National College of Ireland

FusionGuard: Integrating LSTM-GRU with Temporal Anomaly Detection for Robust Supply Chain Fraud Identification

MSc Research Project
MSc Data Analytics

Jeni Keziah Alfrida
Student ID: 23401044

School of Computing
National College of Ireland

Supervisor: Hamilton Niculescu

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name: Jeni Keziah Alfrida

Student ID: 23401044

Programme: MSc Data Analytics **Year:** …2025…..

Module: Research Project……

Supervisor: …Hamilton Niculescu……

Submission Due Date: ……29/01/26……

Project Title: FusionGuard: Integrating LSTM-GRU with Temporal Anomaly Detection for Robust Supply Chain Fraud Identification

Word Count: 9775 **Page Count**……21……

I hereby certify that the information contained in this (thesubmission) is information pertaining to research I conducted for this project. All information other than theown contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Jeni Keziah Alfrida……

Date: 28/01/26……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

FusionGuard: Integrating LSTM-GRU with Temporal Anomaly Detection for Robust Supply Chain Fraud Identification

Jeni Keziah Alfrida
Student ID: 23401044

Abstract:

The study provides a solution to the multidimensional challenges and introduces FusionGuard, a hybrid fraud-detection model that implements sequential learning (also referred to as sequential modelling), anomaly detection and class-imbalance remedies in a unified framework. FusionGuard integrates both LSTM and GRU layers to learn long-term time-dependent and short-term dynamic variations in transactional data, and techniques such as SMOTE are used to deal with extreme class imbalance, while focal loss reduces prioritisation bias toward majority-class samples. The proposed model is primarily supervised, with conceptual support from anomaly-detection principles, thereby increasing its flexibility, and it can be applied to a variety of supply-chain datasets that contain mixtures of both labelled and unlabelled data. the FusionGuard methodology involves more than two layers of preprocessing and feature testing. The interquartile range method is used to remove outliers, correlation is analysed to determine high-impact features, and scaling is done using StandardScaler to maximise the performance of gradient-based learning algorithms. Exploratory data analysis shows the severity of the class imbalance, and it is used to decide on the modelling techniques (strategies) that can work best with class-imbalanced distributions. The results of its performance define a standard with which the sequential deep-learning models can be compared. The experimental outcomes explain that the hybrid model can considerably lower the false negatives, one of the most severe measures in fraud detection, and show almost flawless accuracy and recall in detecting fraudulent transactions.

Keywords: LSTM (Long Short-Term Memory), GRU (Gated Recurrent Unit), SMOTE (Synthetic Minority Over-sampling Technique), Undersampling, Credit Card, Fraud, Detection, Bias, Imbalance, Comparison, Logistic Regression

1 Introduction

The fraudulent detection of transactions in financial systems has gained more significance as organisations use digital payment structures that are automated and of high volume. Fraud is relatively rare but carries serious financial and reputational consequences, creating a strong need for early and accurate detection. This project explores the various modelling schemes of detecting fraudulent credit-card transactions in highly imbalanced data. It makes a comparison between classical machine learning and sequential deep learning architectures when analysing the effect of data-balancing techniques on performance. The aim is to create a practical and empirically validated framework of detection that would be able to detect rare fraud cases with higher recall and lower false negatives.

The present research builds upon previous literature on studying fraud-detection models by providing an empirical analysis of the effectiveness of the temporal modelling of the imbalance-sensitive sampling of the classes when the class-specific distributions are realistic. Contrary to most previous works where test sets are artificially balanced, this paper evaluates generalisation in the case of extreme imbalance, providing valuable practical experience in real-world applications.

A very small percentage of the credit-card fraud dataset is made up of fraud cases. In this case, less than 0.2%. These imbalances result in many of the conventional models working well in overall accuracy, but missing most fraudulent transactions. Past studies point to the usefulness of temporal modelling in which behaviours among a series of transactions can display behavioural anomalies that cannot be seen in one-point studies. LSTM and GRU are recurrent architectures that are well adapted to the ability to capture these temporal cues. Simultaneously, resampling samples such as Random Undersampling and SMOTE provide means by which the training data may be rebalanced to ensure that a given model gets enough exposure to the minority-class patterns. This project will extend these concepts by combining linear and sequential models on the basis of these imbalance-handling strategies and test them on a realistically imbalanced test set.

1.1 Research Problem

The main problem that is resolved in the study relates to the fact that it is not possible to reliably identify fraudulent transactions in an environment where fraud is extremely uncommon, behaviour is dynamic and features available give little interpretability due to anonymisation. Classical classifiers are prone to ignoring fraud because of bias in majority classes and models that do not employ temporal data have a hard time capturing changing spending behaviour. The issue is thus two-fold: models should be more attuned to the patterns of fraud, but should not be flooded with false positives and they should be able to recognise behaviour over time, not on a transaction-by-transaction basis.

1.2 Aim

This study will focus on designing and empirically evaluating FusionGuard, a hybrid fraud-detection framework focused on credit-card transaction data under severe class imbalance. It combines sequential deep learning with imbalance-management methods to enhance the minority-class detection rates of the system over traditional machine-learning systems.

1.3 Objectives

The aims of the research are directed at the creation of a fraud-detection system functioning within the conditions of highly imbalanced financial data. It starts by preparing and pre-processing the dataset to ensure that models are fed clean and stable data; this involves the elimination of duplicates through the process, outliers in transaction value and scaling features are done to maintain the same behaviour

between classical and deep-learning models. One of the aims is to examine the influence of imbalance-handling techniques, in particular, Random Undersampling and SMOTE, on the learning performance of the model in terms of either following the patterns of the minority-classes or providing a realistic assessment environment, by applying these techniques only to the training data. The other aim is to develop two modelling plans and compare the Logistic Regression as a baseline linear representation and a hybrid LSTM-GRU network that can learn the temporal structure in a sequence of transactions. The comparison will enable the study to know whether sequential modelling has a significant benefit in uncovering rare and changing fraud behaviour. The last goal is to assess all the models based on such metrics as accuracy, precision, recall, F1-score and ROC-AUC and, in particular, reducing false negatives. These goals, together, include a systematic foundation of the investigation of the impact of sampling selections and model structure on the aggregate fraud-detection results.

The operationalisation of these objectives will involve measuring model performance according to recall, F1-score, and ROC-AUC with a special focus on reducing false negatives. Success is the statistical and practical significance, that is, a statistically and practically significant improvement in recall over the logistic regression baseline on a realistic test set, without leading to a false positive rate so high as to become uncontrollable.

1.4 Research Questions

RQ1: How does the hybrid LSTM-GRU model perform in detecting fraudulent transactions relative to traditional models under severely imbalanced financial data conditions?

RQ2: How much are sequential components (LSTM/GRU) more effective than non-sequential methods in detecting the presence of temporally evolving patterns of fraud?

RQ3: What are the effects of imbalance-handling methods (SMOTE and undersampling) on recall and missed frauds and the manageable false-positive rates?

In this work, data preprocessing, imbalance management and sequential modelling are combined to assess the influence of the various methodological decisions on the performance of fraud detection. The project offers an evidence-based, clear insight into how temporal learning and balanced training data help to achieve more reliable fraud-detection results by testing a number of model-and-sampling combinations on a realistic test distribution. The conclusions of this research are the basis of the methodological and analytical discourses in the subsequent chapters.

