encompassed an analysis of their merits and drawbacks, the number of identified classes, the datasets utilised, and the measurement metrics employed.

The novel hybrid ML+ DL model, developed in this chapter, consists of combining the ML techniques used in Chapter 4; (DT, RF, SVM, LR, XGBoost, and CatBoost) with DL techniques CNN-BiLSTM-Attention using ensemble techniques and sampling techniques. The novel hybrid ML+ DL model was trained and tested in real-world dataset.

The research work's novelty and contributions were as follows:

1. A new approach including ML techniques and DL techniques which their outputs were merged using Stacking ensemble techniques with and without sampling techniques to classify transaction to cyber fraud and non-fraud.
2. The whole proposed framework as known Hybrid ML+DL with ensemble and sampling techniques model in this research was trained and tested in a real-world dataset and the obtained results indicates the novel hybrid ML+ DL model has high performance in comparison with using only ML or DL in the state-of-the-art techniques and baseline models.

3. A comprehensive comparison between the novel models has been conducted to thoroughly evaluate their respective performances and identify the strengths and weaknesses of each approach. This detailed analysis aims to provide clear insights into how each model handles the complexities of cyber fraud detection, including their ability to accurately identify fraudulent transactions, minimise false positives, and effectively manage class imbalance. This comparison not only highlights the best-performing models but also offers valuable recommendations for future improvements and practical applications in real-world scenarios. In the following, a brief description of the ensemble learning explained and then the novel hybrid ML+ DL classifier is elaborated. Figure 6.1 shows Novel Hybrid Stacking ML + DL model.

6.1 Novel hybrid stacking ML+DL model

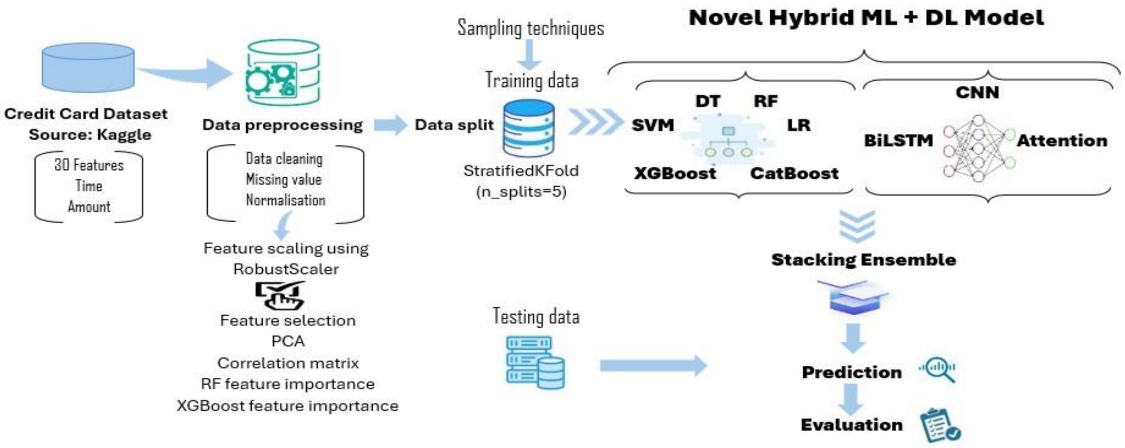


Figure 6.1 *Novel Hybrid Stacking ML + DL model.*

The model introduced in this chapter integrates ML techniques outlined in Chapter 4 with advanced DL architecture that detailed in Chapter 5. Through meticulous refinement, enhancements have been implemented, including the incorporation of additional layers aimed at improving the efficacy of credit card cyber fraud detection. This incorporation leverages the combined effect between ML and DL paradigms, with the goal of achieving heightened accuracy and resilience in identifying fraudulent transactions. The model's design emphasises the utilisation of the strengths inherent in both methodologies to enhance the overall effectiveness of systems. In this chapter, the novel hybrid stacking DL+ML model incorporates several additional layers not used in the chapter 5 basic DL model, including multiple Conv1D layers, Batch Normalisation layers, additional Bidirectional LSTM layers, and an Attention mechanism. These components significantly enhance the model's capacity to extract relevant features. The inclusion of three Conv1D layers with increasing filter sizes enables the novel hybrid stacking ML+ DL model to capture more complex and abstract patterns by progressively deepening feature extraction. This hierarchical approach improves the model's ability to discern subtle differences in input sequences, making it more adept at recognising nuanced variations crucial for accurate classification.

The integration of Batch Normalisation layers after each Conv1D layer further stabilises and accelerates the training process. This leads to faster convergence, improved performance, and mitigates issues like vanishing or exploding gradients, enhancing training reliability. Additionally, the novel hybrid stacking ML+ DL model in this chapter employs two BiLSTM layers with 128 and 64 units, respectively, compared to a single BiLSTM layer with 64 units in the chapter 5 basic model. This allows the novel hybrid stacking ML+ DL model in this chapter to more effectively capture long-term dependencies and contextual information from both past and future states, providing a comprehensive view of the entire sequence and enhancing the ability to learn complex temporal patterns. The incorporation of an Attention mechanism further distinguishes the novel hybrid stacking ML+ DL model in this chapter, as it dynamically weighs the importance of different time steps, focusing on the most relevant parts of the input. This significantly improves the model's interpretability and efficiency, ensuring critical information is prioritised and leading to more precise and contextually relevant classifications. Collectively, these additional components make the novel hybrid stacking ML+ DL model in this chapter more powerful and efficient, enhancing its performance in binary classification tasks by better capturing complex patterns, stabilising training, and prioritising important features.

The heart of the model lies in its architecture, which includes both ML classifiers and a DL model. The classical classifiers include RF, SVM, LR, DT, XGBoost, and CatBoost. These classifiers are trained individually on the training data and evaluated using standard metrics. Additionally, we introduce a CNN-BiLSTM with an attention mechanism. The CNN architecture consists of convolutional layers followed by batch normalisation, max-pooling layers, Bidirectional LSTM layers, dropout layers for regularisation, and a custom AttentionLayer. This AttentionLayer helps the model focus on important features. The CNN is trained alongside the classical classifiers and is evaluated similarly. Moreover, we employ various callbacks during training, such as F1ScoreCallback, ModelCheckpoint, ReduceLROnPlateau, and EarlyStopping, to monitor the model's performance and prevent overfitting. Finally, we build a StackingClassifier that combines the predictions of all classifiers ML and DL, using a Random Forest classifier as the final estimator. This StackingClassifier is trained on the training data and evaluated on the test set. The performance metrics, including accuracy, precision, recall, F1-score, ROC AUC score, and confusion matrix, are computed and visualised for comprehensive analysis. In summary, the novel hybrid stacking ML+ DL model is a hybrid approach that combines the strengths of classical ML algorithms with the representation learning capabilities of a CNN-BiLSTM with an attention, resulting in a robust and effective framework for detecting fraudulent transactions.

6.1.1 ML techniques

ML model in this chapter combines various ML techniques that have been applied in chapter 4, including DT, RF, SVM, XGBoost, CatBoost, and LR, each tailored to address specific challenges in credit card cyber fraud detection. For instance, the DT model utilises recursive partitioning to create a tree-like structure, ensuring accurate classification by analysing features and decision points. On the other hand, RF constructs multiple decision trees to improve predictive accuracy while preventing overfitting. SVM optimise the separation between distinct classes in the feature space, hence improving their effectiveness in binary classification problems. XGBoost employs gradient boosting to iteratively refine predictions, while CatBoost efficiently handles categorical features without pre-processing hassles. Lastly, LR provides a fundamental yet powerful approach to binary classification, offering insights into the relationships between features and the target variable.

6.1.2 DL techniques

The model architecture consists of 18 layers, each contributing uniquely to the feature extraction and classification process:

1-Conv1D Layer: The architecture begins with a 1-dimensional convolutional layer equipped with 32 filters and a kernel size of 3. This layer utilises the ReLU activation function to introduce non-linearity and efficiently capture local patterns in the input sequence. The input shape for this layer is defined as (X_train.shape [1], 1), where X_train.shape [1] represents the number of features in each input sample. By setting the number of filters to 32, the model aims to learn 32 distinct feature maps from the raw input.

2-Batch Normalisation Layer: Immediately following the first Conv1D layer, Batch Normalisation is applied. This technique normalises the activations of the previous layer; by minimising internal covariate shift, the training process is substantially improved in terms of stability and efficiency. It helps in maintaining a consistent distribution of inputs across layers, which is crucial for deep networks.

3-MaxPooling1D Layer: The MaxPooling layer is implemented to downscale the feature maps, with a pool size of 2. This layer reduces the spatial dimensions and retains the most significant features while mitigating the computational load and helping to prevent overfitting.

4-Conv1D Layer: The model then incorporates a second Conv1D layer with 64 filters and a kernel size of 3. This layer continues to capture more complex patterns, building on the representations learned by the first convolutional layer.

5-Batch Normalisation Layer: Like the first set of layers, Batch Normalisation is applied to further stabilise the learning process and reduce the dimensionality of the feature maps.

6-MaxPooling1D Layer: Another MaxPooling layer is applied to downsample the feature maps, retaining the most important features while reducing computational complexity and preventing overfitting.

7-Conv1D Layer: The filter size is increased to 128 in the third Conv1D layer, which also maintains a kernel size of 3. This layer allows the model to learn even more abstract and higher-level features from the input sequence, which are crucial for accurate classification.

8-Batch Normalisation Layer: Batch Normalisation is applied again to ensure stable and efficient learning by normalising the activations.

9-MaxPooling1D Layer: A third MaxPooling layer is applied to further reduce the spatial dimensions of the feature maps, retaining the most critical features.

10-Bidirectional LSTM Layer: Following the convolutional layers, the model employs a Bidirectional LSTM (BiLSTM) layer with 128 units. Unlike traditional LSTMs, The BiLSTM algorithm processes the input sequence in both the forward and backward orientations, allowing the model to incorporate dependencies from both past and future contexts. This bidirectional processing is particularly beneficial for understanding the sequential nature of the data and for improving the context-awareness of the model.

11-Dropout Layer: To prevent overfitting, a Dropout layer with a dropout rate of 0.5 is added after the first BiLSTM layer. Randomly, 50% of the neurons are removed during each training iteration using this technique, which helps in promoting model generalisation by preventing co-adaptation of neurons.

12-Bidirectional LSTM Layer: Another Bidirectional LSTM layer with 64 units is then included, providing a more compact representation of the sequential data while still benefiting from bidirectional context.

13-Dropout Layer: An additional Dropout layer with the same dropout rate of 0.5 is applied after the second BiLSTM layer to further reduce overfitting and improve model robustness.

14-Attention Layer: The attention mechanism is a crucial component that dynamically weighs the importance of different time steps in the sequence. The Attention Layer computes a context vector by focusing on the most relevant parts of the sequence. This layer works by assigning higher weights to the time steps that contribute more significantly to the output prediction. The introduction of this mechanism enhances the model's interpretability and efficiency by allowing it to focus on the most informative segments.

15-First Dense Layer: An attention-driven fully linked layer with 64 neurons and ReLU activation. This layer incorporates features from preceding levels to teach the model complicated feature interactions.

16-Dropout Layer: Another Dropout layer with a rate of 0.5 is applied to prevent overfitting.

17-Second Dense Layer: A second Dense layer with 32 neurons with ReLU activation reduces dimensionality and prioritizes classification criteria.

18-Output Layer: In the final Dense layer, a single neuron is equipped with a sigmoid activation function. This layer generates a probability score ranging from 0 to 1, which denotes the probability of the positive class. The sigmoid activation is particularly suited for binary classification tasks, as it effectively maps the input to a probability distribution. Table 6.1 shows the Deep Learning model structure.

The sequence of layers in this model is meticulously designed to maximise the strengths of different neural network components and to ensure efficient and effective learning. Starting with Conv1D layers is crucial for extracting local patterns. These layers can discover short-term relationships and important local features, which provide the basis for later processing. Introducing Batch Normalisation early stabilises the learning process, The computational intricacy is reduced by MaxPooling, which reduces the dimensionality of the feature maps while retaining essential features. Deeper Conv1D Layers, by progressively increasing the number of filters in subsequent Conv1D layers, the model can capture more complex and abstract patterns. This hierarchical feature extraction is essential for understanding the underlying structure of the data. Following the convolutional layers with BiLSTM layers allows the model to capture long-term dependencies and contextual information from both past and future states. The bidirectional nature of these layers enhances the model's ability to understand the sequential context of the data. Placing the Attention Layer after the BiLSTM layers allows the model to dynamically focus on the most relevant parts of the sequence, improving interpretability and performance. The attention mechanism ensures that the model gives more weight to important time steps, enhancing its predictive capabilities. Dense layers following the attention mechanism integrate and transform the features extracted by the previous layers. These layers, combined with Dropout, reduce the dimensionality and focus on the most relevant aspects for classification, ensuring robust and accurate predictions. Finally, the sigmoid activation function in the output layer is well-suited for binary classification, as it effectively maps the input to a probability score between 0 and 1.

The model's exhaustive comprehension of the data is guaranteed by the combination of Conv1D and BiLSTM layers, which capture both local patterns and long-term dependencies. Batch Normalisation and MaxPooling layers enhance the stability and efficiency of the model, ensuring faster and more reliable training. The attention mechanism enhances the model's interpretability and efficiency by allowing it to focus on the most informative parts of the sequence, improving overall performance. The inclusion of Dropout layers mitigates overfitting, promoting better generalisation to unseen data. In summary, this hybrid model combines the strengths of CNN and BiLSTM architectures with an attention mechanism to effectively capture and prioritise important features in sequential data. This design leads to improved performance in binary classification tasks by leveraging local pattern recognition, long-term dependency modelling, and dynamic attention-based feature weighting. The careful selection and ordering of layers, along with robust algorithmic components, make this model a powerful tool for tackling complex binary classification challenges.

The Adam optimiser is selected due to its efficient management of sparse gradients and adaptive learning rate capabilities. The binary cross-entropy loss function measures the difference between predicted probabilities and class labels, making it suited for binary classification applications. This loss function is particularly effective for models outputting probabilities, ensuring that the model's predictions are calibrated accurately. ReLU activation function is employed in the hidden layers due to its ability to introduce non-linearity. ReLU is computationally efficient and helps in mitigating the vanishing gradient problem, allowing the model to learn more effectively.

Table 6.1. *The Deep Learning structure.*

Layer (Type)	Output Shape	Param #
conv1d (Conv1D)	(None, 15, 32)	128
batch_normalization (BatchNormalization)	(None, 15, 32)	128
max_pooling1d (MaxPooling1D)	(None, 7, 32)	0
conv1d_1 (Conv1D)	(None, 5, 64)	6208
batch_normalization_1 (BatchNormalization)	(None, 5, 64)	256
max_pooling1d_1 (MaxPooling1D)	(None, 2, 64)	0
conv1d_2 (Conv1D)	(None, 2, 128)	24,704
batch_normalization_2 (BatchNormalization)	(None, 2, 128)	512
max_pooling1d_2 (MaxPooling1D)	(None, 1, 128)	0
Bidirectional (Bidirectional)	(None, 1, 256)	263,168
Dropout (Dropout)	(None, 1, 256)	0
bidirectional_1 (Bidirectional)	(None, 1, 128)	164,352
dropout_1 (Dropout)	(None, 1, 128)	0
attention_layer (AttentionLayer)	(None, 128)	129
dense (Dense)	(None, 64)	8,256
dropout_2 (Dropout)	(None, 64)	0
dense_1 (Dense)	(None, 32)	2,080
dense_2 (Dense)	(None, 1)	33
Total params: 469,954 Trainable params: 469,506 Non-trainable params: 448		

Algorithm 6.1: Novel hybrid stacking ML+DL model

```
Procedure Stacking Hybrid ML+DL _model (X, y, n, cv_folds, test_size)
     Pre-process (X, y)
     Split data into (X, y)
     Normalize (X_train, X_test)
     Model ← Create stacking classifier with RandomForestClassifier
                     as meta-classifier()
              ('Random Forest', rf_classifier)
              ('SVM', svm_classifier)
              ('Logistic Regression', lr_classifier)
              ('Decision Tree', dt_classifier)
              ('XGBoost', xgb_classifier)
              ('CatBoost', catboost_classifier)
              ('CNN-BiLSTM-Attention ', CNN-BiLSTM-Attention _classifier)

         Cross-validation: StratifiedKFold with cv_folds
for k ← 0 to n-1 do
     Train StackingClassifier (X_train, y_train)
     Predict y_pred on (X_test)
     Evaluate Model(X_test, y_test)
     Print Evaluation Metrics

end for
End Procedure
```

6.2 Experimental result

Initially, we evaluate the performance of the Novel hybrid stacking ML+ DL model without utilising any sampling techniques. The results from this evaluation are thoroughly detailed in Table 6.2. Additionally, Figure 6.2 presents the confusion matrix.