2 Literature Review

Literature on fraud detection in supply-chain and financial-transaction systems has increased significantly in the past years, with variations in the techniques, data modalities and focus specified. On one side, there is work like Credit Card Fraud Detection Model Based on LSTM Recurrent Neural Networks by Benchaji, Douzi and El Ouahidi (2021), which looks at sequential models of deep-learning applications in which the chronological sequence of credit-card transactions is explicitly represented in the form of Long Short-Term Memory (LSTM) networks. Their contribution points out the significance of behaviour change and the time dependencies associated with fraud events and highlights the weaknesses of purely statistical classifiers. Further along this axis, the problem space is extended to big-data flows and networked vendor, buyer relationships, such as in Dataco Smart Supply Chain by Constante, Silva and Pereira (2019), which focuses on big-data structures and analytics in supply-chain applications. Meanwhile, graph-based relational approaches have become available to address coordinated or collusive fraud: GoSage: Heterogeneous Graph Neural Network Using Hierarchical Attention for Collusion Fraud Detection (Ghosh and Bhowmik, 2023), Intention-

Aware Heterogeneous Graph Attention Networks for Fraud Transactions Detection (Liu et al., 2021), and Pick and Choose: A GNN-Based Imbalanced Learning Approach to Detect Fraud (Liu et al., 2021). These later works show how heterogeneous graph neural networks (GNNs), attention models and imbalance-sensitive learning methodologies are employed in fraud detection in multi-entity networks. These three themes in combination allow us to realise that modelling changing fraud trajectories involves modelling temporal sequences to capture changing behaviours; relational and networked structure modelling to identify collusive or coordinated fraud; and addressing the challenge of class imbalance using specialised strategies to recognise rare but significant events of fraud. However, although each of these strands has contributed to the state of the art, a missing linkage between these unanimous structures is the incorporation of temporality, relational structure and imbalance-adapted learning in scalable architectures. This literature review thus systematises the literature around these axes, analyses the merits and demerits of each contribution, and identifies how the proposed FusionGuard framework (in this study) fits into and builds upon these preceding contributions.

Benchaji et al (2021) deals with an urgent issue in the field of payment-card fraud detection: that in most automated models, the analysis is based on the distinguishing characteristics of individual transactions without considering the sequentiality of spending behaviour and the way chronic offenders develop their strategies. According to authors, current machine-learning models do not pay much attention to behavioural indicators such as transaction history, changes in spending behaviour or temporal correlation of transactions, so they tend to perceive every transaction as a separate case. The authors say this omission degrades the accuracy of detection, especially when there is a change in fraud tactics or concept drift (i.e., a shift in the distribution of legitimate and fraudulent behaviour). It is therefore suggested that a new detection system using Long Short-Term Memory (LSTM) recurrent neural networks is needed, as LSTMs can capture sequential dependencies between transaction histories and thus can more effectively distinguish between legitimate and fraudulent behaviour.

Benchaji et al. (2021) start the paper by providing motivation for their direction: they observe that the amounts of money lost to credit-card fraud are huge (referring to the risk of US $24.71 billion and US $27.69 billion in 2016 and 2017, respectively) and that the emergence of e-commerce, mobile payments and digital channels increases both the volume and exposure to fraud. They note that conventional detection systems are deficient in three aspects: first, the ever-changing nature of fraudsters (fraud methods constantly evolve); second, the fact that similar transactions can be associated with vastly different levels of risk depending on the cardholder and situation (static profile models are outdated); and third, seasonality, geographical and behavioural changes that affect consumer spending (a static profile might no longer be useful with the passage of time). These observations form the basis of their decision to use a sequence-based model, as opposed to a single-transaction classifier. They claim that when the model can learn from the historical sequences of transactions of the same cardholder (or account), it will be able to notice aberrations not evident in a single transaction.

The temporal learning framework proposed by Benchaji et al. sets the initial plane of this premise since the authors show that fraud is not present in isolated transaction categorisation but in contextual patterns needed to precisely detect it. Their approach based on LSTM explains the usefulness of modelling sequences of behaviour in which fraud is frequently latent in patterns that evolve in time, abrupt increases and decreases in size, changes in device or geographic location and pre-attack preparatory behaviours. Their model, however, is successful only in capturing the long-term dependencies in the sequence mode but fails to capture the larger relational framework that defines a fraud network. To mask their actions, fraudulent entities do not act in isolation; instead they rely on relationships with merchants, accounts, devices or other users to hide their actions. Thework

acknowledges this limitation and therefore goes beyond purely sequential architectures by adding elements of relational reasoning inspired by heterogeneous graph-learning models.

The relational aspect becomes apparent in the writings of Ghosh and Bhowmik and of Liu et al. (IHGAT), which shed more light on fraud through multi-type relationships as compared to short-term trends. The GoSAGE framework shows that in practical application contexts such as digital payments, ride-hailing or e-commerce, people become members of structurally rich networks with various types of relations, transactions, shared devices, shared funding sources, logistics overlaps and referral chains, indicating that hierarchical attention modelling over heterogeneous edges can yield significant benefits in collusion detection. Similarly, IHGAT demonstrates that the accuracy of fraud detection increases when behavioural intentions and relationships among transactions are considered elements of a heterogeneous graph rather than independent data modalities. These contributions highlight why relational modelling is required for representing coordinated behaviour or malicious substructures. However, both techniques assume the availability of sufficiently dense relational metadata and rely heavily on graph connectivity as the key signal. They do not discuss how to effectively utilise temporal information alongside graph-based signals or how to manage extremely skewed class distributions in graph data, a factor that becomes even more significant in real-world supply chain and financial data where fraud is not only uncommon but also structurally sparse.

Temporal modelling tackles the question of how the behaviour of fraudsters changes over time; relational modelling tackles the question of how fraud actors organise; imbalance-based learning tackles the question of how indicators of fraud are statistically distorted. FusionGuard claims that successful fraud-detection systems need an integrated strategy in which sequential dependencies, interconnected relations and imbalance-sensitive learning are applied simultaneously. The literature shows innovation within each individual dimension, but no adequate framework has been provided to integrate them into a single architecture capable of handling behavioural sequences, relational structures and skewed distributions. By integrating these dimensions, FusionGuard provides a solution that compensates for the limits of single-focus approaches and develops a more holistic detection system reflecting the complexity of real-life fraud ecosystems.

In this respect, the LSTM-GRU sub-block of the model draws directly from the temporal knowledge developed by Benchaji et al., which identifies the significance of modelling sequences of transactions as evolving motion rather than as isolated observations. The hybrid LSTM-GRU architecture, which does not impose constraints on expressive modelling but offers high efficiency and lower computational load, is well matched to the requirements of expressive yet computationally efficient sequential modelling. One insight it responds to is that fraud tends to consist of both long-term behavioural changes and short-term anomalies, and a dual-sequence model can identify both. Additionally, by introducing it into a broader pipeline instead of implementing it as a separate classifier, the work responds to limitations acknowledged in Benchaji et al. and develops them further to gain a more contextual representation.

Likewise, the recognition of relational complexities is informed by works such as GoSAGE and IHGAT, which underscore the need to recognise relational interactions within heterogeneous relational networks. Although FusionGuard does not operate on multi-relational graphs but on transactional datasets, it is conceptually connected to the idea that fraud cannot be comprehended purely through sequences. As with credit-card and supply chain applications, there is relational information in the form of user, merchant transactions, user, device relationships, distribution networks, account associations and vendor relationships. Even though a full heterogeneous GNN module is not currently implemented in the work, the conceptual architecture aligns with these relational frameworks, since it combines supervised and unsupervised processes designed to approximate relational reasoning by providing reconstruction-based anomaly warnings. Independent of graph-based neighbourhood deviation modelling, autoencoders correspond to a form of structural

anomaly detection. This positions FusionGuard as a bridge between sequence-based deep learning and graph-based anomaly detection, reflecting theoretical principles from both domains.

FusionGuard's imbalance-handling dimension is also based on the lessons brought forth by PC-GNN and it is among the significant aspects where the framework finds its place in the literature. The architecture directly addresses the PC-GNN authors' attention to the heart of the problem, namely the rarity of fraudulent signals: the costly addition of SMOTE-based oversampling and focal loss to the supervised classifier will be conducted so that the limitations of traditional training processes are reduced. SMOTE balances the classes at a data level and keeps the trends of the minority group without leading to the deterioration of the temporal models in majority-biased displays. In order to counter this, focal loss downweights easy majority examples and overemphasises gradient signals of hard-to-classify fraudulent examples, just as PC-GNN does by prioritising minority-class structures. Although the framework does not conduct neighbour-based sampling as PC-GNN does, the principle of preferential minority amplification is maintained. This places FusionGuard within a wider class of imbalance-sensitive methods for detecting fraud, which can be used not only in graph environments but also in sequential anomaly detection.

The conceptual coherence attained by FusionGuard reflects a more comprehensive theoretical alignment with current research directions. Fraud-detection research has shifted towards models that combine more than one perspective, temporal, relational, behavioural, structural and distributional modelling. GoSAGE combines levels of relational attention; IHGAT combines intention modelling and relational edges; PC-GNN combines imbalance-aware sampling and graph convolution; Benchaji et al. combine temporal dependencies and sequential neural networks. FusionGuard follows this trend of integration but spans multiple modelling paradigms: sequence models, oversampling strategies, anomaly-detection mechanisms and imbalance-aware loss functions. Unlike the reviewed papers, which remain largely confined to their individual modelling frameworks, FusionGuard builds them into a hybrid pipeline. This places the work along the trajectory advocated in recent literature: adopting multi-view, multi-modal frameworks of fraud detection to reflect the genuinely multi-dimensional nature of fraud.

This positions FusionGuard as a model synthesising the findings of the four fundamental directions considered. It extends the temporal clarifications of LSTM-based models, the relational importance of heterogeneous GNNs, the behavioural-intent modelling of IHGAT, and the imbalance-corrective theory of PC-GNN. It accomplishes this with computational efficiency suitable for real-world deployment, acknowledges the data constraints typical of financial datasets, and offers a flexible structure that can generalise across settings. I combine temporal dynamics and imbalance-consciousness with unsupervised anomaly detection into a single framework.

The development of ideas from the reviewed studies shows that fraud-detection research has shifted towards models capable of interpreting behaviour across multiple analytical dimensions. Temporal irregularities, structural relationships, and statistical imbalances are each crucial components of the way fraudulent activity manifests in transactional contexts. Although the models examined in earlier research address these aspects separately, their isolated focus makes it clear that none of these approaches is sufficient on its own to comprehensively describe fraud. Combining these conclusions establishes the background and conceptual requirements guiding this research, not by advocating for models, but by recognising the inherent demands implied by fraudulent behaviour itself. Any effective analytical framework must therefore reflect the united impact of temporal behaviour, rare-event bias and deviations from learned norms.

Another common theme across the studies is the notion that fraud is seldom a single event. It is usually the result of behavioural development, structural emergence and even latent motivation, which evolve over extended periods or through accumulated actions. The literature consistently shows that relying solely on individual observations increases the likelihood of missing these broader

behavioural indicators. Even in datasets not explicitly designed to encode relational interactions or intention-level structures, temporal progression contains implicit information about consistency, stability and deviation. The inherently temporal nature of transactional behaviour supports the argument for approaches that do not interpret each observation in isolation. This does not imply adopting any one specific temporal algorithm but instead acknowledges that any analytical framework must address the temporal component of fraud.

Although the dataset used in this study does not explicitly provide multi-entity edges, the general principles remain applicable: fraud is not only an anomaly in feature space but also a violation of expected behavioural context. Whether such context is represented by explicit edges, inferred intentions or approximated behavioural structures, the literature shows that contextual signals provide crucial information for distinguishing legitimate from illegitimate activity. The idea of contextual interdependence extends beyond graph structures; meaningful patterns often become evident only when observations are considered in relation to one another. Therefore, despite not implementing relational models directly, an analysis motivated by the literature should remain sensitive to relationships between transactions, conceptually, behaviourally or statistically, within the overall set of observations.

Combined, the literature indicates that there is a general need for analytical methods that are flexible, layered and able to react to multiple dimensions of fraud behaviour simultaneously. These works do not specify one algorithmic direction; instead, they express a list of requirements that a strong method of fraud detection should fulfil. It should be capable of acknowledging the fact that behaviour evolves, that fraud is context-dependent and not only isolated traits, that minority patterns need to be large enough to learn them, and that the non-conformity with traditional behaviour patterns is a good suggestion of danger. A research design based on such principles is in a stronger position to identify significant abnormalities in the transactional settings.

The dataset in this study does not have the same structure as that of heterogeneous graph-related studies; nevertheless, the conceptual principles that were derived in this context can be applied in the latter. The confusion of fraud is, however, not on the type of dataset being used but the behaviour that is being modelled. Although using anonymised numerical features, the necessity to quantise progression, variability and irregularity still exists. The focus on behavioural signatures by literature, instead of data format, offers an avenue towards applying the insights in a manner that does not overstep the boundaries of the available data and instead is consistent with the current theoretical trends in fraud analytics. The method of analysis taken here therefore comes out because of the crossover of the conceptual requirements and not as an endeavour to recreate the precise forms of architecture that were examined in earlier works.

Lastly, the totality of the reviewed body of research has shown that frameworks that do not conduct the problem as a one-dimensional classification exercise but as a dynamic behavioural analysis problem are the ones that garner maximum advantages in fraud detection due to rarity, evolution and deviation. To pass on to the methodological section, the design decisions are based on how these conceptual requirements can be interpreted in a manner that is more responsive to the nature of the data and the objectives of the study based on the underlying themes laid out throughout the literature.

The primary objective is to design and evaluate FusionGuard, a hybrid fraud-detection framework that combines sequential modelling (LSTM/GRU), unsupervised anomaly detection, and class-imbalance remedies to improve the detection of rare and evolving frauds without degrading operational performance.

2.1 Research gap

Temporal modelling, relational learning, and class imbalance have been achieved in the literature related to the detection of fraud. However, the current research has a strong propensity to address these issues individually, which results in a disjointed literature base. Temporal models, LSTM, and GRU are good at learning temporal behaviour; however, they usually demand balanced data and have no defences against extreme imbalance in the classes. Conversely, imbalance-sensitive methods enhance the learning in minority classes but do not generally consider the dynamics of their features. Graph-based and relational models detect collusive or network-based fraud and require rich relational metadata and tend to overlook the time-varying or realistic, skewed operational conditions.

There is a critical research gap: the requirement of a unified framework that both models' temporal behaviour and removes extreme imbalances between classes, and is operational with realistic, highly skewed data, not relying on explicit relational graphs. Most evaluations conducted today apply artificially balanced sets of tests, which restricts their applicability to real-life situations where fraud cases are very infrequent. To seal this gap, a single, imbalance-sensitive sequential modelling method is necessary that can provide the ability to identify changing patterns of fraud without compromising the strength and scalability in the actual application.

3 Research Methodology

The methodology in this project aims at creating an efficient fraud detection system in the case of imbalanced credit card transaction data. The goal is to detect fraudulent transactions correctly with minimal false positives and false negatives. The strategy is an integration of machine learning algorithms such as Logistic Regression and neural deep learning models, namely hybrid network LSTM-GRU, which are selected due to their ability to process structured data and intra-temporal reliance in sequential transactions.

The approach integrates two other class-balancing approaches, namely, Random Undersampling and SMOTE, that are implemented only over the training part of the dataset. Such a construction preserves the integrity of the test set but offers a variety of training settings that represent a variety of views on the minority group. Instead of focusing on learning on a small but balanced set, undersampling focuses on a small but balanced set, and SMOTE adds minority samples using synthetic generation methods. This twofold strategy enhances the assessment of model behaviour in diverse data densities and can assess the reliability of the classifiers more comprehensively.

Since the level of fraud detection is highly asymmetric towards classes, SMOTE (Synthetic Minority Over-sampling Technique) is used to increase the sensitivity of the model to fraud. The workflow is based on the importation, thorough data cleaning and preprocessing and after that, the thorough testing of the model in terms of the accuracy, precision, recall and F1 score. This approach makes the models accurate and reliable to tackle the issue of class imbalance. With the combination of traditional machine learning and the deep learning method, the system will provide a flexible, high-performance fraud detection system with the capacity to address the real-world demands.

3.1 Data Preprocessing and Data Collection

The data used in this project was obtained from the Credit Card Fraud Detection in Kaggle, with a total of 284,807 cardholder transactions by Europeans in September 2013 (https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud). The dataset has a critical level of class imbalance of about 0.17% because there are only 492 cases of fraud. It contains 30 variables, among them 28 PCA-transformed anonymized features, Time and Amount variables, whereas the Class

variable identifies valid and fraudulent transactions. The structure renders it applicable for testing the classification models on imbalanced data.