Table 6.2 *Novel hybrid ML+DL model performance.*

Model	Accuracy	Precision	Recall	F1 score	AUC
Novel hybrid stacking ML+DL	99.97%	97.62%	83.67%	90.11%	91.83%

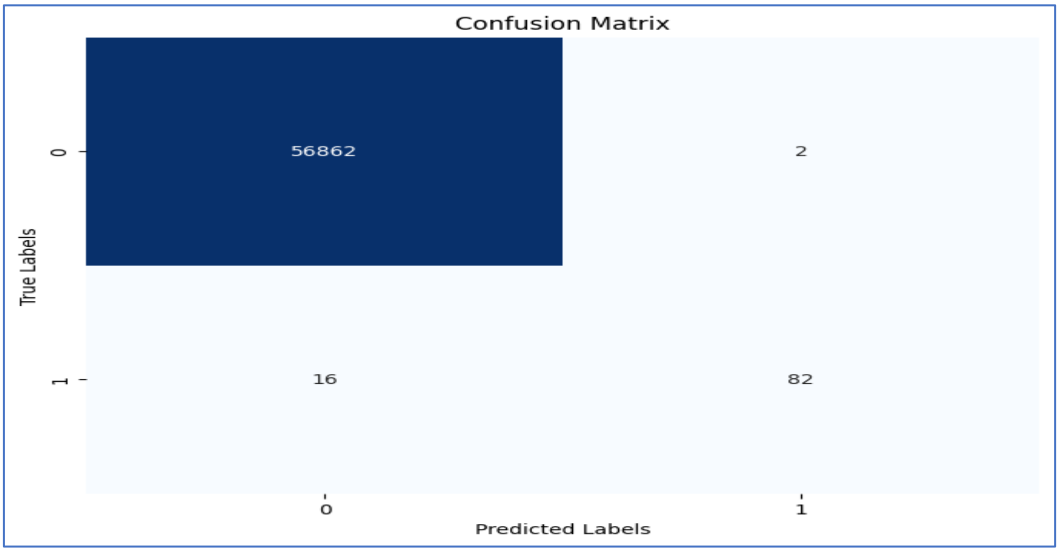


Figure 6.2 Confusion matrix Novel hybrid ML+DL model.

The novel hybrid ML+DL model exhibits outstanding performance across several assessment measures. With an accuracy of 99.97%, the model showcases its ability to accurately classify instances, which is crucial in credit card cyber fraud detection where even minor errors can have significant consequences. Additionally, the model achieves a high precision of 97.62%, indicating a low rate of false positives. This is particularly advantageous in cyber fraud detection, where correctly identifying fraudulent transactions is paramount to minimising financial losses for both customers and financial institutions.

Moreover, the model exhibits a commendable recall of 83.67%, highlighting its capability to capture a high proportion of actual positive cases. In cyber fraud detection scenarios, where the number of fraudulent transactions is typically much lower than legitimate ones, a high recall ensures that the model effectively identifies fraudulent activities, thereby enhancing cyber fraud detection efficiency. The F1 score of 90.11% further emphasises the model's balanced performance between precision and recall, demonstrating its ability to manage false positives and false negatives. The AUC value of 91.83% underscores the model's ability to distinguish between fraudulent and legitimate transactions effectively. A high AUC suggests that the model performs well across various threshold values, further reinforcing its reliability in making accurate predictions.

The novel hybrid ML+DL model excels in its intricate design, which effectively integrates the advantages of both classic ML and DL techniques. The ML techniques incorporated, such as DT, RF, SVM, XGBoost, CatBoost, and LR, offer a diverse set of tools tailored to address specific challenges in credit card cyber fraud detection. These techniques leverage various algorithms and strategies to effectively capture patterns and make accurate predictions. Furthermore, the DL techniques utilised, including CNN and BiLSTM networks, are well-suited for handling sequential data such as transaction sequences. The integration of attention mechanisms further enhances the model's ability to focus on relevant segments of the input sequence, improving interpretability and performance. Overall, the novel hybrid ML+ DL model benefits from comprehensive feature extraction, stability, and efficiency, dynamic attention mechanisms, robustness, and generalisation capabilities. It is an effective tool for addressing intricate binary classification challenges, particularly in credit card cyber fraud detection scenarios, due to its high-level architecture, meticulous layer sequencing, and robust algorithmic components.

6.2.1 Comparison between the three novel models

This analysis evaluates the performance of three models: the novel hybrid stacking ML+DL model (described in this chapter), the hybrid ML model discussed in Chapter 4, and the hybrid CNN-BiLSTM model presented in Chapter 5. Table 6.3 displays the performance metrics of the three models.

Table 6.3 *Comparison performance.*

Model	Accuracy	Precision	Recall	F1 score
Hybrid Stacking ML model (chapter 4)	99.96%	98.73%	79.59%	88.14%
Hybrid CNN-BiLSTM model (chapter 5)	99.96%	93.02%	81.63%	86.96%
Hybrid Stacking ML+DL (chapter 6)	99.97%	97.62%	83.67%	90.11%

In terms of accuracy, all three models exhibit exceptionally high performance. The novel Hybrid Stacking ML+DL achieves the highest accuracy at 99.97%, demonstrating its enhanced capacity for accurate instance classification. ML model from Chapter 4 closely follows with an accuracy of 99.963%, while the DL model from Chapter 5 shows a slightly lower accuracy of 99.958%. This suggests that while all models are highly accurate, the hybrid stacking ML+DL model slightly edges out the others in overall correct classifications.

Precision measures the proportion of true positives among the predicted positives, and the novel hybrid stacking ML+DL model leads with a precision of 97.62%. This high precision indicates that the hybrid stacking ML+DL model has the low rate of false positives. The ML model from Chapter 4 also performs well in this regard, with high precision of 98.73%. The DL model from Chapter 5, however, has a lower precision of 93.02%, suggesting a relatively higher rate of false positives compared to the other models.

Recall, which measures the proportion of actual positives correctly identified, is highest in the novel hybrid stacking ML+DL at 83.67%. This demonstrates its effectiveness in capturing true positive cases. Both the ML and DL models from Chapter 4 and Chapter 5, respectively, show identical recall values of 79.59% and 81.63%. This indicates that while they are proficient at identifying actual positives, they do not perform as well as the Hybrid model in this metric.

The F1 score is another critical metric where the novel hybrid stacking ML+DL model excels, achieving 90.11%. This score indicates that the model is the most reliable in terms of precision and recall, as it exhibits a balanced performance. The ML model from Chapter 4 follows with a little lower F1 score of 88.14%, still demonstrating strong performance. The DL model from Chapter 5 has the lowest F1 score of 86.96%, indicating that it is less balanced in handling precision and recall compared to the other models.

In summary, The novel hybrid stacking ML+DL stands out as the most balanced and robust model among the three, particularly excelling in its F1 score of 90.11%. This high F1 score reflects a harmonious balance between precision and recall, indicating that the Hybrid model minimises false positives effectively. Such a balance is crucial in many practical applications where both the identification of true cases and the avoidance of false alarms are equally important. Therefore, the novel hybrid stacking ML+DL model’s superior performance across multiple metrics, especially its F1 score, highlights its potential as a highly effective tool for tackling class imbalance in complex datasets.

The novel hybrid stacking ML+DL model demonstrates superior performance compared to the most advanced models in terms of accuracy and F1 score on the European dataset. While some models achieve high accuracy individually, the novel hybrid approach, which combines ML and DL through stacking, achieves the highest accuracy of 99.97% and a competitive F1 score of 90.11%. This suggests that the integration of both ML and DL techniques in a stacking framework enhances the model's predictive capabilities, offering promising results for card cyber fraud detection. Table 6.4 shows comparison performance with existing models.

Table 6.4 *Comparison performance with existing models.*

Study Ref.	Model	Accuracy	F1score	Dataset
Muaz et al. (2020)	Stacked ensemble	78%	---	European cards
Alghofaili et al. (2020)	LSTM	99.95%	---	European cards
Najadat et al. (2020)	BiLSTM-MaxPooling-BiGRU-MaxPooling	91.37%	---	European cards
Nguyenet al. (2020)	LSTM	---	84.85%	European cards
Owolafe et al. (2021)	LSTM-RNN	99.58 %	88.76%	European cards
Agarwal et al. (2021)	CNN-BiLSTM	95%	---	European cards
Alarfaj et al. (2022)	CNN	99.72%	---	European cards
Arun & Rajesh, (2022)	BEPO-OGRU	94.78%	---	European cards
Malik et al. (2022)	Adaboost+LGBM	---	77%	European cards
Handa et al. (2022)	Hybrid ensemble	99.90%	---	European cards
Yamini et al. (2023)	CNN-ELM	98.7%	---	European cards
Al Balawi & Aljohani, (2023)	CNN	99.81	83.72%	European cards
Maheshwari et al. (2023)	RNN-LSTM-Attention	99.4%	---	European cards
Jahnavi et al. (2024)	Hybrid DT -LR-RF	98.1%	---	European cards
Paulraj, (2024)	hybrid CNN-RNN	99.20%	---	European cards
Alamri & Ykhlef, (2024)	Hybrid ML+BCBSMOTE	---	85.20%	European cards
Novel Hybrid ML+DL model	Stacking ML+DL	99.97%	90.11%	European cards

6.2.2 The novel hybrid stacking ML+DL model with sampling

The Novel Hybrid Stacking ML+DL Model employs a variety of resampling techniques to address the issue of class imbalance in our dataset. Class imbalance is a common issue in real-world applications such as cyber fraud detection, where the minority class (e.g., fraudulent transactions) is significantly underrepresented compared to the majority class. To address this, we employ Borderline-SMOTE (Borderline Synthetic Minority Over-sampling Technique), ROS (RandomOverSampler), Tomek Link, ENN, and SMOTEENN (SMOTE + ENN) as strategies to rebalance our dataset and enhance the performance of our hybrid stacking ML+DL model.

A. Borderline-SMOTE

In the novel hybrid stacking ML+DL model, we have applied Borderline-SMOTE to address the class imbalance in the dataset. Borderline-SMOTE is an advanced variant of SMOTE specifically designed to improve the classifier's performance on the minority class by generating synthetic samples close to the decision boundary where most misclassifications occur. Borderline-SMOTE first identifies the minority class instances that are difficult to classify correctly. These are typically the instances that are near the decision boundary between classes. Instead of generating synthetic samples randomly for all minority class instances, Borderline-SMOTE generates new samples specifically for these borderline instances. This helps in making the decision boundary clearer and more accurate. By focusing on the borderline instances, Borderline-SMOTE enhances the classifier’s ability to learn the decision boundary more effectively, thereby improving the model's ability to generalise to unseen data. Borderline-SMOTE is applied to the training data to balance the classes before training the novel hybrid stacking ML+DL model.

B. RandomOverSampler (ROS)

In the novel hybrid stacking ML+DL model, we have applied RandomOverSampler to address the class imbalance in the dataset. ROS is a straightforward and effective method that balances the classes by randomly duplicating examples from the minority class until the classes are equally represented. This process increases the number of minority class samples, balancing the dataset. ROS first identifies the minority class in the dataset. It then randomly selects and duplicates instances from the minority class until the number of samples in the minority class equals the number of samples in the majority class. By duplicating the minority class samples, the dataset becomes balanced, helping the classifier to learn from a more representative sample of each class. ROS is applied to the training data to balance the classes before training the ML and DL components of the novel hybrid stacking ML+DL model.

C. Tomek Links

In the novel Hybrid Stacking ML+DL model, we have applied Tomek Links to address class imbalance in the dataset. Tomek Links is a data cleaning technique that aims to improve class separability by removing instances that are ambiguously located near the decision boundary between classes. Tomek Links are pairs of instances from different classes that are each other's nearest neighbours. These pairs often lie near the decision boundary, representing ambiguous or noisy points. Removing such pairs can make the classes more distinct and improve classifier performance. For each pair of instances from different classes, Tomek Links checks if they are the nearest neighbours of each other. If such a pair is found, one or both instances can be removed. This process helps in cleaning the dataset by eliminating borderline instances that may confuse the classifier.

D. Edited Nearest Neighbours (ENN)

In the novel hybrid stacking ML+DL model, we have applied ENN to tackle the class imbalance in the dataset. ENN, like Tomek Links, is a data cleaning technique aimed at enhancing class separability by identifying and removing instances that are ambiguous or noisy, particularly those near the decision boundary between classes. ENN is a data cleaning technique used for handling imbalanced datasets. It works by iteratively examining each instance in the dataset and removing those that are misclassified based on their nearest neighbours. ENN identifies instances where the class label differs from most of its k nearest neighbours. Once identified, these instances are removed from the dataset. This process helps in reducing noise and improving the distinction between classes. ENN operates in two main steps: Selection of Nearest Neighbours: For each instance, ENN identifies its k nearest neighbours based on a chosen distance metric. Instance Removal: If an instance's class label differs from the majority class of its k nearest neighbours, it is removed. This step is repeated iteratively to enhance the quality of the dataset. By integrating ENN into our model, we ensure that the classifier is trained on a more refined dataset that emphasises clear class boundaries and reduces the influence of noisy or ambiguous instances. This approach enhances the model's ability to accurately identify minority class instances, such as fraudulent transactions in this context, leading to more robust and reliable predictions.

E. SMOTEENN

In the novel hybrid stacking ML+DL model, we apply SMOTEENN (SMOTE + ENN) to handle class imbalance in the dataset. SMOTEENN combines the SMOTE and ENN to generate synthetic samples for the minority class and then clean the dataset by removing ambiguous or noisy instances near the decision boundary. SMOTEENN aims to balance class distribution by increasing the number of minority class samples. ENN clean the dataset and identifies and removes instances that are misclassified based on their nearest neighbours, especially those near the decision boundary. By removing noisy or ambiguous instances, ENN enhances the distinction between classes. SMOTEENN operates in two main steps: SMOTE:

Synthetic samples are generated for the minority class to increase its representation in the dataset. This step addresses the imbalance issue by oversampling the minority class. In our hybrid stacking ML+DL model, SMOTEENN is applied to the training set to oversample the minority class using SMOTE and then clean the dataset using ENN to remove noisy or borderline instances.

We employed several resampling techniques; oversampling, undersampling, and combination of oversampling and undersampling to address the significant class imbalance present in the dataset. Results from our hybrid stacking ML+DL model integrated with the several resampling techniques are shown in Table 6.5. Figure 6.3 shows the f1 score for the sampling techniques with novel hybrid stacking ML+DL model. Figure 6.4 shows the performance of the sampling techniques with novel Hybrid Stacking ML+DL model.