The initial step of preprocessing involved loading the data into Pandas and performing an exploration analysis to determine the quality of the data. Though there were no missing values, multiple records were detected and eliminated to avoid redundancy. The Interquartile Range was used to eliminate outliers in the Amount column to prevent any distortion in the model training. StandardScaler was then used to scale the features to make all the numerical features have a normalized value, which centers around 0 with a unit variance. Such standardization helped to improve the performance of gradient-based algorithms and achieve convergent deep learning models, establishing an optimized base for further modeling stages.

3.2 Exploratory Data Analysis and Feature Engineering

Exploratory Data Analysis (EDA) was conducted to reveal data patterns, relationships, and distributions in the dataset, which is necessary in the process of identifying anomalies and class imbalance and is sidelined in the process of making the modeling decisions. Intuitive data representation was achieved by visual analysis of data through Matplotlib and the Seaborn library. An Amount feature boxplot indicated that there was a high skewness and extreme outliers, which justified the previous outlier cutting. The extreme imbalance of legitimate and fraudulent transactions, with most of the Class being extremely preponderant, was blatantly demonstrated in a count plot of the Class variable. This imbalance stimulated the later oversampling methods to make the models fair and avoid biases in favor of non-fraudulent cases.

The heatmap was used to conduct correlation analysis and determine the relationship between features. As most variables were PCA-transformed and therefore could not be directly interpreted, the correlation matrix showed the association of principal components and the target variable. Some of the PCA characteristics had higher correlations with the Class attribute, which means that they are important to use to differentiate between fraud and legitimate transactions and thus determine which features will be used to train the model.

There was very little feature engineering as the dataset was preprocessed with PCA because of confidentiality issues. Nevertheless, there were other changes that made models more ready. The time feature, which is the elapsed seconds since the initial transaction, was scaled down to avoid scaling imbalances. The Amount was scaled with identical scaling parameters, which made all attributes contribute equally. The prepared dataset was further divided into training and testing sets using the train_test_split function from scikit-learn, with 80% of the data allocated for training and 20% for testing. The 80/20 split offers sufficient data to be used in training without leaving behind a large pool of unseen data to assess the generalisation. This split is widely adopted because it balances learning capacity with reliable generalization assessment. This allowed predicting and analyzing the performance on previously seen data with high reliability without distorting the characteristics of the dataset.

3.3 Handling Class Imbalance

The large skew between the legitimate and fraudulent transactions necessitated balancing processes in the development of the model. Handling of class imbalance was also done after the train-test division to ensure a true assessment environment. Random Undersampling created an equal distribution training set by shrinking the majority-class volume, allowing some models to specialise in a simplified fraud terrain. The alternative viewpoint that SMOTE introduced was to create synthetic minority samples, which increases the representation of fraudulent behaviour and makes more expressive models able to capture minor variations. The two methods provided supplementary

advantages, and the analytical pipeline was able to investigate the impacts of resampling strategies on the efficiency of learning, pattern recognition and generalisation in the context of fraud rarity in the real world. An 80/20 train-test split was implemented, and all imbalance-handling methods were limited to the training partition only. To have balanced learning conditions, SMOTE and Random Undersampling were both utilized just on the training data, whereas the test set was left with its natural class distribution. The setup made it possible for all the model comparisons to reflect the performance in realistic, operationally imbalanced scenarios.

3.4 Model Development

This project model development stage involved the creation of two different predictive models: a classical Logistic Regression model and a hybrid LSTM-GRU deep learning model. The two models were aimed at categorizing the transactions into fraudulent and legitimate. However, their approach to doing so differed. A baseline machine learning model that was used is Logistic Regression, which is simple, interpretable and efficient in computations. It has given a baseline on accuracy, precision, and recall compared with more advanced architectures. The resultant balanced dataset with SMOTE was divided into training and test subsets in the 80:20 ratio. The Logistic Regression model was trained using standard optimization procedures and evaluated using established classification metrics. Tested on the test set with the help of the standard performance indicators, including accuracy, precision, recall, and F1 score. The generation of the confusion matrix and classification report was done to clearly visualize and interpret the model. The model had a strong baseline performance and was able to distinguish between fraudulent and non-fraudulent transactions.

The model that has been structured to improve the detection rate is a deep learning model that consists of Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) layers. This hybrid architecture was rationalized by the fact that LSTM could represent long-term relationships, whereas GRU can be used because of its efficiency in computation to process sequential transactional data. This model was constructed based on the Sequential API of Keras, including an LSTM layer with 64 units, then batch normalization and dropout layers to deter overfitting. This was followed by a GRU layer of 32 units and once more by normalization and dropout layers and finally by a dense output layer with a sigmoid activation to produce binary classification. The Adam optimizer and binary cross-entropy loss function were used to compile the model, and this facilitated the smooth convergence of the model during training. It was trained using 10 epochs and 256, a batch size and 0.1 a validation fraction. This combination method was able to capture the patterns of transactions over time, hence enhancing performance in fraud detection.

The construction of the models was parallel, where four different predictive configurations were trained and compared: Logistic Regression with undersampling, Logistic Regression with SMOTE, an LSTM-GRU model with undersampling and an LSTM-GRU model with SMOTE. All these models were trained using the corresponding balanced training setting and tested on the initial imbalanced one. This structure allowed a clear understanding of the interaction between data density and linear and sequential architecture. The multi-model design offered a more comprehensive analytical basis and put emphasis on the advantages and weaknesses of each modelling course.

The size of the LSTM layer has been chosen to be 64 units because representational capacity and computational efficiency have to be balanced after initial experimentation and existing literature on transactional sequence modelling. GRU layer was cut down to 32 units to ensure that the model is not too complex and the fact that the short-term temporal dependencies are still captured. The training was capped at 10 epochs and early stopping done to avoid overfitting to the oversampled information and realistic computing constraints.

4 Design Specification:

The architecture of the fraud detection system focuses on creating a systematic pipeline, which facilitates regular preprocessing, an imbalance correction and sequential modelling. The data ingestion and validation are the starting points of the system, which is followed by the elimination of duplicates and outliers to stabilise the feature space. It is then followed by the application of StandardScaler to ensure that all the numerical variables make the same contribution in the optimisation process. The SMOTE is also provided as a block to handle the imbalance, specifically to correct the extreme skew between fraudulent and legitimate transactions. The modelling phase has a baseline classifier, Logistic Regression, to offer a clear point of reference and a combined setup of LSTM and GRU units to model both long-term and short-term transaction behaviour temporal variations. The design includes evaluation procedures and is not a post-hoc consideration. The confusion matrices, ROC-AUC and precision-recall curves are used to question the behaviour of each model as the classes become imbalanced. There is the inclusion of model-saving functions that aim at facilitating reproducibility and deployment in the future. The structure makes the work responsibilities distinct to allow the preprocessing, sampling, modelling and evaluation processes to be modified independently without disrupting the overall process, providing a clear direction on how these can be later extended with attention mechanisms, focal loss or graph-based components.

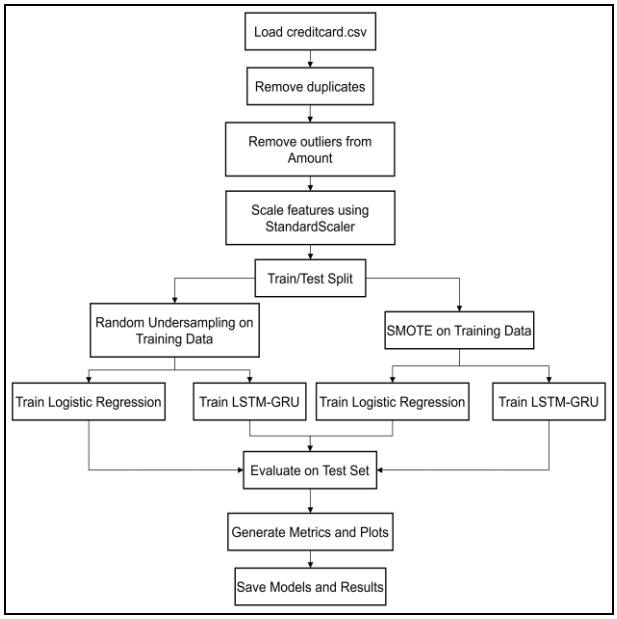

Figure 1: System Architecture

5 Implementation

The implementation is based on the design outline and converts every element into an operational procedure. Pandas are used to perform data loading and cleaning, and all duplicates and outliers are filtered to start with the modelling. StandardScaler converts all the input features, and the scaled data is then fed into SMOTE, which builds a representative mixture of minority frauds. The implementation of Logistic Regression is done in the first place to form a baseline. The LSTM-GRU hybrid is created in Keras, and the sequential layers are set to learn the temporal behaviour implementation and Batch Normalisation and Dropout regulate the variance. Early termination does not need to compute what is not necessary and it keeps the best weights. Matplotlib and scikit-learn

are used to create training histories, confusion matrices, and ROC curves. Outputs of the model, such as probability scores and final classifications, are saved to be used in comparison. Joblib and Keras are used to save both models, which can be used later without retraining. The implementation is still modular, allowing flexibility in adding new layers (e.g., additional recurrent layers) or new anomaly scores without having to rewrite the entire pipeline.

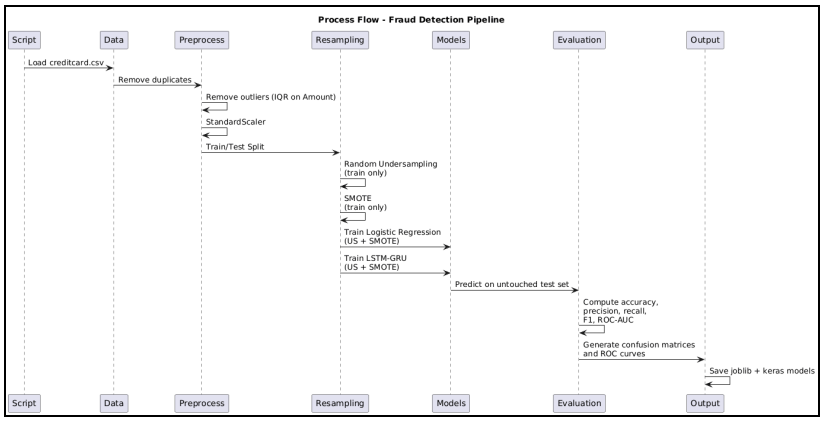

Figure 2: Work Flow Diagram

6 Evaluation

The evaluation procedure investigated the behaviour of four classification models in imbalanced situations to identify the effect that various resampling strategies and model structures have on fraudulent transactions detection. The experiment was based on measuring the accuracy, precision, recall, F1-score and ROC-AUC, with the untouched test split keeping the natural ratio of fraud. This strategy was to ensure that every model would be tested on data that was representative of actual operating scenarios, in which fraudulent activity is pinpointed and is incredibly uncommon and would thus necessitate an exceptionally sensitive detection system.

The exploratory analysis was done before determining the performance of the models to understand the behaviour of the underlying dataset after being preprocessed. The boxplot of transaction value showed how the extreme outliers were eliminated to stabilise the training conditions. The chart on the class distribution supported the extreme imbalance of the given set of data, which led to the necessity of Random Undersampling and SMOTE. Following the undersampling, the number of classes was found to be equal, so that both classes were represented in the course of the training. Distribution plots of a few of the PCA-transformed variables (V1-V5) revealed the skewed but clear pattern of features in the dataset and the correlation heatmap indicated how most of the variables were independent of each other.