Table 6.5 *Performance of resampling techniques.*

Model		Accuracy	Precision	Recall	F1 score
With sampling	Hybrid Stacking ML+DL+ Bordersmote	94.90%	99.95%	89.85%	94.63%
	Hybrid Stacking ML+DL+ROS	93.36%	99.96%	86.75%	92.89%
	Hybrid Stacking ML+DL+Tomek	99.95%	1.0	71.43%	83.33%
	Hybrid Sacking ML+DL+ENN	99.96%	98.70%	77.55%	86.86%
	Hybrid Stacking ML+DL+SMOTEEEN	94.24%	99.92%	88.56%	93.90%
Without sampling	Hybrid Stacking ML+DL	99.97%	97.62%	83.67%	90.11%

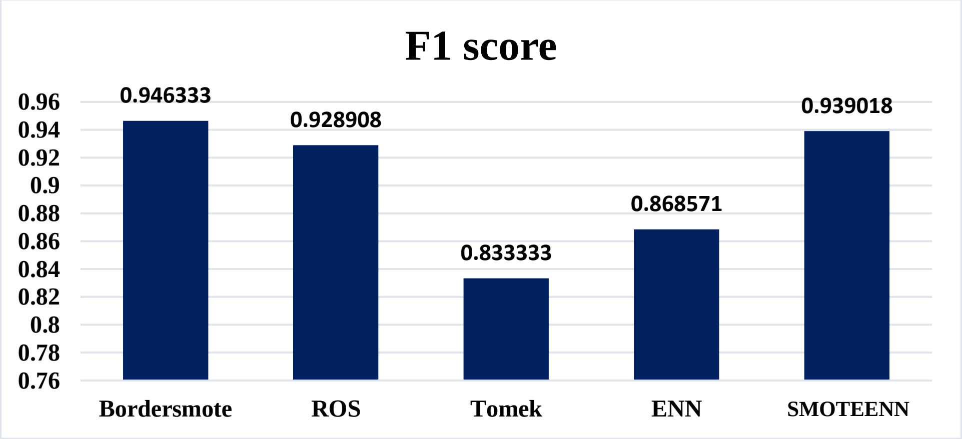


Figure 6.3 *F1 score of sampling with the novel hybrid ML+DL model.*

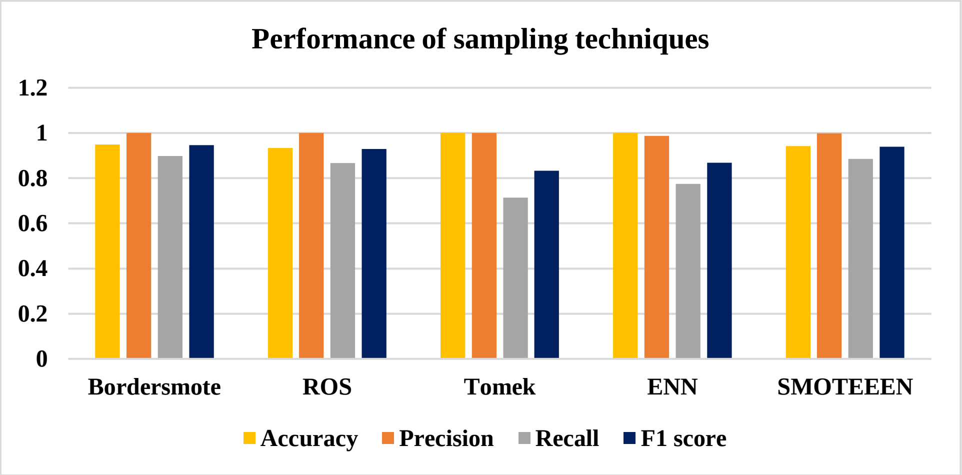


Figure 6.4 *Performance of sampling with the novel hybrid ML+DL model.*

6.2.2.1 Novel hybrid stacking ML+DL model with Borderline-SMOTE

The results from our hybrid stacking ML+DL model integrated with Borderline-SMOTE are highly encouraging and reflect substantial improvement in both the detection and classification of fraudulent transactions. The hybrid model achieved an impressive accuracy of 94.90%. This metric indicates the proportion of total transactions (both fraudulent and legitimate) that were correctly identified by the model. Although accuracy is a commonly reported metric, it can be misleading in imbalanced datasets like ours. Therefore, it is critical to consider additional performance metrics for a comprehensive evaluation. The model's precision is exceptionally high at 99.95%. This high precision value indicates that the model is highly effective at minimising false positives, ensuring that legitimate transactions are seldom incorrectly flagged as fraudulent. This is particularly crucial in practical applications, where false cyber fraud alerts can lead to customer dissatisfaction and unnecessary operational costs.

The recall of the hybrid model stands at 89.85%. While the recall is slightly lower than the precision, it still reflects a strong ability to identify a significant portion of fraudulent transactions. A higher recall value is essential to minimise the risk of undetected cyber frauds, thereby enhancing the security and trustworthiness of the financial system. The F1 score, which is the harmonic mean of precision and recall, is 94.63%. An F1 score of 94.63% demonstrates the model's overall robustness and effectiveness in distinguishing between fraudulent and legitimate transactions. The high precision and recall values achieved by the hybrid stacking ML+DL model with Borderline-SMOTE indicate a well-balanced performance. The model excels at both correctly identifying cyber frauds and minimising false positives, addressing the critical challenge of class imbalance in cyber fraud detection. The slightly lower recall compared to precision suggests that while the model is highly conservative in flagging cyber frauds, it might miss some fraudulent transactions. This compromise is frequently deemed to be permissible in the context of cyber fraud detection where the cost of false positives (legitimate transactions flagged as fraud) is higher than the cost of false negatives (fraudulent transactions not flagged).

6.2.2.2 Novel stacking hybrid ML+DL model with ROS

The results from our hybrid stacking ML+DL+ROS model indicate a commendable performance in identifying fraudulent transactions, as demonstrated by the following key metrics. The hybrid stacking ML+DL+ROS model achieved an accuracy of 93.36%. This accuracy indicates the proportion of correctly identified transactions. The precision of the model is remarkably high at 99.96%. This high precision suggests that the model is highly effective at minimising false positives, ensuring that legitimate transactions are rarely misclassified as fraudulent.

The model's recall stands at 86.75%. Although the recall is slightly lower than the precision, it still indicates that the model successfully identifies a substantial number of fraudulent transactions. High recall is essential in cyber fraud detection to minimise the number of undetected fraudulent activities. The ability to accurately and consistently differentiate between fraudulent and non-fraudulent transactions is demonstrated by an F1 score of 92.89%. The hybrid stacking ML+DL+ROS model 's ability to achieve a high F1 score indicates that it effectively manages the trade-off between identifying cyber fraud and minimising false positives.

6.2.2.3 Novel stacking hybrid ML+DL model with Tomek Links

We applied a hybrid ML-DL model integrated with Tomek Link’s undersampling technique to tackle class imbalance. The hybrid stacking ML+DL+Tomek model's performance, as indicated by the key metrics, provides insightful perspectives on its efficacy in detecting fraudulent transactions. The hybrid stacking ML+DL+Tomek model achieved an outstanding accuracy of 99.95%. The precision score for this model is a perfect 1.0, or 100%. A precision of 100% indicates that the model does not produce any false positives, ensuring that all flagged transactions are indeed fraudulent. This is crucial for maintaining customer trust and operational efficiency, as legitimate transactions are never erroneously flagged as fraudulent.

The recall rate of 71.43% suggests that while the model is excellent at confirming flagged cyber fraud cases, it does miss some fraudulent transactions. This is an important consideration, as higher recall is crucial in the domain of cyber fraud detection to ensure that as many fraudulent activities as possible are identified. The F1 score, at 83.33%, balances the model's perfect precision with its moderate recall, providing a single metric that reflects the overall robustness of the model. This score highlights that the model, while excelling in avoiding false positives, could benefit from strategies aimed at improving recall. Tomek Link's undersampling technique, used in this model, plays a significant role in these outcomes. By identifying and removing Tomek links (pairs of instances from opposite classes that are closest to each other), this technique effectively cleans the boundary between classes, helping the model to better distinguish between fraudulent and non-fraudulent transactions. However, this process also reduces the number of instances, which might contribute to the slightly lower recall.

6.2.2.4 Novel stacking hybrid ML+DL model with ENN

The performance metrics of the hybrid stacking ML+DL+ENN model reveal valuable insights into its efficiency in identifying fraudulent transactions. The hybrid stacking ML+DL+ENN model reached an exceptional accuracy of 99.96%. A precision of 98.70% suggests that the model is very reliable in identifying true cyber fraud cases with minimal false positives, which is essential for maintaining customer trust and operational efficiency. The recall of the hybrid stacking ML+DL+ENN model is 77.55%. While the recall is not as high as the precision, it indicates that the model successfully identifies a substantial portion of fraudulent activities. An F1 score of 86.86% reflects a reliable balance between precision and recall. The hybrid stacking ML+DL+ENN model shows impressive performance, particularly with its high precision and accuracy.

6.2.2.5 Novel stacking hybrid ML+DL model with SMOTEENN

The hybrid stacking ML+DL+SMOTEENN model attained a notable accuracy of 94.24%. This high accuracy indicates that the model is proficient in correctly classifying both fraudulent and non-fraudulent transactions. The model's precision stands at 99.93%. A precision of 99.93% indicates that the model is extremely reliable in identifying genuine cyber fraud cases with minimal false positives. The recall is 88.56%. While not as high as the precision, this recall rate suggests that the model successfully identifies a significant portion of fraudulent activities. An F1 score of 93.90% reflects a strong balance between precision and recall, indicating that the model performs well in both accurately identifying cyber fraud and minimising false positives. In summary, the hybrid stacking ML+DL+SMOTEENN model demonstrates impressive performance, particularly with its high precision and accuracy. This makes it a reliable method for cyber fraud detection with minimal false positives, essential for operational efficiency and customer satisfaction. These results underscore the potential of advanced hybrid techniques like SMOTEENN in enhancing cyber fraud detection systems, providing a solid foundation for further research and practical applications in this field.

6.2.3 Comparison of performance based on F1 score with and without sampling

In our evaluation, the F1 score—a metric that balances precision and recall—served as a critical measure of our hybrid stacking ML+DL model 's performance with and without various sampling techniques. The results indicate a notable variation in the model's performance depending on the sampling method applied.

With Sampling, the hybrid stacking ML+DL+Bordersmote model achieved the highest F1 score of 94.63%. According to this outcome, Bordersmote is notably effective in improving the model's capabilities to identify true positives while maintaining a low false-positive rate. It demonstrates a well-balanced performance, excelling in both precision and recall. Hybrid stacking ML+DL+SMOTEEEN followed closely with an F1 score of 93.90%. Although slightly lower than that of Bordersmote, this score still signifies a robust balance between precision and recall. SMOTEENN effectively handles class imbalance, contributing significantly to the model's performance. Hybrid stacking ML+DL+ROS resulted in an F1 score of 92.89%. While strong, this score is slightly lower than those of Bordersmote and SMOTEENN, suggesting that ROS is less efficient in balancing precision and recall. Nevertheless, ROS remains a valuable technique for enhancing the model's recall capabilities. Hybrid stacking ML+DL+ENN produced an F1 score of 0.868571. Although lower than the aforementioned methods, this score indicates that ENN does improve the model's performance, but not to the same extent as Bordersmote or SMOTEENN. Hybrid stacking ML+DL+Tomek had the lowest F1 score among the sampled methods at 83.33%. This score suggests that while Tomek Links improve the model's performance, they are less effective in achieving a balanced precision and recall compared to the other sampling techniques.

Without Sampling, the novel hybrid stacking ML+DL model, without any sampling techniques applied, achieved an F1 score of 90.11%. This is a commendable score, indicating that the model performs well even without sampling. However, it is noticeably lower than the F1 scores achieved with Bordersmote and SMOTEENN, highlighting the importance of addressing class imbalance. The comparison clearly shows that applying sampling techniques, particularly Bordersmote and SMOTEENN, significantly enhances the model's performance in terms of the F1 score. These methods effectively balance precision and recall, leading to higher F1 scores and, consequently, more reliable and efficient cyber fraud detection. While the novel hybrid stacking ML+DL model performs well on its own, the inclusion of appropriate sampling methods, such as Bordersmote and SMOTEENN, can further optimise its performance, making it more adept at identifying fraudulent activities while minimising false positives.

6.3 Chapter summary

This chapter presented a comparative analysis of hybrid stacking ML+DL model tailored for cyber fraud detection in credit card. To address the limitations of each ML/DL technique and achieve encouraging detection outcomes, it is recommended to contemplate the integration of ML and DL algorithms to achieve favourable detection results. Several scholarly articles have proposed integrating DL methods with conventional ML methods in order to improve the accuracy of credit card cyber fraud detection from unbalanced datasets. The hybrid stacking ML+DL model, developed in this chapter, consists of combining the ML techniques used in Chapter 4; (DT, RF, SVM, LR, XGBoost, and CatBoost) with improved structure of DL techniques CNN-BiLSTM-Attention used in Chapter 5 using ensemble techniques and Sampling techniques. The hybrid stacking ML+DL model was trained and tested in real-world dataset.

Without sampling, the novel hybrid stacking ML+DL model exhibits outstanding performance across multiple evaluation metrics, making it a robust solution for credit card cyber fraud detection. With an accuracy of 99.97%, the model demonstrates its exceptional ability to accurately classify instances, a critical requirement in cyber fraud detection where even minor errors can have significant consequences. The high precision of 97.62% underscores the model's efficiency in minimising false positives, which is crucial for correctly identifying fraudulent transactions and minimising financial losses. Furthermore, the model achieves a commendable recall of 83.67%, highlighting its capability to capture a significant proportion of actual cyber fraud cases. This balance between precision and recall, reflected in an F1 score of 90.11% signifies the model's robustness in handling both false positives and false negatives.

With sampling techniques, the hybrid stacking ML+DL+Bordersmote model exhibits outstanding performance across multiple evaluation metrics, making it a reliable solution for the detection of credit card cyber fraud. The model demonstrates its exceptional ability to accurately classify instances, a critical requirement in cyber fraud detection where even minor errors can have significant consequences. The high precision of 99.95% underscores the model's efficiency in minimising false positives, which is crucial for correctly identifying fraudulent transactions and minimising financial losses. Furthermore, the model achieves a commendable recall of 89.85%, highlighting its capability to capture a significant proportion of actual cyber fraud cases. This balance between precision and recall, reflected in an F1 score of 94.63% signifies the model's robustness in handling both false positives and false negatives.

The strength of the novel hybrid stacking ML+DL model lies in its sophisticated architecture that leverages both traditional ML techniques and advanced DL approaches. Incorporating a variety of ML techniques such as DT, RF, SVM, XGBoost, CatBoost, and LR, the model effectively captures patterns and makes accurate predictions. Additionally, the integration of CNNs and BiLSTM networks, equipped with attention mechanisms, enables the model manage transaction sequences. This comprehensive approach ensures dynamic feature extraction, stability, efficiency, and interpretability, making the novel hybrid stacking ML+DL model a a potent instrument for addressing intricate binary classification challenges in the detection of credit card cyber fraud.

CHAPTER 7 - CONCLUSIONS AND FUTURE WORK

This chapter concludes with a concise description of the problems offered solutions for the research and discussed the contributions. The research's limitations and prospective research are then discussed.

7.1 Conclusion remark

Detecting credit card cyber fraud is paramount for maintaining financial security and integrity, necessitating the continuous advancement of techniques to combat evolving fraudulent tactics. Traditional methods have limitations in effectively identifying fraudulent transactions, prompting the exploration of ML and DL techniques. These sophisticated approaches aim to bolster accuracy, adaptability, and overall performance in detecting fraudulent activities, thereby safeguarding financial institutions and consumers. However, challenges, such as imbalanced datasets, dynamic cyber fraud strategies, and the need for robust model generalisation persist, underscoring the need for ongoing research and innovation in this critical domain.

Although ML and DL algorithms have shown promising results in cyber fraud detection, they face challenges such as class imbalance, overfitting, and scalability issues. In addition, the nature of data, including unbalanced class distributions and evolving cyber fraud patterns, poses further challenges to the accuracy and reliability of cyber fraud detection systems.

In order to overcome these challenges and promote the prospective applications of cyber fraud detection technology in the financial sector, This research developed and assessed three innovative ML and DL cyber fraud detection models with the objective of improving the efficiency and accuracy of cyber fraud detection. The previous literature and recent advancements in cyber fraud detection methodologies were meticulously reviewed and analysed to address the key research questions posed in Chapter 1, Section 2.

Utilising the insights derived from Chapter 2, which delved into recent innovations in credit card cyber fraud detection methodologies, a thorough investigation was conducted into the application of ML and DL techniques to analyse credit card transaction data. This inquiry revealed that traditional ML methodologies fell short of adequately discerning and selecting pivotal features crucial for accurate cyber fraud detection in credit card transactions.

The contributions of this aspect of the research are multifaceted:

It provides a comprehensive exploration of recent advancements in cyber fraud detection methodologies, offering insights into the most effective ML and DL models for cyber fraud detection in credit card transactions. Through meticulous experimentation and analysis, this research contributes to the advancement of cyber fraud detection technologies, paving the way for more robust and reliable cyber fraud detection systems in the financial sector.

To address these challenges and bolster the efficacy of cyber fraud detection systems in the financial sector, this research undertakes the development and evaluation of three novel cyber fraud detection models, each designed to enhance the accuracy and efficiency of cyber fraud detection algorithms in credit card transactions. The contributions and findings of this research are outlined below:

Contribution: Three novel ML and DL -based classifiers were created and evaluated, including the hybrid ML model along with stacking ensemble, Hybrid DL model (CNN-BiLSTM) model, and the Hybrid ML+ DL model with stacking ensemble. These models were rigorously tested using real-world credit card transaction dataset to assess their effectiveness in cyber fraud detection. The novel hybrid ML model embodies a comprehensive fusion of decision trees (DT), random forest (RF), support vector machine (SVM), logistic regression (LR), XGBoost, and CatBoost with ensemble learning techniques. This approach is particularly effective in improving the accuracy of cyber fraud detection by leveraging the combined capabilities of multiple base models. Through meticulous feature selection and ensemble stacking, the novel hybrid ML Approach demonstrated robust performance in distinguishing between fraudulent and non-fraudulent transactions. A novel DL model-based CNN-BiLSTM model represents a paradigm shift in cyber fraud detection methodologies, leveraging the intrinsic capabilities of convolutional neural networks (CNN) and bidirectional long short-term memory (BiLSTM) networks. This model epitomises sophistication and captures spatial and temporal patterns in credit card transaction data. By seamlessly integrating feature extraction and sequential analysis, the CNN-BiLSTM model achieves unparalleled accuracy in identifying fraudulent transactions, thus fortifying the resilience of cyber fraud detection systems against evolving threats. The hybrid stacking ML+DL signifies a groundbreaking integration of ML and DL techniques, employing stacking ensemble and sampling methods to increase classification accuracy. By synergistically combining the outputs of ML classifiers with DL architectures, this model exhibited exceptional performance in real-world dataset testing. Surpassing conventional baseline models and state-of-the-art techniques, the ensemble hybrid ML and DL models embody the pinnacle of the cyber fraud detection process, paving the way for enhanced security in financial transactions.

Comparing the performance of the three created models, it is evident that each brings its own strengths to the table. The hybrid ML model, shown in Chapter 4, impresses its high precision, suggesting its ability to correctly classify relevant instances while minimising false positives. However, its recall rate lagged that of the other models, indicating a potential limitation in capturing all relevant instances within the dataset. Conversely, the DL model featuring the CNN-BiLSTM architecture presented in Chapter 5 demonstrates a balance between precision and recall, showcasing its capability to effectively identify relevant instances while maintaining a lower false positive rate. Nonetheless, its precision falls short compared to that of the hybrid ML model, suggesting a slightly higher tendency for false positives.

The combination of ML and DL techniques in the hybrid stacking ML + DL model, as illustrated in Chapter 6, offers a compelling synthesis of strengths from both domains. With high precision and recall rates, this hybrid stacking ML + DL model shows robust performance in correctly classifying relevant instances, while minimising both false positives and false negatives. Its precision surpasses that of the CNN-BiLSTM model, indicating an improvement in reducing false positives, whereas its recall exceeds that of the hybrid ML model, implying enhanced coverage of relevant instances within the dataset. Thus, the hybrid stacking ML+DL model has emerged as the most promising candidate, leveraging the complementary strengths of ML and DL to achieve superior classification accuracy and reliability.

In summary, the novel hybrid stacking ML+DL model stands out as the most balanced and robust among the three, particularly with an F1 score of 90.11% without sampling and 94.63% with sampling. This high F1 score reflects a harmonious balance between precision and recall, indicating that the model not only accurately identifies a high proportion of true positives, but also effectively minimises false positives. Such a balance is crucial in many practical applications, where both the identification of true cases and the avoidance of false alarms are equally important. Therefore, the superior performance of the novel hybrid stacking ML+DL model across multiple metrics, especially its F1 score, highlights its potential as a highly effective tool for tackling class imbalance in complex datasets.

In conclusion, the hybrid stacking ML+DL models presented in this research offer promising prospects for enhancing cyber fraud detection capabilities. By leveraging automated cyber fraud detection models, financial institutions can bolster their proficiency in detecting and mitigating fraudulent activities, thereby safeguarding the financial ecosystem and ensuring a secure environment for businesses and consumers. Furthermore, the application of artificial intelligence in cyber fraud detection holds immense potential for streamlining diagnosis processes, reducing operational costs, and improving overall efficiency in financial transaction monitoring and analysis.

7.2 Current limitation

The section highlighted the various constraints observed within this project and credit card cyber fraud detection systems, including the following:

1- Limited Dataset Availability: Credit card cyber fraud detection models often face the challenge of scarce datasets being available for training and evaluation. Privacy concerns and proprietary restrictions contribute to the limited access to diverse datasets, potentially hindering the generalisability of the model findings.

2- Addressing Class Imbalances and Overfitting: Mitigating class imbalances and overfitting is crucial for enhancing the model generalisation and robustness. Employing advanced sampling techniques and regularisation methods can alleviate these challenges and ensure optimal performance across diverse datasets.

3- Exploration of Feature Selection and Ensemble Strategies: Novel feature selection techniques and ensemble learning strategies hold promise for optimising cyber fraud detection accuracy. Leveraging advanced algorithms, such as genetic algorithms and swarm intelligence, can facilitate the identification of critical features and enhance model interpretability.

4- Real-time Implementation and Scalability: Investigating the real-time implementation and scalability of the proposed models is imperative for seamless deployment in financial institutions and online payment systems. Embracing cloud-based architectures and distributed computing frameworks can streamline model deployment and easily accommodate dynamic transaction volumes.

7.3 Suggestion for future work

Although this thesis has effectively tackled numerous challenges in credit card cyber fraud detection, there remains a need for further exploration, which may encompass the following:

1- Investigate alternative DL methodologies, such as Generative Adversarial Networks (GANs), Restricted Boltzmann Machines (RBMs), and autoencoders, to enhance cyber fraud detection in credit card transactions.

2- Building upon the promising results obtained from the newly developed combined ML + DL model, there is significant potential for implementing real-time cyber fraud detection systems.

3- Advancing cyber fraud detection algorithms by evaluating them across diverse datasets containing various forms of fraudulent transactions. This approach could improve the efficiency and accuracy of feature extraction, thus enabling broader real-time applications within the financial sector and online payment systems.

4- Comprehensive algorithms can integrate all relevant cyber fraud indicators, including transactional patterns, user behaviours, and contextual information. Incorporating additional contextual data beyond transaction records could further enhance the precision and efficacy of cyber fraud detection systems.

5- Overcoming the challenge of limited access to comprehensive cyber fraud datasets is imperative. Future research should prioritise the validation of models on multiple datasets to ensure consistent performance across various data sources, thus enhancing the generalisability of the proposed methodologies to different conditions and types of fraudulent activities.

6- As the number of dimensions in a dataset increases, the amount of data required to generalize accurately grows exponentially. In the future, the researcher can adopt quantum algorithms to train on large volumes of financial data, leveraging more sophisticated machine learning models. This approach could facilitate the identification of critical features and enhance model interpretability, providing a more efficient way to handle complex datasets and improving the overall performance of financial prediction models.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., & Isard, M. (2016). {TensorFlow}: a system for {Large-Scale} machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI 16).

Abd El-Naby, A., Hemdan, E. E.-D., & El-Sayed, A. (2023). An efficient fraud detection framework with credit card imbalanced data in financial services. Multimedia Tools and Applications, 82(3), 4139-4160. https://doi.org/https://doi.org/10.1007/s11042-022-13434-6

Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113. https://doi.org/https://doi.org/10.1016/j.jnca.2016.04.007

Abdulsalami, B., Kolawole, A., Ogunrinde, M., Lawal, M., Azeez, R., & Afolabi, A. (2019). Comparative analysis of back-propagation neural network and K-means clustering algorithm in fraud detection in online credit card transactions. Fountain Journal of Natural and Applied Sciences, 8(1). https://doi.org/https://doi.org/10.53704/fujnas.v8i1.315.

Adebayo, O. S., Favour-Bethy, T. A., Otasowie, O., & Okunola, O. A. (2023). Comparative Review of Credit Card Fraud Detection using Machine Learning and Concept Drift Techniques. Int. J. Comput. Sci. Mob. Comput, 12, 24-48. https://doi.org/https://doi.org/10.47760/ijcsmc.2023.v12i07.004.

Adityasundar, N., SaiAbhigna, T., & Lakshman, B. (2020). Credit card fraud detection using machine learning classification algorithms over highly imbalanced data. Journal of Science & Technology (JST), 5(3), 138-146. https://doi.org/https://doi.org/10.46243/jst.2020.v5.i3.pp138-146.

Afriyie, J. K., Tawiah, K., Pels, W. A., Addai-Henne, S., Dwamena, H. A., Owiredu, E. O., Ayeh, S. A., & Eshun, J. (2023). A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. Decision Analytics Journal, 6, 100163. https://doi.org/https://doi.org/10.1016/j.dajour.2023.100163.

Aftab, A., Shahzad, I., Anwar, M., Sajid, A., & Anwar, N. (2023). Fraud Detection of Credit Cards Using Supervised Machine Learning. Pakistan Journal of Emerging Science and Technologies (PJEST, 4(3). https://doi.org/https://doi.org/10.58619/pjest.v4i3.114.

Agarwal, A., Iqbal, M., Mitra, B., Kumar, V., & Lal, N. (2021). Hybrid CNN-BILSTM-attention based identification and prevention system for banking transactions. NVEO-NATURAL VOLATILES & ESSENTIAL OILS Journal| NVEO, 2552-2560.

Agarwal, A., Rana, A., Verma, N., & Gupta, K. (2021). Enhancement of classification techniques using principal component analysis and class imbalance handling methods in credit card defaulter detection. International Journal of Forensic Engineering, 5(1), 1-18. https://doi.org/ https://doi.org/10.1504/ijfe.2021.117383

Ahirwar, A., Sharma, N., & Bano, A. (2020). Enhanced SMOTE & fast random forest techniques for credit card fraud detection. Solid State Technology, 63(6), 4721-4733.

Ahmadi, S. (2023). Open AI and its Impact on Fraud Detection in Financial Industry. Sina, A.(2023). Open AI and its Impact on Fraud Detection in Financial Industry. Journal of Knowledge Learning and Science Technology ISSN, 2959-6386. https://doi.org/https://doi.org/10.60087/jklst.vol2.n3.p281

Al-Alawi, A. I., & Al-Bassam, M. S. A. (2020). The significance of cybersecurity system in helping managing risk in banking and financial sector. Journal of Xidian University, 14(7), 1523-1536. https://doi.org/https://doi.org/10.37896/jxu14.7/174

Al-Hashedi, K. G., & Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 40, 100402. https://doi.org/https://doi.org/10.1016/j.cosrev.2021.100402

Al Balawi, S., & Aljohani, N. (2023). Credit-card fraud detection system using neural networks. Int. Arab J. Inf. Technol., 20(2), 234-241. https://doi.org/https://doi.org/10.34028/iajit/20/2/10

Al Rubaie, E. M. H. (2021). Improvement in credit card fraud detection using ensemble classification technique and user data. International Journal of Nonlinear Analysis and Applications, 12(2), 1255-1265. https://doi.org/10.22075/IJNAA.2021.5228

Ala'raj, M., Abbod, M. F., & Majdalawieh, M. (2021). Modelling customers credit card behaviour using bidirectional LSTM neural networks. Journal of Big Data, 8(1), 69. https://doi.org/https://doi.org/10.1186/s40537-021-00461-7.

Alam, T. M., Shaukat, K., Hameed, I. A., Luo, S., Sarwar, M. U., Shabbir, S., Li, J., & Khushi, M. (2020). An investigation of credit card default prediction in the imbalanced datasets. IEEE Access, 8, 201173-201198. https://doi.org/https://doi.org/10.1109/access.2020.3033784.

Alamri, M., & Ykhlef, M. (2024). Hybrid Undersampling and Oversampling for Handling Imbalanced Credit Card Data. IEEE Access. https://doi.org/https://doi.org/10.1109/access.2024.3357091.

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M., & Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms. IEEE Access, 10, 39700-39715. https://doi.org/https://doi.org/10.1109/access.2022.3166891.

Aleesa, A., Zaidan, B., Zaidan, A., & Sahar, N. M. (2020). Review of intrusion detection systems based on deep learning techniques: coherent taxonomy, challenges, motivations, recommendations, substantial analysis and future directions. Neural Computing and Applications, 32, 9827-9858. https://doi.org/https://doi.org/10.1007/s00521-019-04557-3.

Alenzi, H. Z., & Aljehane, N. O. (2020). Fraud detection in credit cards using logistic regression. International Journal of Advanced Computer Science and Applications, 11(12). https://doi.org/https://doi.org/10.14569/ijacsa.2020.0111265.

Alghofaili, Y., Albattah, A., & Rassam, M. A. (2020). A financial fraud detection model based on LSTM deep learning technique. Journal of Applied Security Research, 15(4), 498-516. https://doi.org/https://doi.org/10.1080/19361610.2020.1815491.

Alhowaide, A., Alsmadi, I., & Tang, J. (2020). PCA, Random-forest and pearson correlation for dimensionality reduction in IoT IDS. 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS).

Alias, M. S. A., Ibrahim, N., & Zin, Z. M. (2019). Comparative study of machine learning algorithms and correlation between input parameters. International Journal of Integrated Engineering, 11(4).

Almarshad, F. A., Gashgari, G. A., & Alzahrani, A. I. (2023). Generative Adversarial Networks-Based Novel Approach for Fraud Detection for the European Cardholders 2013 Dataset. IEEE Access. https://doi.org/https://doi.org/10.1109/access.2023.3320072.

Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232. https://doi.org/https://doi.org/10.3390/electronics12010232.

Amusan, E., Alade, O., Fenwa, O., & Emuoyibofarhe, J. (2021). Credit card fraud detection on skewed data using machine learning techniques. Lautech Journal of Computing and Informatics, 2(1), 49-56.

Arfeen, A. A., & Khan, B. M. A. (2023). Empirical analysis of machine learning algorithms on detection of fraudulent electronic fund transfer transactions. IETE Journal of Research, 69(11), 7920-7932. https://doi.org/https://doi.org/10.1080/03772063.2022.2048700.