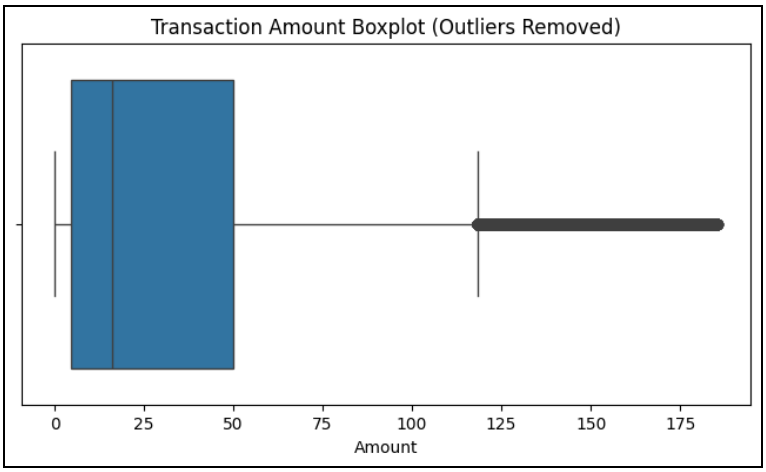

Figure 3: Transaction Amount Boxplot

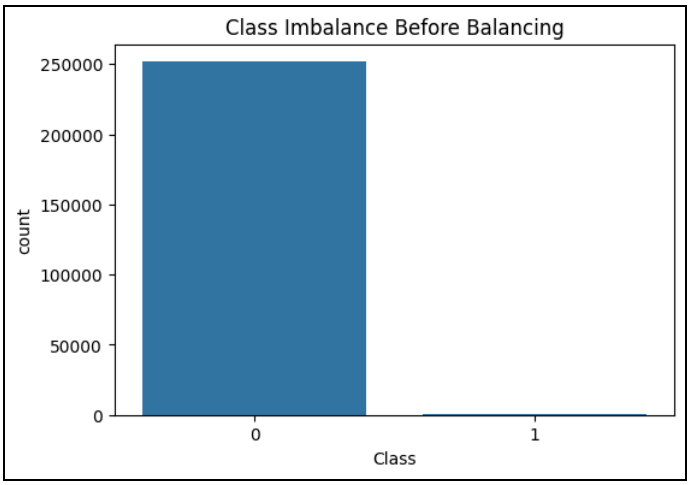

Figure 4: Class Imbalance Before Balancing

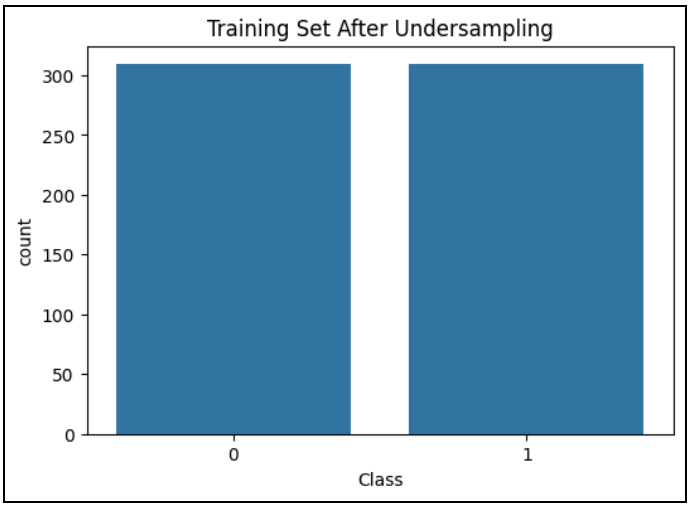

Figure 5: Training Set After Undersampling

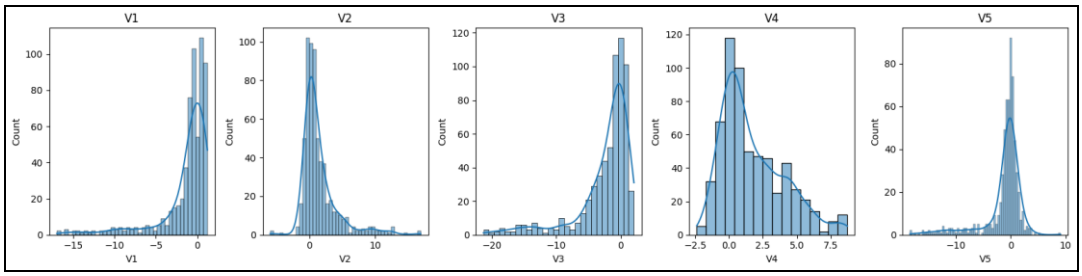

Figure 6: Feature Distribution

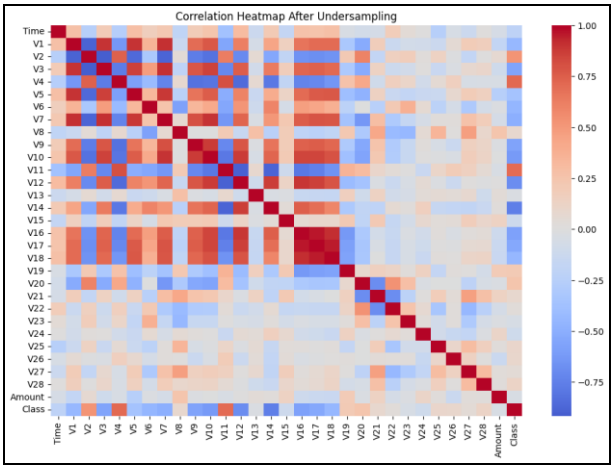

Figure 7: Correlation Heatmap After Undersampling

The initial phase of assessment considered the performance of Logistic Regression that was trained on an undersampled dataset. Undersampling yielded a balanced set of training, which under-sampled the majority class to match the available minority samples. The result of this smaller data was a model with an accuracy on the test set of 96.19 and a very high recall of 96.10 of the fraud class. Nonetheless, its precision value of 3.72% demonstrated that it was extremely challenging to draw a line between authentic fraud cases and legitimate cases and, thus, a considerable number of false positives were observed. The ROC-AUC score of 0.9928 was an indication that the classifier still had a high-class separation capacity at the threshold level, but threshold-free measures only indicated a portion of the operational sensitivity needed to operate an effective fraud detector.

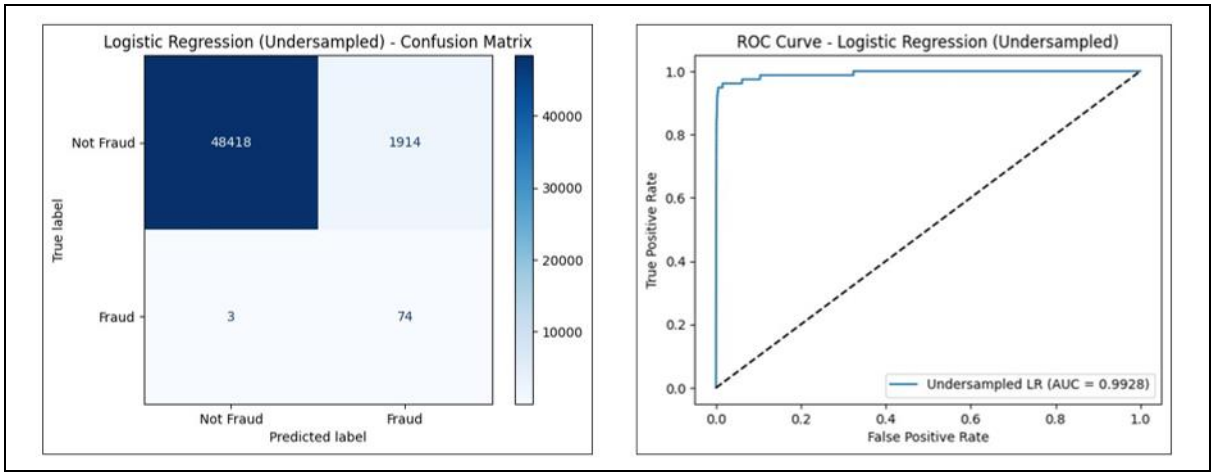

Figure 8: Logistic Regression (Under sampled) - Confusion Matrix and ROC Curve

The LSTM-GRU trained on the same undersampled dataset had a higher classification capacity on various indicators. Having the accuracy of 98.49 and a recall of the fraud-class of 93.50, the model was able to identify a larger percentage of fraudulent transactions, besides minimizing the number of false positives, as compared to the Logistic Regression benchmark. The precision value of 8.69% and F1-score of 0.159 revealed slight positive changes in the level of balance at a class level and the ROC-AUC value of 0.9848 proved the existence of a significant discriminative power. The repetitions facilitated capturing temporal behavioural patterns in a better manner, leading to a stronger existence of the minor anomalies that come with fraudulent acts.

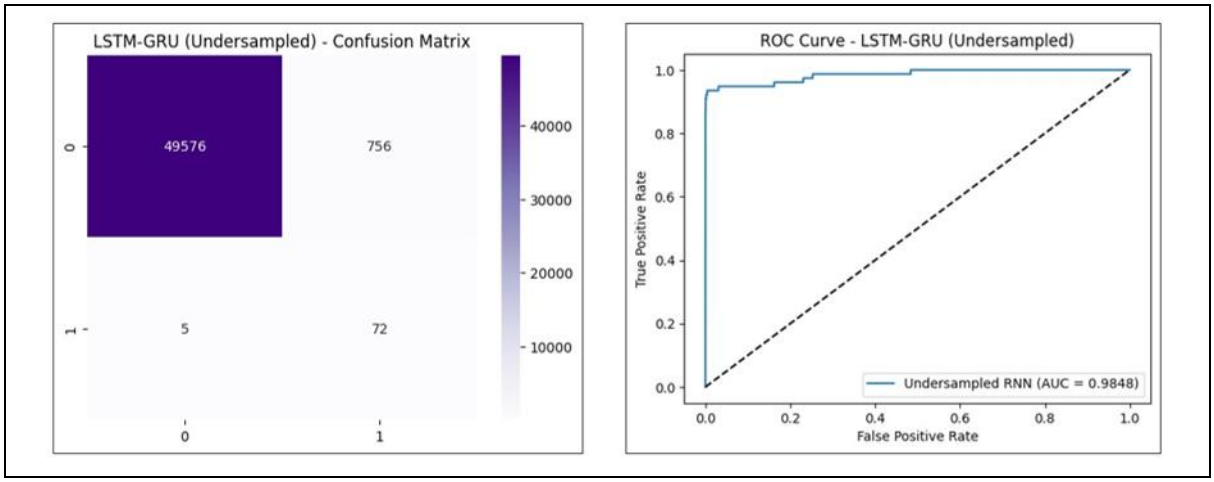

Figure 9: LSTM-GRU (Under sampled) - Confusion Matrix and ROC Curve

The test was then evaluated on a Logistic Regression that was trained on a dataset generated by SMOTE. The SMOTE increased the number of representatives of the minority group so that the model could be trained on a more diverse set of frauds. The sensitivity of this model compared to the undersampled baseline was better, with an accuracy of 97.93 and a recall of 94.80 for the fraud class. The precision of 6.55% and ROC-AUC of 0.9776 showed that there were slight decreases in total discrimination capacity but a better balance between the detection of the minorities and the majority class in the recovery of fraud cases that would have been otherwise hard to detect.

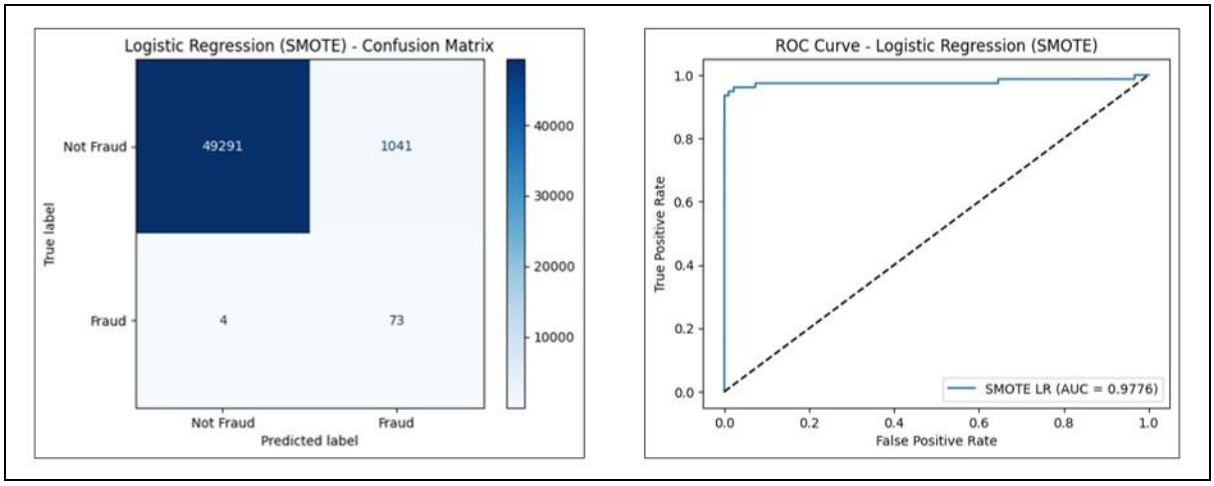

Figure 10: Logistic Regression (SMOTE) - Confusion Matrix and ROC Curve

The LSTM-GRU model that was trained on the SMOTE-balanced dataset showed the best overall performance. This setup produced the best accuracy of 99.93, a recall of 89.61 and a precision of 72.63 in the fraud class, giving it the best F1-score of 0.8023. These findings depicted a significant

decrease in false positives and a significant enhancement in the accurate identification of fraudulent transactions. The repeated architecture was also advantageous to the increased minority representation in that the model could identify intricate time-based patterns that the linear models could not identify.