Arun, G. K., & Venkatachalapathy, K. (2020). Intelligent feature selection with social spider optimization based artificial neural network model for credit card fraud detection. IIOABJ, 11(2), 85-91.

Arya, M., & Sastry G, H. (2020). DEAL–'Deep Ensemble ALgorithm'framework for credit card fraud detection in real-time data stream with Google TensorFlow. Smart Science, 8(2), 71-83. https://doi.org/https://doi.org/10.1080/23080477.2020.1783491.

Asaad, R. R., & Saeed, V. A. (2022). A Cyber Security Threats, Vulnerability, Challenges and Proposed Solution. Applied computing Journal, 227-244. https://doi.org/https://doi.org/10.52098/acj.202260.

Aschi, M., Bonura, S., Masi, N., Messina, D., & Profeta, D. (2022). Cybersecurity and fraud detection in financial transactions. In Big data and artificial intelligence in digital finance: Increasing personalization and trust in digital finance using big data and AI (pp. 269-278). Springer. https://doi.org/https://doi.org/10.1007/978-3-030-94590-9_15.

Asha, R., & KR, S. K. (2021). Credit card fraud detection using artificial neural network. Global Transitions Proceedings, 2(1), 35-41. https://doi.org/https://doi.org/10.1016/j.gltp.2021.01.006.

Askari, S. M. S., & Hussain, M. A. (2020). IFDTC4. 5: Intuitionistic fuzzy logic based decision tree for E-transactional fraud detection. Journal of Information Security and Applications, 52, 102469. https://doi.org/https://doi.org/10.1016/j.jisa.2020.102469.

Aswathy, M., & Samuel, L. (2019). Credit card fraud detection using hybrid models. International Research Journal of Engineering and Technology, 6, 2019.

Ata, O., & Hazim, L. (2020). Comparative analysis of different distributions dataset by using data mining techniques on credit card fraud detection. Tehnički vjesnik, 27(2), 618-626. https://doi.org/https://doi.org/10.17559/tv-20180427091048.

Azam, Z., Islam, M. M., & Huda, M. N. (2023). Comparative analysis of intrusion detection systems and machine learning based model analysis through decision tree. IEEE Access. https://doi.org/ https://doi.org/10.1109/access.2023.3296444.

Bagga, S., Goyal, A., Gupta, N., & Goyal, A. (2020). Credit card fraud detection using pipeling and ensemble learning. Procedia computer science, 173, 104-112. https://doi.org/ https://doi.org/10.1016/j.procs.2020.06.014.

Bandyopadhyay, S., Thakkar, V., Mukherjee, U., & Dutta, S. (2021). Emerging approach for detection of financial frauds using machine learning. https://doi.org/https://doi.org/10.20944/preprints202108.0028.v1.

Bandyopadhyay, S. K., & Dutta, S. (2020). Detection of fraud transactions using recurrent neural network during COVID-19: fraud transaction during COVID-19. Journal of Advanced Research in Medical Science & Technology (ISSN: 2394-6539), 7(3), 16-21. https://doi.org/https://doi.org/10.24321/2394.6539.202012.

Bansal, D., Bhatia, M., Atrey, A., & Yadav, A. K. (2024). Perspective of Cybersecurity and Ethical Hacking with Vulnerability Assessment and Exploitation Tools. In Big Data Analytics Framework for Smart Grids (pp. 98-111). CRC Press. https://doi.org/https://doi.org/10.1201/9781032665399-6.

Barahim, A., Alhajri, A., Alasaibia, N., Altamimi, N., Aslam, N., & Khan, I. U. (2019). Enhancing the credit card fraud detection through ensemble techniques. Journal of Computational and Theoretical Nanoscience, 16(11), 4461-4468. https://doi.org/https://doi.org/10.1166/jctn.2019.8619.

Benchaji, I., Douzi, S., & El Ouahidi, B. (2021). Credit card fraud detection model based on LSTM recurrent neural networks. Journal of Advances in Information Technology, 12(2). https://doi.org/https://doi.org/10.12720/jait.12.2.113-118.

Bengio, Y., & Grandvalet, Y. (2003). No unbiased estimator of the variance of k-fold cross-validation. Advances in Neural Information Processing Systems, 16.

Berhane, T., Melese, T., Walelign, A., & Mohammed, A. (2023). A Hybrid Convolutional Neural Network and Support Vector Machine-Based Credit Card Fraud Detection Model. Mathematical Problems in Engineering, 2023(1), 8134627. https://doi.org/ https://doi.org/10.1155/2023/8134627.

Bhowmik, A., Sannigrahi, M., Chowdhury, D., Dwivedi, A. D., & Mukkamala, R. R. (2022). Dbnex: Deep belief network and explainable ai based financial fraud detection. 2022 IEEE International Conference on Big Data (Big Data).

Borse, D. D., Patil, S. H., & Dhotre, S. (2021). Credit card fraud detection using naive Bayes and robust scaling techniques. International Journal, 10(1), 1-5. https://doi.org/https://doi.org/10.30534/ijatcse/2021/311012021.

Btoush, E., Zhou, X., Gururaian, R., Chan, K. C., & Tao, X. (2021). A survey on credit card fraud detection techniques in banking industry for cyber security. 2021 8th International Conference on Behavioral and Social Computing (BESC).

Btoush, E. A. L. M., Zhou, X., Gururajan, R., Chan, K. C., Genrich, R., & Sankaran, P. (2023). A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Computer Science, 9, e1278. https://doi.org/https://doi.org/10.7717/peerj-cs.1278.

Bwalya, D., & , & Phiri, J. (2023). Fraud Detection in Mobile Banking Based on Artificial Intelligence. In Computer Science On-line Conference (537–554). https://doi.org/https://doi.org/10.1007/978-3-031-35314-7_48.

Can, B., Yavuz, A. G., Karsligil, E. M., & Guvensan, M. A. (2020). A closer look into the characteristics of fraudulent card transactions. IEEE Access, 8, 166095-166109. https://doi.org/https://doi.org/10.1109/access.2020.3022315.

Carbo-Valverde, S., Cuadros-Solas, P., & Rodríguez-Fernández, F. (2020). A machine learning approach to the digitalization of bank customers: Evidence from random and causal forests. Plos one, 15(10), e0240362. https://doi.org/https://doi.org/10.1371/journal.pone.0240362.

Carcillo, F., Le Borgne, Y.-A., Caelen, O., Kessaci, Y., Oblé, F., & Bontempi, G. (2021). Combining unsupervised and supervised learning in credit card fraud detection. Information sciences, 557, 317-331. https://doi.org/https://doi.org/10.1016/j.ins.2019.05.042.

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020). A comprehensive survey on support vector machine classification: Applications, challenges and trends. Neurocomputing, 408, 189-215. https://doi.org/https://doi.org/10.1016/j.neucom.2019.10.118.

Checkland, P., & Holwell, S. (1998). Information, systems and information systems: making sense of the field. John Wiley & Sons, Inc. https://doi.org/https://doi.org/10.2307/3010006.

Chen, C.-T., Lee, C., Huang, S.-H., & Peng, W.-C. (2024). Credit Card Fraud Detection via Intelligent Sampling and Self-supervised Learning. ACM Transactions on Intelligent Systems and Technology. https://doi.org/https://doi.org/10.1145/3641283.

Chen, J. I.-Z., & Lai, K.-L. (2021). Deep convolution neural network model for credit-card fraud detection and alert. Journal of Artificial Intelligence, 3(02), 101-112. https://doi.org/https://doi.org/10.36548/jaicn.2021.2.003.

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2020). Graph neural network for fraud detection via spatial-temporal attention. IEEE Transactions on Knowledge and Data Engineering, 34(8), 3800-3813. https://doi.org/https://doi.org/10.1109/access.2021.3074243.

Cheon, M.-j., Lee, D., Joo, H. S., & Lee, O. (2021). Deep learning based hybrid approach of detecting fraudulent transactions. Journal of Theoretical and Applied Information Technology, 99(16), 4044-4054. https://scholarworks.bwise.kr/hanyang/handle/2021.sw.hanyang/141261.

Cherif, A., Badhib, A., Ammar, H., Alshehri, S., Kalkatawi, M., & Imine, A. (2023). Credit card fraud detection in the era of disruptive technologies: A systematic review. Journal of King Saud University-Computer and Information Sciences, 35(1), 145-174. https://doi.org/https://doi.org/10.1016/j.jksuci.2022.11.008.

Choubey, R., & Gautam, P. (2020). Combined technique of supervised classifier for the credit card fraud detection. Shodah Sarita, 7, 27-32.

Chowdari, G. B., & Chowdari, G. B. (2021). Supervised machine learning algorithms for detecting credit card fraud. EPRA International Journal of Research & Development, 6(7). https://doi.org/https://doi.org/10.36713/epra7636.

Creado, Y., & Ramteke, V. (2020). Active cyber defence strategies and techniques for banks and financial institutions. Journal of Financial Crime, 27(3), 771-780. https://doi.org/https://doi.org/10.1108/jfc-01-2020-0008.

Daliri, S. (2020). Using harmony search algorithm in neural networks to improve fraud detection in banking system. Computational Intelligence and Neuroscience, 2020(1), 6503459. https://doi.org/ https://doi.org/10.1155/2020/6503459.

Dang, T. K., Tran, T. C., Tuan, L. M., & Tiep, M. V. (2021). Machine learning based on resampling approaches and deep reinforcement learning for credit card fraud detection systems. Applied Sciences, 11(21), 10004. https://doi.org/ https://doi.org/10.3390/app112110004.

Darem, A. A., Alhashmi, A. A., Alkhaldi, T. M., Alashjaee, A. M., Alanazi, S. M., & Ebad, S. A. (2023). Cyber threats classifications and countermeasures in banking and financial sector. IEEE Access, 11, 125138-125158. https://doi.org/https://doi.org/10.1109/access.2023.3327016.

Das, T. A., Lagade, K. C., Girase, M. P., & Patole, R. (2020). Credit card fraud detection system using data mining. Artificial & Computational Intelligence, 3(3), 2020.
DEB, K., Ghosal, S., & Bose, D. (2021). A comparative study on credit card fraud detection. https://doi.org/https://doi.org/10.31224/osf.io/8ctxd.

DeepaShree, N., Lakshmi, S. V., Alagundagi, T., Bhumika, R., & Myageri, S. (2019). Supervised machine learning algorithms for credit card fraudulent transaction detection. International Journal of Research in Engineering, Science and Management, 2(6), 2019.

Deepika, S., & Senthil, S. (2019). Credit card fraud analysis using robust space invariant artificial neural networks (RSIANN). International Journal of Recent Technology and Engineering (IJRTE), 8(2), 2277-3878. https://doi.org/https://doi.org/10.35940/ijrte.b2315.078219.

Ding, Y., Kang, W., Feng, J., Peng, B., & Yang, A. (2023). Credit card fraud detection based on improved Variational Autoencoder Generative Adversarial Network. IEEE Access. https://doi.org/https://doi.org/10.1109/access.2023.3302339.

Divakar, K., & Chitharanjan, K. (2019). Performance evaluation of credit card fraud transactions using boosting algorithms. Int. J. Electron. Commun. Comput. Eng. IJECCE, 10(6), 262-270.

Diwan, T. D. (2021). An investigation and analysis of cyber security information systems: latest trends and future suggestion. Information Technology in Industry, 9(2), 477-492. https://doi.org/https://doi.org/10.17762/itii.v9i2.372.

Douzas, G., Bacao, F., & Last, F. (2018). Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE. Information sciences, 465, 1-20. https://doi.org/https://doi.org/10.1016/j.ins.2018.06.056.

Dwivedi, A. K. (2021). Fraud detection in credit card transactions using anomaly detection. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 12(12), 837-846.

Dzakiyullah, N. R., Pramuntadi, A., & Fauziyyah, A. K. (2021). Semi-supervised classification on credit card fraud detection using autoencoders. Journal of Applied Data Sciences, 2(1), 01-07. https://doi.org/https://doi.org/10.47738/jads.v2i1.16.

El Hlouli, F. Z., Riffi, J., Mahraz, M. A., Yahyaouy, A., El Fazazy, K., & Tairi, H. (2024). Credit Card Fraud Detection: Addressing Imbalanced Datasets with a Multi-phase Approach. SN Computer Science, 5(1), 173. https://doi.org/https://doi.org/10.1007/s42979-023-02559-6.

Elluri, L., Mandalapu, V., Vyas, P., & Roy, N. (2023). Recent Advancements in Machine Learning for Cybercrime Prediction. Journal of Computer Information Systems, 1-15. https://doi.org/ https://doi.org/10.1080/08874417.2023.2270457.

Elreedy, D., Atiya, A. F., & Kamalov, F. (2023). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. Machine Learning, 1-21. https://doi.org/https://doi.org/10.1007/s10994-022-06296-4.

Esenogho, E., Mienye, I. D., Swart, T. G., Aruleba, K., & Obaido, G. (2022). A neural network ensemble with feature engineering for improved credit card fraud detection. IEEE Access, 10, 16400-16407. https://doi.org/https://doi.org/10.1109/access.2022.3148298.

Fakiha, B. (2023). Forensic Credit Card Fraud Detection Using Deep Neural Network. Journal of Southwest Jiaotong University, 58(1). https://doi.org/https://doi.org/10.35741/issn.0258-2724.58.1.33.

Fang, W., Li, X., Zhou, P., Yan, J., Jiang, D., & Zhou, T. (2021). Deep learning anti-fraud model for internet loan: Where we are going. IEEE Access, 9, 9777-9784. https://doi.org/10.1109/ACCESS.2021.3051079.

Faraj, A. A., Mahmud, D. A., & Rashid, B. N. (2021). Comparison of different ensemble methods in credit card default prediction. UHD Journal of Science and Technology, 5(2), 20-25.

Faridpour, M., & Moradi, A. (2020). A novel method for detection of fraudulent bank transactions using multi-layer neural networks with adaptive learning rate. International Journal of Nonlinear Analysis and Applications, 11(2), 437-445. https://doi.org/10.22075/IJNAA.2020.4576.

Forough, J., & Momtazi, S. (2021). Ensemble of deep sequential models for credit card fraud detection. Applied Soft Computing, 99, 106883. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106883.

Gable, G. G. (1994). Integrating case study and survey research methods: an example in information systems. European journal of information systems, 3, 112-126. https://doi.org/https://doi.org/https://doi.org/10.1057/ejis.1994.1.

Geoffrey, M. (2019). Esential of research design and metodology (42).

Gewers, F. L., Ferreira, G. R., Arruda, H. F. D., Silva, F. N., Comin, C. H., Amancio, D. R., & Costa, L. d. F. (2021). Principal component analysis: A natural approach to data exploration. ACM Computing Surveys (CSUR), 54(4), 1-34. https://doi.org/https://doi.org/10.1145/3447755.

Ghaleb, F. A., Saeed, F., Al-Sarem, M., Qasem, S. N., & Al-Hadhrami, T. (2023). Ensemble Synthesized Minority Oversampling based Generative Adversarial Networks and Random Forest Algorithm for Credit Card Fraud Detection. IEEE Access. https://doi.org/https://doi.org/10.1109/access.2023.3306621.

Gupta, A., Lohani, M., & Manchanda, M. (2021). Financial fraud detection using naive bayes algorithm in highly imbalance data set. Journal of Discrete Mathematical Sciences and Cryptography, 24(5), 1559-1572. https://doi.org/https://doi.org/10.1080/09720529.2021.1969733.

Habibpour, M., Gharoun, H., Mehdipour, M., Tajally, A., Asgharnezhad, H., Shamsi, A., Khosravi, A., & Nahavandi, S. (2023). Uncertainty-aware credit card fraud detection using deep learning. Engineering Applications of Artificial Intelligence, 123, 106248. https://doi.org/https://doi.org/10.1016/j.engappai.2023.106248.

Hammed, M., & Soyemi, J. (2020). An implementation of decision tree algorithm augmented with regression analysis for fraud detection in credit card. International Journal of Computer Science and Information Security (IJCSIS), 18(2), 79-88.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36. https://doi.org/https://doi.org/10.1148/radiology.143.1.7063747.

Harikrishna, J., Rupa, C., & Gireesh, R. (2022). Deep learning-based real-time object classification and recognition using supervised learning approach. Sentimental Analysis and Deep Learning: Proceedings of ICSADL 2021.

Harwani, H., Jain, J., Jadhav, C., & Hodavdekar, M. (2020). Credit card fraud detection technique using hybrid approach: an amalgamation of self organizing maps and neural networks. International Research Journal of Engineering and Technology (IRJET), 7(2020).

Hassan, M., Aziz, L. A.-R., & Andriansyah, Y. (2023). The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Reviews of Contemporary Business Analytics, 6(1), 110-132.

Hema, A., & Muttipati, A. (2020). Machine learning methods for discovering credit card fraud. International Research Journal of Computer Science, 8(1), 1-6.

Hilal, W., Gadsden, S. A., & Yawney, J. (2022). Financial fraud: a review of anomaly detection techniques and recent advances. Expert systems With applications, 193, 116429. https://doi.org/https://doi.org/10.1016/j.eswa.2021.116429.

Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. Computing in science & engineering, 9(03), 90-95. https://doi.org/https://doi.org/10.1109/MCSE.2007.55
Husejinovic, A. (2020). Credit card fraud detection using naive Bayesian and c4. 5 decision tree classifiers. Husejinovic, A.(2020). Credit card fraud detection using naive Bayesian and C, 4, 1-5.

Hussein, N. K., Abbas, A. R., & Mahdi, B. S. (2021). Fraud classification and detection model using different machine learning algorithm. Tech-Knowledge, 1(1), 13-22.

Ileberi, E., Sun, Y., & Wang, Z. (2021). Performance evaluation of machine learning methods for credit card fraud detection using SMOTE and AdaBoost. IEEE Access, 9, 165286-165294. https://doi.org/https://doi.org/10.1109/access.2021.3134330.

Islam, M. A., Uddin, M. A., Aryal, S., & Stea, G. (2023). An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. Journal of Information Security and Applications, 78, 103618. https://doi.org/https://doi.org/10.1016/j.jisa.2023.103618.

Itoo, F., Meenakshi, & Singh, S. (2021). Comparison and analysis of logistic regression, Naïve Bayes and KNN machine learning algorithms for credit card fraud detection. International Journal of Information Technology, 13(4), 1503-1511. https://doi.org/https://doi.org/10.1007/s41870-020-00430-y.

Jaiswal, S., Brindha, R., & Lakhotia, S. (2021). Credit card fraud detection using isolation forest and local outlier factor. Annals of the Romanian Society for Cell Biology, 4391-4396. https://doi.org/https://doi.org/10.55041/ijsrem14371.

Janapareddy, D., & Yenduri, N. C. (2023). Credit Card Approval Prediction: A comparative analysis between logistic regressionclassifier, random forest classifier, support vectorclassifier with ensemble bagging classifier.

Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. Electronic Markets, 31(3), 685-695. https://doi.org/https://doi.org/10.1007/s12525-021-00475-2.

Järvinen, P. (2005). Action research as an approach in design science.

Javaid, N., Akbar, M., Aldegheishem, A., Alrajeh, N., & Mohammed, E. A. (2022). Employing a machine learning boosting classifiers based stacking ensemble model for detecting non technical losses in smart grids. IEEE Access, 10, 121886-121899. https://doi.org/ https://doi.org/10.1109/access.2022.3222883.

Jonnalagadda, V., Gupta, P., & Sen, E. (2019). Credit card fraud detection using Random Forest Algorithm. International Journal of Advance Research, Ideas and Innovations in Technology, 5(2), 1-5.

Kalid, S. N., Ng, K.-H., Tong, G.-K., & Khor, K.-C. (2020). A multiple classifiers system for anomaly detection in credit card data with unbalanced and overlapped classes. IEEE Access, 8, 28210-28221. https://doi.org/https://doi.org/10.1109/access.2020.2972009.

Karthik, R., Navinkumar, R., Rammkumar, U., & Mothilal, K. (2019). Supervised machine learning algorithms for credit card fraudulent transaction detection. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 2019, 2456-3307. https://doi.org/https://doi.org/10.32628/cseit195274.

Karthik, V., Mishra, A., & Reddy, U. S. (2022). Credit card fraud detection by modelling behaviour pattern using hybrid ensemble model. Arabian Journal for Science and Engineering, 47(2), 1987-1997. https://doi.org/ https://doi.org/10.1007/s13369-021-06147-9.

Karthika, J., & Senthilselvi, A. (2023). Smart credit card fraud detection system based on dilated convolutional neural network with sampling technique. Multimedia Tools and Applications, 82(20), 31691-31708. https://doi.org/https://doi.org/10.1007/s11042-023-15730-1.

Kasasbeh, B., Aldabaybah, B., & Ahmad, H. (2022). Multilayer perceptron artificial neural networks-based model for credit card fraud detection. Indonesian Journal of Electrical Engineering and Computer Science, 26(1), 362-373. https://doi.org/https://doi.org/10.11591/ijeecs.v26.i1.pp362-373.

Kazdin, A. E. (2016). Methodology: What it is and why it is so important. https://doi.org/https://doi.org/10.1037/14805-001.

Ketkar, N., Moolayil, J., Ketkar, N., & Moolayil, J. (2020). Deep Learning with Python: Learn Best Practices of Deep Learning Models with PyTorch. Apress LP. https://doi.org/https://doi.org/10.1007/978-1-4842-5364-9_6.

Khalid, A. R., Owoh, N., Uthmani, O., Ashawa, M., Osamor, J., & Adejoh, J. (2024). Enhancing credit card fraud detection: an ensemble machine learning approach. Big Data and Cognitive Computing, 8(1), 6. https://doi.org/https://doi.org/10.3390/bdcc8010006.

Khan, S., Alourani, A., Mishra, B., Ali, A., & Kamal, M. (2022). Developing a credit card fraud detection model using machine learning approaches. International Journal of Advanced Computer Science and Applications, 13(3). https://doi.org/https://doi.org/10.14569/ijacsa.2022.0130350.

Kırelli, Y., Arslankaya, S., & Zeren, M. T. (2020). Detection of credit card fraud in e-commerce using data mining. Avrupa Bilim ve Teknoloji Dergisi(20), 522-529. https://doi.org/https://doi.org/10.31590/ejosat.747399.

Klein, H. K., & Lyytinen, K. (1985). The poverty of scientism in information systems. Research methods in information systems, 131-161.

Kraiem, M. S., Sánchez-Hernández, F., & Moreno-García, M. N. (2021). Selecting the suitable resampling strategy for imbalanced data classification regarding dataset properties. An approach based on association models. Applied Sciences, 11(18), 8546. https://doi.org/ https://doi.org/10.3390/app11188546.

Krawczyk, B. (2016). Learning from imbalanced data: open challenges and future directions. Progress in Artificial Intelligence, 5(4), 221-232. https://doi.org/https://doi.org/10.1007/s13748-016-0094-0.

Krichen, M. (2023). Convolutional neural networks: A survey. Computers, 12(8), 151. https://doi.org/https://doi.org/10.3390/computers12080151.

Krishna, K. G., Kulkarni, P., & Natraj, N. (2023). Use of Big Data Technologies for Credit Card Fraud Prediction. 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS).

Kumar, M. D., Mubarak, A., & Dhanush, M. (2020). Credit card fraud detection using Bayesian belief network. International Journal of Research in Engineering, Science and Management, 3(7), 316-319. https://journal.ijresm.com/index.php/ijresm/article/view/86.

Kumar, R., Student, P., & Budihul, R. (2020). An efficient approach for credit card fraud detection. International Journal of Innovative Science and Research Technology, 5(4), 2020.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444. https://doi.org/https://doi.org/10.1038/nature14539.

Leedy, P., & Ormrod, J. (2005). Practical research: Pearson Custom. Sl: Pearson.

Leevy, J. L., Johnson, J. M., Hancock, J., & Khoshgoftaar, T. M. (2023). Threshold optimization and random undersampling for imbalanced credit card data. Journal of Big Data, 10(1), 58. https://doi.org/https://doi.org/10.1186/s40537-023-00738-z.

Li, C., Ding, N., Zhai, Y., & Dong, H. (2021). Comparative study on credit card fraud detection based on different support vector machines. Intelligent Data Analysis, 25(1), 105-119. https://doi.org/10.3233/IDA-195011.

Li, Z., Zhang, H., Masum, M., Shahriar, H., & Haddad, H. (2020). Cyber fraud prediction with supervised machine learning techniques. Proceedings of the 2020 ACM Southeast Conference.

Lim, K. S., Lee, L. H., & Sim, Y.-W. (2021). A review of machine learning algorithms for fraud detection in credit card transaction. International Journal of Computer Science & Network Security, 21(9), 31-40. https://doi.org/https://doi.org/10.22937/IJCSNS.2021.21.9.4.

Lin, T.-H., & Jiang, J.-R. (2021). Credit card fraud detection with autoencoder and probabilistic random forest. Mathematics, 9(21), 2683.

Maheshwari, V. C., Osman, N. A., & Aziz, N. (2023). A Hybrid Approach Adopted for Credit Card Fraud Detection Based on Deep Neural Networks and Attention Mechanism. Journal of Advanced Research in Applied Sciences and Engineering Technology, 32(1), 315-331. https://doi.org/https://doi.org/10.37934/araset.32.1.315331.

Makolo, A., & Adeboye, T. (2021). Credit card fraud detection system using machine learning. International Journal of Information Technology and Computer Science, 4, 24-37. https://doi.org/https://doi.org/10.5815/ijitcs.2021.04.03.

Manlangit, S., Azam, S., & Shanmugam, B. (2019). Novel machine learning approach for analyzing anonymous credit card fraud patterns. International Journal of Electronic Commerce Studies, 10(2), 175-202. https://doi.org/ https://doi.org/10.7903/ijecs.1732.

Mansourifar, H., & Shi, W. (2020). Deep synthetic minority over-sampling technique. arXiv preprint arXiv:2003.09788.

McKinsey, & Company. (2022, October 27). New survey reveals $2 trillion market opportunity for cybersecurity technology and service providers. Retrieved 02-02 from https://www.mckinsey.com/capabilities/risk-and-resilience/our-insights/cybersecurity/new-survey-reveals-2-trillion-dollar-market-opportunity-for-cybersecurity-technology-and-service-providers.

Meenakshi, B. D., Janani, B., Gayathri, S., & Indira, N. (2019). Credit card fraud detection using random forest. International Research Journal of Engineering and Technology (IRJET), 6(3), 2019.

Merchantcostconsulting. (2024). Credit Card Fraud Statistics 21- Merchantcostconsulting. Retrieved 12-12-2023 from https://merchantcostconsulting.com/lower-credit-card-processing-fees/credit-card-fraud-statistics.

Metz, C. E. (1978). Basic principles of ROC analysis. Seminars in nuclear medicine,
Mienye, I. D., & Sun, Y. (2023). A deep learning ensemble with data resampling for credit card fraud detection. IEEE Access, 11, 30628-30638. https://doi.org/https://doi.org/10.1109/access.2023.3262020.

Mijwil, M. M., & Salem, I. E. (2020). Credit card fraud detection in payment using machine learning classifiers. Asian Journal of Computer and Information Systems (ISSN: 2321–5658), 8(4). https://doi.org/https://doi.org/10.24203/ajcis.v8i4.6449.

Mim, M. A., Majadi, N., & Mazumder, P. (2024). A soft voting ensemble learning approach for credit card fraud detection. Heliyon, 10(3). https://doi.org/https://doi.org/10.1016/j.heliyon.2024.e25466.

Misra, S., Thakur, S., Ghosh, M., & Saha, S. K. (2020). An autoencoder based model for detecting fraudulent credit card transaction. Procedia computer science, 167, 254-262. https://doi.org/https://doi.org/10.1016/j.procs.2020.03.219.

Mittal, S., & Tyagi, S. (2020). Computational techniques for real-time credit card fraud detection. Handbook of Computer Networks and Cyber Security: Principles and Paradigms, 653-681. https://doi.org/https://doi.org/10.1007/978-3-030-22277-2_26.

Mohanty, S., Sharma, S., Pattnaik, P. K., & Hol, A. (2023). A comprehensive review on cyber security and online banking security frameworks. Risk Detection and Cyber Security for the Success of Contemporary Computing, 1-22. https://doi.org/https://doi.org/10.4018/978-1-6684-9317-5.ch001.

Muaz, A., Jayabalan, M., & Thiruchelvam, V. (2020). A comparison of data sampling techniques for credit card fraud detection. International Journal of Advanced Computer Science and Applications, 11(6). https://doi.org/https://doi.org/10.14569/ijacsa.2020.0110660.

Mytnyk, B., Tkachyk, O., Shakhovska, N., Fedushko, S., & Syerov, Y. (2023). Application of artificial intelligence for fraudulent banking operations recognition. Big Data and Cognitive Computing, 7(2), 93. https://doi.org/https://doi.org/10.3390/bdcc7020093.

Nalayini, C., Katiravan, J., Sathyabama, A., Rajasuganya, P., & Abirami, K. (2022). Identification and Detection of Credit Card Frauds Using CNN. International Conference on Computers, Management & Mathematical Sciences.

Nama, F. A., Obaid, A.J. and Alrammahi, A.A.H. (2023). Credit Card Fraud Detection and Classification Using Deep Learning with Support Vector Machine Techniques Swaroop, A., Polkowski, Z., Correia, S.D., Virdee, B. (eds) Proceedings of Data Analytics and Management.

Narayan, V., & Ganapathisamy, S. (2022). Hybrid Sampling and Similarity Attention Layer in Bidirectional Long Short Term Memory in Credit Card Fraud Detection. International Journal of Intelligent Engineering & Systems, 15(6). https://doi.org/https://doi.org/10.22266/ijies2022.1231.04.

Nguyen, T. T., Tahir, H., Abdelrazek, M., & Babar, A. (2020). Deep learning methods for credit card fraud detection. arXiv preprint arXiv:2012.03754. https://doi.org/ https://doi.org/10.48550/arXiv.2012.03754.

Nwogu, E., & Nwachukwu, E. (2019). An improved hybrid system for the prediction of debit and credit card fraud. Computing, Information Systems & Development Informatics Journal, 10(3). https://doi.org/ https://doi.org/10.22624/AIMS/CISDI/V10N3P8.

Oluwasanya, O., & Braimah Joachim, A. (2023). Credit card fraud detection using logistic regression and isolation forest algorithms. UNIZIK Journal of Engineering and Applied Sciences, 2(1), 187-195.

Omair, B., & Alturki, A. (2020). A systematic literature review of fraud detection metrics in business processes. IEEE Access, 8, 26893-26903. https://doi.org/https://doi.org/10.1109/access.2020.2971604.

Oumar, A. W., & Augustin, P. (2019). Credit card fraud detection using ANN. International Journal of Innovative Technology and Exploring Engineering, 8(7), 313-316.

Owolafe, O., Ogunrinde, O. B., & Thompson, A. F.-B. (2021). A long short term memory model for credit card fraud detection. In Artificial Intelligence for Cyber Security: Methods, Issues and Possible Horizons or Opportunities (pp. 369-391). Springer. https://doi.org/https://doi.org/10.1007/978-3-030-72236-4_15.

Ozbayoglu, A. M., Gudelek, M. U., & Sezer, O. B. (2020). Deep learning for financial applications: A survey. Applied Soft Computing, 93, 106384. https://doi.org/https://doi.org/10.1016/j.asoc.2020.106384.