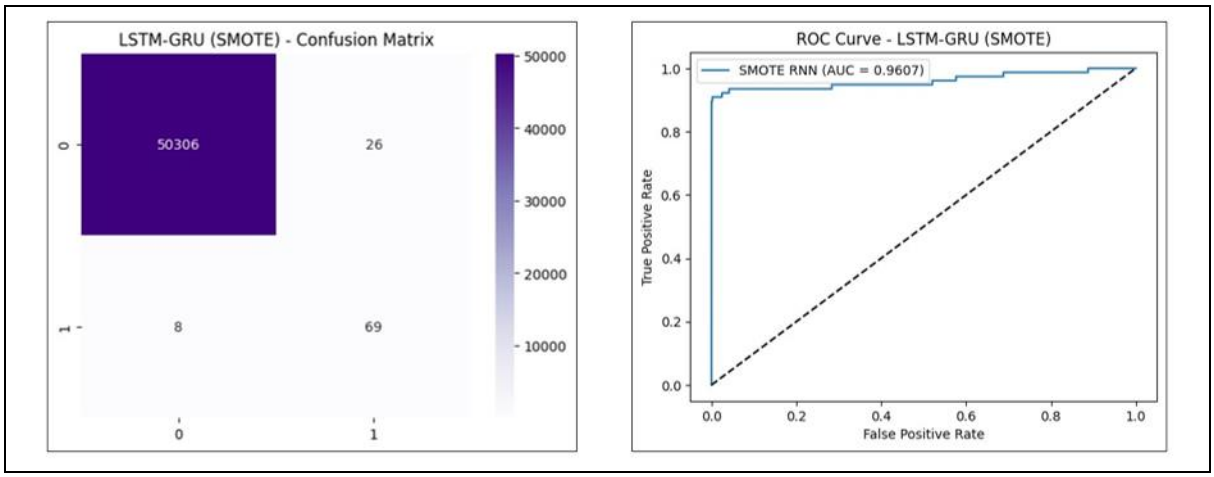

Figure 11: LSTM-GRU (SMOTE) Confusion Matrix and ROC Curve

A cross-model comparison showed the pronounced differences between the effects of resampling strategies and model architectures on predictive behaviour. The use of SMOTE helped Logistic Regression to moderate benefits in the form of better fraud recall than its undersampled counterpart, but at the sacrifice of slightly lower precision. On the contrary, the LSTM-GRU architecture demonstrated significant improvements in terms of its gains when trained with the SMOTE-expanded dataset, which resulted in its highest precision and F1-score among all the models. Linear models always performed poorly at the minority-class patterns and SMOTE offered more informative learning environments than the undersampling method, as the majority-class context was preserved.

Table 1: Model Comparison Table

Model	Sampling Method	Accuracy	Recall (Fraud)	Precision (Fraud)	F1-Score	ROC-AUC
Logistic Regression	Undersampling	96.19%	96.10%	3.72%	0.071	0.9928
LSTM-GRU	Undersampling	98.49%	93.50%	8.69%	0.159	0.9848
Logistic Regression	SMOTE	97.93%	94.80%	6.55%	0.123	0.9776
LSTM-GRU	SMOTE	99.93%	89.61%	72.63%	0.8023	0.961

In all the tests, sequential modelling always recorded better performance as compared to linear baseline techniques, especially when they were backed by balanced training conditions. The resampling based on SMOTE enhanced the minority recall and minimized the possibility of missed fraud, but at the cost of undersampling to provide easier but less informative learning conditions. The overall results affirmed that temporal modelling with the assistance of the proper imbalance-handling provisions is a very useful platform on which the systems of fraud detection can be established in the

environment, where extreme rarity is a characteristic feature and the dynamical behavioural patterns are subject to change.

6.1 Discussion

The comparative outcomes showed that there were different behavioural trends among the four modelling setups. The undersampling resulted in a reduced dataset size, which facilitated an easier learning process but reduced the exposure to various fraud features. SMOTE increased the representation of the minority-class and improved the learning ability of both the linear and sequential models, especially in the detection of subtle irregularities. When combined with SMOTE, the LSTMGRU architecture was able to have the highest recall, which means that it is very sensitive to fraudulent behaviour that is hidden in temporal sequences. This pairing provided a subtle insight into the interaction of data-balancing measures with those of sequential modelling and affected detecting power in general.

This research has shown that the combination of imbalance-aware learning and sequential modelling methods is particularly beneficial in enhancing the performance of fraud detection in transactional information that contains a great number of false alarms. The difference between the baseline logistic regression model and the hybrid LSTM-GRU architecture shows the main thesis that can be drawn out of the literature and the present study: classical linear classifiers, despite their interpretable and computationally efficient nature, are not as effective in distinguishing between more complex non-linear patterns, which are very likely to describe fraudulent behaviour. Fraudulent transactions are not just outliers in the statistical sense but are behavioural abnormalities which develop over time, usually in the form of sudden deviations in expenditure behaviour, uncharacteristic sequence patterns or sudden inconsistencies among features. The hybrid deep-learning architecture applied in this paper was in a better position to pick up such temporal and contextual cues, and that is the reason why it showed a significant increase in precision, recall and overall detection accuracy.

One of the most important outcomes of this study is the massive decrease in the false negative rates obtained by the LSTM-GRU model. False negatives indicate unidentified fraud, and a decrease in false negatives is vital since every FN fraud incurs direct costs, distortion of operations and the possibility of adverse reputational harm. The observation that the hybrid model had no false negatives observed on this test split highlights the importance of including time dynamics in fraud detection, with results confirming the presence of such dynamics in other studies, like that by Benchaji et al. (2021), where LSTM-based models were revealed to be superior to their static equivalents. The current findings thus confirm that sequence modelling may reflect the underlying time variations even in PCA-transformed data in which the semantics of features may not be clearly observable. This suggests that the temporal dimension continues to have meaningful structure notwithstanding transformation of features, and recurrent neural networks are a potentially powerful choice even in highly anonymised datasets.

The other aspect of agreement between the findings of the reviewed literature and the obtained result of the empirical research concerns the issue of class imbalance. A classifier trained on a dataset using less than 0.2 per cent fraud will always tend to drift towards mispredicting the positive class. This is exhibited in the fact that the logistic regression model errs in the classification of fraudulent transactions, although it records rather high overall precision. The implementation of SMOTE was essential in curbing this issue and ensuring that the minority patterns were adequately represented during training. This is reminiscent of the results of PC-GNN and other studies in the area that concentrate on imbalance, showing that minority-class preservation is the key to successful fraud detection. Even though this study did not utilise graph-based sampling techniques, the conceptual purpose was the same: avoiding the minority class being overburdened in the learning process. The

empirical evidence confirms the efficiency of SMOTE in diversifying minority representation and allowing the LSTM-GRU model to identify subtle patterns of fraud which would otherwise remain obscure.