Patel, K. (2023). Credit Card Analytics: A Review of Fraud Detection and Risk Assessment Techniques. International Journal of Computer Trends and Technology, 71(10), 69-79. https://doi.org/https://doi.org/10.14445/22312803/ijctt-v71i10p109.

Patil, S., Nemade, V., & Soni, P. K. (2018). Predictive modelling for credit card fraud detection using data analytics. Procedia computer science, 132, 385-395. https://doi.org/https://doi.org/10.1016/j.procs.2018.05.199.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., & Dubourg, V. (2011). Scikit-learn: Machine learning in Python. the Journal of machine Learning research, 12, 2825-2830.

Pereira, R. M., Costa, Y. M., & Silla Jr, C. N. (2020). MLTL: A multi-label approach for the Tomek Link undersampling algorithm. Neurocomputing, 383, 95-105. https://doi.org/https://doi.org/10.1016/j.neucom.2019.11.076.

Perwej, Y., Abbas, S. Q., Dixit, J. P., Akhtar, N., & Jaiswal, A. K. (2021). A systematic literature review on the cyber security. International Journal of scientific research and management, 9(12), 669-710. https://doi.org/https://doi.org/10.18535/ijsrm/v9i12.ec04.

Potula, S. R., Selvanambi, R., Karuppiah, M., & Pelusi, D. (2023). Artificial intelligence-based cyber security applications. In Artificial Intelligence and Cyber Security in Industry 4.0 (pp. 343-373). Springer. https://doi.org/https://doi.org/10.1007/978-981-99-2115-7_16.
Prabhakar, E., Kumar, N., Ponnar, K., Suresh, A., & Jayandhiran, R. (2019). Credit card fraud detection using boosted stacking. South Asian Journal of Engineering and Technology, 8 No. 1 (2019)(1), 149-153.

Priya, G. J. a. S. (2021). Fraud Detection and Prevention Using Machine Learning Algorithms: A Review 7th International Conference on Electrical Energy Systems (ICEES).

Rai, A. K., & Dwivedi, R. K. (2020). Fraud detection in credit card data using machine learning techniques. Machine Learning, Image Processing, Network Security and Data Sciences: Second International Conference, MIND 2020, Silchar, India, July 30-31, 2020, Proceedings, Part II 2.

Ramchoun, H., Idrissi, M. J., Ghanou, Y., & Ettaouil, M. (2017). Multilayer Perceptron: Architecture Optimization and training with mixed activation functions. Proceedings of the 2nd international Conference on Big Data, Cloud and Applications.

Ramisetty, U. M., Gundavarapu, V. N. K., Mishra, A., & Bali, S. K. (2022). Analysis of Fraud Detection Prediction Using Synthetic Minority Over-Sampling Technique. International Conference on Artificial Intelligence Techniques for Electrical Engineering Systems (AITEES 2022).

Reddy, S. T. S., & Sriramya, P. (2023). Comparison of the Support Vector Classifier algorithm with the Decision Tree algorithm for Credit Card Fraud Detection with the Goal of Improving Accuracy. Journal of Survey in Fisheries Sciences, 10(1S), 2304-2313. https://doi.org/https://doi.org/10.17762/sfs.v10i1S.463.

Research, S., & Department. (2023). 'Global cybersecurity spending 2017-2022'. Statista. Retrieved 10-11-2023 from https://www.statista.com/statistics/991304/worldwide-cybersecurity-spending/.

Rijsbergen, V. (1979). Information retrieval; ; Butterworth, 1978. J. librariansh., 11, 237.

Rtayli, N., & Enneya, N. (2020). Enhanced credit card fraud detection based on SVM-recursive feature elimination and hyper-parameters optimization. Journal of Information Security and Applications, 55, 102596. https://doi.org/https://doi.org/10.1016/j.jisa.2020.102596.

Sadgali, I., Sael, N., & Benabbou, F. (2021). Bidirectional gated recurrent unit for improving classification in credit card fraud detection. Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 21(3), 1704-1712. https://doi.org/https://doi.org/10.11591/ijeecs.v21.i3.pp1704-1712.

Salameh, R., & Lutfi, K. (2021). The role of artificial intelligence on limiting Jordanian commercial banks cybercrimes. Accounting, 7(5), 1147-1156. https://doi.org/https://doi.org/10.5267/j.ac.2021.2.024.

Saraf, S., & Phakatkar, A. (2022). Detection of Credit Card Fraud using a Hybrid Ensemble Model. International Journal of Advanced Computer Science and Applications, 13(9), 464-474. https://doi.org/https://doi.org/10.14569/ijacsa.2022.0130953.

Saranya, T., Sridevi, S., Deisy, C., Chung, T. D., & Khan, M. A. (2020). Performance analysis of machine learning algorithms in intrusion detection system: A review. Procedia computer science, 171, 1251-1260. https://doi.org/https://doi.org/10.1016/j.procs.2020.04.133.

Sarker, I. H., Furhad, M. H., & Nowrozy, R. (2021). Ai-driven cybersecurity: an overview, security intelligence modeling and research directions. SN Computer Science, 2(3), 173. https://doi.org/https://doi.org/10.1007/s42979-021-00557-0.

Sarumathi, R., & Saraswathy, G. (2022). Comparative Analysis of Classification Techniques for Credit Card Fraud Detection. History Manuscript Reference No: IRJCS/RS/, 9, 16-25. https://doi.org/ https:/ / doi.org/ 10.26562/ irjcs.2022.v0902.004.

Sasikala, G., Laavanya, M., Sathyasri, B., Supraja, C., Mahalakshmi, V., Mole, S. S., Mulerikkal, J., Chidambaranathan, S., Arvind, C., & Srihari, K. (2022). An innovative sensing machine learning technique to detect credit card frauds in wireless communications. Wireless Communications and Mobile Computing, 2022(1), 2439205. https://doi.org/ https://doi.org/10.1155/2022/2439205.

Sayan, D., Apratim (2020), Comparative Study of Dimensional Reductional Techniques: Principal Compenent Analysis and AutoEncoders. International Research Journal of Computer Science, 7, 291-299. https://doi.org/ https:/ / doi.org/ 10.26562/ irjcs.2020.v0712.001.

Seera, M., Lim, C. P., Kumar, A., Dhamotharan, L., & Tan, K. H. (2024). An intelligent payment card fraud detection system. Annals of operations research, 334(1), 445-467. https://doi.org/https://doi.org/10.1007/s10479-021-04149-2.

Shabbir, A., Shabir, M., Javed, A. R., Chakraborty, C., & Rizwan, M. (2022). Suspicious transaction detection in banking cyber–physical systems. Computers & Electrical Engineering, 97, 107596. https://doi.org/https://doi.org/10.1016/j.compeleceng.2021.107596.

Sharifani, K., & Amini, M. (2023). Machine learning and deep learning: A review of methods and applications. World Information Technology and Engineering Journal, 10(07), 3897-3904.

Sharma, P., Prasad, J. S., & Ahamed, S. K. (2024). An efficient cyber threat prediction using a novel artificial intelligence technique. Multimedia Tools and Applications, 1-17. https://doi.org/https://doi.org/10.1007/s11042-024-18169-0.

Shaukat, K., Luo, S., Varadharajan, V., Hameed, I. A., Chen, S., Liu, D., & Li, J. (2020). Performance comparison and current challenges of using machine learning techniques in cybersecurity. Energies, 13(10), 2509. https://doi.org/ https://doi.org/10.3390/en13102509.

Shekar, G. K., & Ramakrisha, R. S. (2021). Semisupervised algorithms based creditcard fraud detection using majority voting. International Journal of Advanced Research in Science and Technology, 11(1), 23-30.

Singh, A., & Jain, A. (2019). Financial fraud detection using bio-inspired key optimization and machine learning technique. International Journal of Security and Its Applications, 13(4), 75-90. https://doi.org/http://dx.doi.org/10.33832/ijsia.2019.13.4.08.

Singh, A., Ranjan, R. K., & Tiwari, A. (2022). Credit card fraud detection under extreme imbalanced data: a comparative study of data-level algorithms. Journal of Experimental & Theoretical Artificial Intelligence, 34(4), 571-598. https://doi.org/https://doi.org/10.1080/0952813x.2021.1907795.

Soh, W. W., & Yusuf, R. M. (2019). Predicting credit card fraud on a imbalanced data. International Journal of Data Science and Advanced Analytics, 1(Vol. 1 No. 1 (2019)), 12-17.

Soni, K. B., Chopade, M., & Vaghela, R. (2021). Credit card fraud detection using machine learning approach. Appl. Inf. Syst. Manag, 4(2), 71-76. https://doi.org/https://doi.org/10.15408/aism.v4i2.20570.

Statistics, A. B. o. (2024). Personal Fraud. ABS. Retrieved 05-05-2024 from <https://www.abs.gov.au/statistics/people/crime-and-justice/personal-fraud/latest-release>.

Sujatha, M. (2019). A comparative study of credit card fraud detection using machine learning for United Kingdom dataset. International Journal of Computer Science and Information Security (IJCSIS), 17(9), 2019.

Susman, G. (1983). Action_Research. Beyond Method: Strategies for Social Research. In: Newbury Park, California: Sage Publications.

Taha, A. Y., Tiun, S., Abd Rahman, A. H., & Sabah, A. (2021). Multilabel over-sampling and under-sampling with class alignment for imbalanced multilabel text classification. Journal of Information and Communication Technology, 20(3), 423-456. https://doi.org/https://doi.org/10.32890/jict2021.20.3.6.

Taye, M. M. (2023). Theoretical understanding of convolutional neural network: Concepts, architectures, applications, future directions. Computation, 11(3), 52. https://doi.org/ https://doi.org/10.3390/computation11030052.

Tingfei, H., Guangquan, C., & Kuihua, H. (2020). Using variational auto encoding in credit card fraud detection. IEEE Access, 8, 149841-149853. https://doi.org/https://doi.org/10.1109/access.2020.3015600.

Trivedi, N. K., Simaiya, S., Lilhore, U. K., & Sharma, S. K. (2020). An efficient credit card fraud detection model based on machine learning methods. International Journal of Advanced Science and Technology, 29(5), 3414-3424.

Unogwu, O. J., & Filali, Y. (2023). Fraud detection and identification in credit card based on machine learning techniques. Wasit Journal of Computer and Mathematics Science, 2(3), 16-22. https://doi.org/ https://doi.org/10.31185/wjcms.185.

Veigas, K. C., Regulagadda, D. S., & Kokatnoor, S. A. (2021). Optimized stacking ensemble (OSE) for credit card fraud detection using synthetic minority oversampling model. Indian Journal of Science and Technology, 14(32), 2607-2615. https://doi.org/https://doi.org/10.17485/ijst/v14i32.807 .

Vengatesan, K., Kumar, A., Yuvraj, S., Kumar, V., & Sabnis, S. (2020). Credit card fraud detection using data analytic techniques. Advances in Mathematics: Scientific Journal, 9(3), 1185-1196. https://doi.org/https://doi.org/10.37418/amsj.9.3.43 .

Voican, O. (2021). Credit Card Fraud Detection using Deep Learning Techniques. Informatica Economica, 25(1). https://doi.org/https://doi.org/10.24818/issn14531305/25.1.2021.06.

Wang, L., Han, M., Li, X., Zhang, N., & Cheng, H. (2021). Review of classification methods on unbalanced data sets. IEEE Access, 9, 64606-64628.

Weng, W., Wang, D.-H., Chen, C.-L., Wen, J., & Wu, S.-X. (2020). Label specific features-based classifier chains for multi-label classification. IEEE Access, 8, 51265-51275. https://doi.org/https://doi.org/10.1109/access.2020.2980551.

Wu, E., Cui, H., & Welsch, R. E. (2020). Dual autoencoders generative adversarial network for imbalanced classification problem. IEEE Access, 8, 91265-91275. https://doi.org/https://doi.org/10.1109/access.2020.2994327.

Xin, Y., Kong, L., Liu, Z., Chen, Y., Li, Y., Zhu, H., Gao, M., Hou, H., & Wang, C. (2018). Machine learning and deep learning methods for cybersecurity. IEEE Access, 6, 35365-35381. https://doi.org/https://doi.org/10.1109/access.2018.2836950.

Yamini, K., Anitha, V., Polepaka, S., Chauhan, R., Varshney, Y., & Singh, M. (2023). An Intelligent Method for Credit Card Fraud Detection using Improved CNN and Extreme Learning Machine. 2023 8th International Conference on Communication and Electronics Systems (ICCES).

Yao, J., Pan, Y., Yang, S., Chen, Y., & Li, Y. (2019). Detecting fraudulent financial statements for the sustainable development of the socio-economy in China: a multi-analytic approach. Sustainability, 11(6), 1579. https://doi.org/https://doi.org/10.3390/su11061579.

Yuan, Y., Wei, J., Huang, H., Jiao, W., Wang, J., & Chen, H. (2023). Review of resampling techniques for the treatment of imbalanced industrial data classification in equipment condition monitoring. Engineering Applications of Artificial Intelligence, 126, 106911. https://doi.org/https://doi.org/10.1016/j.engappai.2023.106911.

Zhang, D., Bhandari, B., & Black, D. (2020). Credit card fraud detection using weighted support vector machine. Applied Mathematics, 11(12), 1275. https://doi.org/ https://doi.org/10.4236/am.2020.1112087.

Zhang, X., Han, Y., Xu, W., & Wang, Q. (2021). HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Information sciences, 557, 302-316. https://doi.org/ https://doi.org/10.1016/j.ins.2019.05.023.

Zioviris, G., Kolomvatsos, K., & Stamoulis, G. (2022). Credit card fraud detection using a deep learning multistage model. The Journal of Supercomputing, 78(12), 14571-14596. https://doi.org/https://doi.org/10.1007/s11227-022-04465-9.

APPENDIX A

1. Publications

No	Paper	Paper type	
1	Btoush, E., Zhou, X., Gururaian, R., Chan, K.C. and Tao, X., 2021, October. A survey on credit card fraud detection techniques in banking industry for cyber security. In 2021 8th International Conference on Behavioral and Social Computing (BESC) (pp. 1-7). IEEE. https://doi.org/10.1109/BESC53957.2021.9635559	Conference	Published
2	Btoush, E.A.L.M., Zhou, X., Gururajan, R., Chan, K.C., Genrich, R. and Sankaran, P., 2023. A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Computer Science, 9, p.e1278. doi: https://doi.org/10.7717/peerj-cs.1278	(Q1)Journal	Published
3	Optimising Security: A Hybrid CNN-BiLSTM Model for Credit Card Cyber Fraud Detection	Conference / CBD2024 The Twelfth International Conference on Advanced Cloud and Big.	Accepted On 02/09/2024.
4	Twitter-Based Cyber Threat Detection Through IDCNN and BiLSTM Integration	Conference / CBD2024 The Twelfth International Conference on Advanced Cloud and Big	Accepted On 02/09/2024.
5	Credit Card Cyber Fraud Detection Using Machine Learning: A Comparative Analysis of Resampling Techniques.	Journal/ Resubmitted in IEEE ACCESS. 21/07/2024.	Under review
6	Component-Based e-organization Application Sustainability and its Behavioral Security Modeling.	Conference / Annual Computer Security Applications Conference (ACSAC). Submitted on 04/09/2024.	Under review

A Survey on Credit Card Fraud Detection Techniques in Banking Industry for Cyber Security

Eyad Btoush
School of Business
University of Southern Queensland
Brisbane, Australia
EyadAbdelLatif.A.Q.MarazqahBtoush@usq.edu.au

Xujuan Zhou
School of Business
University of Southern Queensland
Brisbane, Australia
xujuan.zhou@usq.edu.au

Raj Gururajan
School of Business
University of Southern Queensland
Brisbane, Australia
raj.gururajan@usq.edu.au

KC Chan
School of Business
University of Southern Queensland
Brisbane, Australia
kc.chan@usq.edu.au

XiaoHui Tao
School of science
University of Southern Queensland
Brisbane, Australia
xiaohui.tao@usq.edu.au

2021 8th International Conference on Behavioral and Social Computing (BESC) | 978-1-6654-0023-7/21/$31.00 ©2021 IEEE | DOI: 10.1109/BESC53957.2021.9635559

***Abstract*— The technological revolution is accelerating due to a number of key enabling technologies, such as Artificial Intelligence (AI)/Machine Learning (ML), big data, blockchain, cloud computing, Internet of Thing (IoT). With the broad adoption of ever-improving internet technology, cyber security is of great importance in the banking industry due to the rising number of cyber attacks and crimes. Credit card fraud is one of the most serious threats facing the banking industry worldwide. Credit card fraud is expanding at an alarming rate and has developed into a significant problem, particularly as the volume of financial transactions involving credit cards continues to expand. In this paper, we have reviewed various credit card fraud detection techniques that can strengthen the defense against a range of frauds. Additionally, we analysed the findings and reported the research challenges. Finally, we compared various techniques and highlighted their advantages and disadvantages. This will help provide guidance for determining the most appropriate techniques for credit card fraud detection.**

Keywords—Fraud, Credit card, Machine learning, Artificial intelligence, Credit card fraud detection, Banking.

I. Introduction

The development of information technology (IT) has had a profound impact on the banking industry. However, as it evolved, so too have the variety of ways in which individuals fell victims of various attacks. The majority of banking application system transactions are now completed via credit cards or online net banking and responding to client demands via a range of channels, including online, mobile, web, and Internet of Things, all of which pose new vulnerabilities [1].

Banks holding large amount of client data have made themselves a prime target for hackers. As a result, banks have been at the forefront of enterprise cybersecurity.In the past decade, cyber security has gained popularity in the banking industry. Over the last 13 years, the cybersecurity market has grown approximately 35 times. The market was predicted to reach $170.4 billion in 2022[2], and cybercrime is expected to cost the global economy $6 trillion USD in 2021. Worldwide cybercrime expenses would increase by 15% each year over the next five years, reaching $10.5 trillion USD annually by 2025[3].

Credit card fraud is one of the major challenges for the banking industry, resulting in billions of dollars in annual losses. The banking industry has made enhancing cyber security protection a priority due to the rising number of cyber threats and attacks. To detect credit card fraud, several monitoring and detection systems have been developed. However, with the threat landscape constantly evolving, it is essential to equip banks with smart and novel technologies for threat management [4].

This paper contains a review of credit card fraud detection techniques. The rest of this paper is structured as follows: Section two is an overview of Credit card fraud, Section three defines the credit card fraud detection problem, Section four reviews the relevant studies and the algorithms used in previous researches, Section five is discussion of related works and the final section concludes the paper and summarises future works .

II. Credit Card Fraud

Individuals' modes of payment have shifted dramatically as a result of the advancement of modern technology. The use of online payment methods such as Online Banking, Debit Card, and Credit Card has grown. Credit card fraud has grown extremely prevalent in the modern day, with several cases reported in recent years due to the rise in cybercrime. Credit card fraud is a sort of identity theft that occurs when someone other than you makes an unauthorised payment using your credit card or account information. A credit card fraud can occur as a result of a stolen, misplaced, or counterfeit card. Credit card fraud is one of the most serious hazards facing individuals and banking industry worldwide, particularly as the volume of financial transactions involving credit cards continues to expand. Credit card fraud is expanding at an alarming rate and has developed into a significant issue in the banking industry [5].

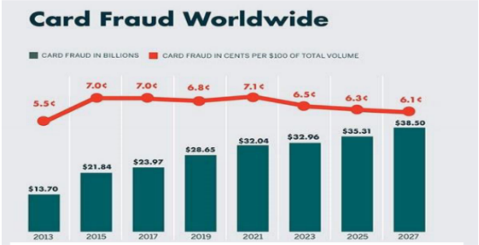


Figure 1 Global credit card fraud losses

PeerJ
Computer Science

A systematic review of literature on credit card cyber fraud detection using machine and deep learning

Eyad Abdel Latif Marazqah Btoush[1], Xujuan Zhou[1], Raj Gururajan[1,2], Ka Ching Chan[1], Rohan Genrich[1] and Prema Sankaran[3]

[1] School of Business, University of Southern Queensland, Toowoomba, QLD, Australia
[2] School of Computing, SRM Institute of Science and Technology, Chennai, India
[3] School of Management, Presidency University, Bangalore, India

ABSTRACT

The increasing spread of cyberattacks and crimes makes cyber security a top priority in the banking industry. Credit card cyber fraud is a major security risk worldwide. Conventional anomaly detection and rule-based techniques are two of the most common utilized approaches for detecting cyber fraud, however, they are the most time-consuming, resource-intensive, and inaccurate. Machine learning is one of the techniques gaining popularity and playing a significant role in this field. This study examines and synthesizes previous studies on the credit card cyber fraud detection. This review focuses specifically on exploring machine learning/deep learning approaches. In our review, we identified 181 research articles, published from 2019 to 2021. For the benefit of researchers, review of machine learning/deep learning techniques and their relevance in credit card cyber fraud detection is presented. Our review provides direction for choosing the most suitable techniques. This review also discusses the major problems, gaps, and limits in detecting cyber fraud in credit card and recommend research directions for the future. This comprehensive review enables researchers and banking industry to conduct innovation projects for cyber fraud detection.

Subjects Algorithms and Analysis of Algorithms, Artificial Intelligence, Data Mining and Machine Learning, Security and Privacy
Keywords Machine learning, Deep learning, Cyber security, Credit card cyber fraud, Bank industry, Artificial intelligence

Submitted 28 December 2022
Accepted 15 February 2023
Published 17 April 2023

Corresponding author
Eyad Abdel Latif Marazqah Btoush, EyadAbdelLatif.A.Q.MarazqahBtoush@usq.edu.au

Academic editor
Vladimir Simic

Additional Information and Declarations can be found on page 55

DOI 10.7717/peerj-cs.1278

INTRODUCTION

The banking industry has been profoundly impacted by the evolution of information technology (IT). Credit card and online net banking transactions, which are currently the majority of banking system transactions, all present additional vulnerabilities (*Jiang & Broby, 2021*). Hackers have increasingly targeted banks with enormous quantities of client data. Therefore, banks have been in the forefront of cyber security for business. In the past thirteen years, cyber security industry expanded fast. The market is predicted to be valued 170.4 billion in 2022 (*Morgan, 2019*). In the next three years, the cost of cybercrime is expected to rise by 15% every year, finally exceeding $10.5 trillion USD each year by 2025 (*Morgan, 2020*).

How to cite this article Marazqah Btoush EAL, Zhou X, Gururajan R, Chan KC, Genrich R, Sankaran P. 2023. A systematic review of literature on credit card cyber fraud detection using machine and deep learning. PeerJ Comput. Sci. 9:e1278 DOI 10.7717/peerj-cs.1278

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2022.Doi Number

Credit Card Cyber Fraud Detection Using Machine Learning: A Comparative Analysis of Resampling Techniques

Eyad Btoush[1], Xujuan Zhou[1], Raj Gururajan[1,2], KC Chan[1], Omar Alsodi[1]

[1]School of Business, University of Southern Queensland, Toowoomba, QLD, Australia
[2]School of Computing, SRM Institute of Science and Technology, Chennai, India

Corresponding author: Eyad Btoush (e-mail: EyadAbdelLatif.A.Q.MarazqahBtoush@usq.edu.au).

ABSTRACT The prevalence of online transactions and extensive adoption of credit card payments have contributed to the rise of credit card cyber fraud in modern society. These trends are propelled by technological advancements, which provide fraudulent actors with more opportunities. Fraudsters exploit victim's financial vulnerabilities by obtaining illegal access to sensitive credit card information through deceptive means, such as phishing, fraudulent phone calls, and fraudulent SMS messages. This study addresses these critical challenges by leveraging Machine Learning(ML) techniques including Random Forest(RF), Decision Tree(DT), Logistic Regression(LR) , K-Nearest Neighbors (KNN), Support Vector Machine (SVM), Extreme Gradient Boosting (XGBoost), and Categorical Boosting(CatBoost), and various sampling techniques such as Tomek Link, Synthetic Minority Oversampling Technique(SMOTE), Edited Nearest Neighbor(ENN), Tomek+ENN, and SMOTE+ENN to accurately predict and detect cases of cyber fraud in credit card transactions. We performed a comprehensive comparison analysis to evaluate the performance of these ML techniques in term of accuracy, precision, recall, F1 score, and ROC-AUC score. The objective of this study was to increase the cyber fraud detection effectiveness and thus to achieve an increased financial security leveraging the different ML techniques.

.

INDEX TERMS Machine learning, Credit card cyber fraud, Cyber fraud, Resampling techniques.

I. INTRODUCTION

Credit card usage has experienced an enormous increase in recent times, owing to the constant advancement of technology. Consequently, this surge in usage has led to a progressive increase in the occurrence of cyber fraud associated with credit cards. Credit cards are the prevailing mode of payment in the modern business environment, encompassing both minor- and large-scale industries. The extensive integration of credit card cyber fraud has rendered it a widespread concern in numerous industries. The banking industry has experienced substantial transformation and benefited substantially from technological advancements. Nevertheless, the issue of credit card cyber fraud has been expedited owing to the fast rate of cyber adoption.

The banking industry faces significant difficulties owing to credit card cyber fraud, which causes financial losses amounting to billions of dollars annually (Malik et al., 2022; Krishna et al., 2023). In recent years, the acceptability of credit cards and other online payments has increased, and consequently, credit card cyber fraud has increased. This makes cyber fraud detection an essential industry. Traditional approaches to address this problem are insufficient (Ni et al, 2023; Zhu et al., 2023). The algorithms for the conventional method have been formulated by experts in the cyber fraud domain. In addition, a proactive approach is required to counteract cyber fraud. All sectors have endeavored to implement Machine Learning (ML) based solutions owing to their widespread adoption, rapidity, and efficacy. It has been established that the ML approach is among the most efficient approaches in this area (Mniai et al. 2023). The implementation of ML techniques has the potential to enhance the accuracy of monitoring system anomaly detection and threat reduction (Ala'raj et al., 2021). ML techniques facilitate extensive consumer behavior surveillance and individual profiling, allowing for the tracking of their actions and the provision of actionable insights to aid in the mitigation of risks and cyber fraud. (Priya & Saradha, 2021). To improve the data processing

Optimizing Security: A Hybrid CNN-BiLSTM Model for Credit Card Cyber Fraud Detection

Eyad Btoush
The School of Business
University of Southern Queensland
Brisbane, Australia
EyadAbdelLatif.A.Q.MarazqahBtoush@usq.edu.au

Xujuan Zhou
The School of Business
University of Southern Queensland
Brisbane, Australia
Xujuan.Zhou@usq.edu.au

Raj Gururajan
The School of Business
University of Southern Queensland
Brisbane, Australia
Raj.Gururajan@usq.edu.au

KC Chan
The School of Business
University of Southern Queensland
Brisbane, Australia
kc.chan@unisq.edu.au

Omar Alsodi
The School of Business
University of Southern Queensland
Brisbane, Australia
Omar.Alsodi@usq.edu.au

Abstract**—***The banking industry has long recognised the importance of developing efficient credit card cyber fraud detection systems. Despite these efforts, businesses face rising credit card cyber fraud due to technological development. The growing frequency of cybersecurity data breaches has made credit card cyber fraud detection systems less efficient at detecting advanced fraud. Although Machine Learning(ML) algorithms have been employed to detect credit card cyber fraud, no cyber fraud detection system has yet to achieve a high level of efficiency. Using DL techniques for credit card fraud detection has yielded better performance than traditional algorithms. It tackled the problem of detecting unexpected and sophisticated cyber fraud patterns. This paper proposes a novel Hybrid CNN-BiLSTM model to effectively address credit card cyber fraud. The novel hybrid model uses DL techniques like convolutional neural network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). We conducted a thorough comparison of the accuracy, precision, recall, F1 score, and ROC-AUC score. The experiments demonstrate that the innovative CNN-BiLSTM model surpasses the performance of each individual model. The integration of CNN and BiLSTM techniques represents substantial progress in the domain of cyber fraud detection, thereby boosting the security of financial transactions*

Keywords—*Machine Learning, Deep Learning, Credit Card Cyber Fraud, Cyber Fraud.*

I. Introduction

The banking industry has greatly benefited from the development of technologies and has undergone considerable transformation. However, the rapid increase in technology usage has exacerbated cyber fraud using credit cards. Credit card cyber fraud refers to the act of committing illegal financial transactions by fraudsters as a means of identity theft. Most banking application system transactions are currently conducted using credit cards or online net banking (Patil et al. 2018), leading to increased vulnerability to new attacks and techniques. Thus, cyber fraud is far more serious in the banking industry (Mienye& Sun 2023). As cybercrimes of increasing sophistication become increasingly prevalent in infiltrating digital lives, there is an imperative need to innovate credit card cyber fraud detection. This innovation should be built around the sophisticated techniques utilized in cybersecurity and firmly grounded in the fundamental concepts of safeguarding against cyber threats (Almarshad et al. 2023). Credit card cyber fraud is a major challenge for the banking industry, resulting in billions of dollars in annual losses(Krishna et al. 2023). Several detection systems have been developed to detect credit card cyber frauds. However, as the threat landscape constantly evolves, it is essential to equip banks with smart and novel technologies for threat management(Seera et al. 2024). ML techniques have been employed in the domain of credit card cyber fraud detection to facilitate data processing and analysis of data(Soni et al. 2021). However, no cyber fraud detection system has yet to achieve a high level of efficiency.

DL plays a crucial role in identifying credit card cyber frauds. Utilizing DL approaches can enhance the probability of monitoring systems to accurately identify anomalies and mitigate possible threats (Ala'raj et al. 2021). The advantage of DL is that it enables the efficient replacement of features manually using unsupervised or semi-supervised feature learning and hierarchical feature extraction (Xin et al., 2018; Agarwal et al., 2021). The primary advantage of DL over traditional ML is its higher performance on large datasets.

Despite the application of ML algorithms efforts to mitigate cyber fraud have been ineffective, revealing challenges, such as false positives and negatives. False negatives cause financial losses for both cardholders and institutions. These difficulties continue to exist because of unbalanced data, adversarial attacks, concept dispersion, and model interpretability issues. This study aims to address these challenges by employing DL techniques to enhance the detection of credit card cyber fraud. Specifically, we applied and evaluated the performance of DL algorithms, including Convolutional Neural Network (CNN) and Bidirectional Long Short-Term Memory (BiLSTM). Additionally, we developed novel Hybrid CNN-BiLSTM model. To determine the performance of the algorithms in terms of accuracy, precision, recall, F1 score, and Roc_Auc score for credit card cyber fraud detection, we conducted a comparative analysis of these DL techniques. Our study provides a comprehensive assessment of these techniques. The objective of this study is to develop a resilient model that can accurately detect and combat instances of credit card cyber fraud.

II. Cyber Fraud in Credit Card

Credit card cyber fraud has grown extremely prevalent in the modern day, with several cases reported in recent years owing to the rise in cybercrime. The term "cyber fraud" pertains to individuals engaged in cybercriminal activities, utilising technological resources to execute cyberattacks that lead to the compromise of sensitive data, such as credit card information. These compromised data can then be used for fraudulent purposes. Credit card cyber fraud is expanding at an alarming rate and has become a significant issue in the banking industry(Karthika& Senthilselvi 2023).

Criminals in the field of cyber fraud are constantly enhancing their techniques for avoiding detection, and are now integrating novel methods to circumvent credit cyber