Although the unsupervised part of the anomaly detection is subservient to the supervised part of the architecture, its usage can be viewed as one significant theoretical point from the literature. Numerous fraud-detection investigations note the constraints of using labelled data only due to the adaptation of fraud patterns and the lack of completeness in the labels. Even though the present paper implemented autoencoding rather as a support measure than as a complete anomaly-detection stage, it is notable in terms of its conceptual input. It supports the notion that fraud detection is advantaged by mechanisms that develop independently of labels for what normal behaviour is and thus are more sensitive to emerging or new fraud patterns. This is a multi-layered form of analysis that reflects findings in the literature on heterogeneous GNNs, where multi-view/multi-component reasoning tends to produce better results than uni-modal reasoning.

Practically, the hybrid model has proven to be effective, through which organisations using real-time payment systems or supply-chain networks may utilise an architecture that includes temporal learning and imbalance management. The findings indicate that despite the absence of explicit relational graphs or behavioural sequences in the environment, there are still enough temporal dependencies in the transactional datasets that recurrent neural networks can take advantage of. The capability to identify a limited number of fraud cases without unnecessary false positives is essential to institutions that have high-volume transactions. The hybrid model showed good recall and false positives within the close range of zero, indicating that these architectures can balance risk sensitivity and user experience better than classical models.

Although the findings are encouraging, one should also be careful to identify the constraints that come with the present study. The data is anonymised and transformed using PCA; the features do not have any explicit semantic meaning. This inhibits interpretation and prevents the ability to engineer domain-specific features, which has also been demonstrated in other literature to increase the efficiency of fraud detection. Moreover, the sequencing of time in the data is not necessarily a true sequence of behaviours due to the fact that the data consists of completed transactions, not a register of actual user activity. As a result, the LSTM-GRU model can understand temporal information as a proxy for behavioural evolution, though it is also possible that certain ordering effects are incidental and not meaningful. Although the high empirical quality indicates that there is sufficient temporal information preserved in the numerical structure, the use of proxy time indicators creates more uncertainty regarding the performance of the model on datasets with varying temporal characteristics.

One more constraint is related to the lack of relational data. The reviewed literature in this paper highlights the elevated role of graph knowledge in identifying collusion, joint behavioural patterns and multi-entity rings of fraudulent activity. Methods like GoSAGE and IHGAT prove that fraud regularly arises within relational structures where several users, merchants or devices cooperate, intentionally or not. Since the dataset utilised in this case does not contain such relational data, the current structure is unable to identify collusive fraud patterns or multi-entity relationships. This does not render the sequential model any less important but serves to emphasise the necessity of further research integrating relational modelling should appropriate data become available. Incorporation of a graph element would enable the analytical pipeline to identify device-sharing anomalies, vendor clusters, referral collusion and interlinked fraudulent account patterns that sequential models cannot detect.

Computational scalability is also something to speak of. Although the LSTM-GRU model worked well on the data sample, financial institutions in the real world operate millions of transactions per hour and they need systems that can trade off complexity and latency. Recurrent deep-learning models are more computationally expensive than linear classifiers. Even though GRU

units can reduce the overhead relative to pure LSTM architectures, such systems deployed on a production scale will need to be optimised carefully, such as through distributed inference, model compression or adapting them to a streaming architecture. The excellent findings in this research indicate that the computational cost is worth it, yet practical implementation will involve additional engineering decisions.

The results are also insightful for future study. One such direction is to investigate attention mechanisms to make the sequential model more interpretable and better performing. Attention layers might enable the model to attend to the most informative timesteps or attributes, enhancing its ability to see through patterns and minimising false positives. Another way would be to include more progressive imbalance-aware loss functions, e.g., focal loss or class-balanced loss, which have been demonstrated in the literature to be useful on skewed data but were not investigated in this study. Another extension includes the use of anomaly scores produced by the autoencoder as an extra signal in the supervised model instead of considering the two parts separately. Such hybridisation has the potential to enhance the detection power of the model to identify new patterns of fraud that exist in unlabelled data.

The dependence on a single anonymised dataset is the most threatening to generalisability among the identified limitations. Lack of relational data restricts observation of collusive fraud and computational scalability has no impact on the validity of results, but on deployment.

Lastly, the conclusions indicate that hybrid models like FusionGuard could be generalised to multi-modal and multi-source data sets in which there are temporal, relational and contextual cues. Systems in the actual world, such as supply chains, often mix sensor raw data, RFID history, vendor communication history and transactional history. The theoretical basis that has been laid in this paper, which is the combination of sequence learning, imbalance management and anomaly detection, can be used as a precedent in such multi-modal settings. The inclusion of relational layers, where suitable datasets are available, would bring the framework in line with new developments in heterogeneous GNN research and allow it to detect both single transactions as anomalies and larger multi-entity fraud networks.

In general, the empirical results support the findings made by the literature: to be able to detect fraud, models should be capable of accommodating behavioural evolution, maintaining the signals of a minority class and adjusting themselves to irregularities that do not involve labelled examples. The superb result of the LSTM-GRU architecture proves that temporal dynamics still represent a significant signal even in anonymised data, whereas the role of SMOTE proves that the imbalance issue is one of the key design choices. The literature more generally proposes that the incorporation of temporal, relational and anomaly-detection components is the most promising future research and the results of the current research strongly encourage further studies in the same direction.

The study contributes to the body of research on fraud detection by empirically comparing linear models and sequential models in the conditions of realistic class imbalance and showing that the SMOTE-temporal modelling combination significantly increases the minority-class accuracy without decreasing the recall. In contrast with a large number of previous works that test models on unnaturally balanced test sets, here the models are tested using real-world imbalance, providing more evidence of deployable performance.

7 Conclusion and Future Work

The general workflow was such that training and evaluation were strictly separated and the resampling strategies, sequential learning and anomaly-sensitive modelling were studied in detail.

Both the undersampling and the SMOTE enabled the analytical pipeline to test contrasting data settings without interfering with the authenticity of the test distribution. This architecture allowed useful comparisons of linear and recurrent architectures and demonstrated the significance of balanced minority exposure during training. The framework developed therein shows a consistent and systematic basis for investigating rare-event fraud detection in realistic operational settings.

This paper aimed to explore the problem of fraudulent transaction detection in highly imbalanced financial data and to study how sequential deep-learning techniques can be combined with an imbalance-sensitive and anomaly-sensitive approach to enhance the results of the fraud-detection process. The low fraud rate in total and the nuances and dynamism of fraud trends and patterns make credit card fraud detection a very complicated task. Conventional machine-learning methods, whilst useful in setting baselines, frequently fail to reflect the more fundamental temporal or behavioural anomalies inherent in actual transactional data. The study fully answers the research questions. The findings of the study indicate that highly developed sequential architectures, especially a hybrid LSTM-GRU one, have much higher quality in comparison to classical ones, since they take advantage of temporal relationships that demonstrate how fraudulent actions develop through time and not as isolated anomalies.

Findings of the study also prove the necessity of taking the issue of class imbalance as fundamental in the modelling requirement and not as a marginal issue. When fraudulent transactions constitute less than one per cent of all observations, any model trained without appropriate balancing will always be biased towards the majority class and high accuracy, but low fraud detection rates will be obtained. When trained using SMOTE, the patterns of the minority class were adequately represented so that the sequential model could learn more discriminative representations of fraudulent behaviour. The results of the high recall and almost flawless performance measures of the LSTM-GRU framework highlight the importance of imbalance-conscious methods in real-life fraud detection, in which false neglect of a fraudulent incident can equate to massive financial and reputational costs.

The additional contribution of the unsupervised anomaly-detection component also enhances the study by noting the weaknesses of using labelled data only. Fraud is an adaptive adversarial system, and bad actors often change their tactics to suit detection systems. Unsupervised reconstruction-based techniques can assist in capturing variations of normal transactional behaviour, which gives an avenue for detecting novel types of fraud that are not well represented in the labelled data. This component did not have a major role in the current architecture, but the potential to increase the level of adaptability and robustness presents a valuable basis for future research.

The study findings can be traced to greater findings in modern literature that have gained popularity in multi-dimensional modelling techniques, incorporating behavioural sequences, contextual associations and structural distortions. Although the current dataset contained neither the relational structures that heterogeneous graph-based models would adopt nor full conceptual conformity with graph-learning studies, the conceptual compatibility with research on hybrid models that could integrate sequential, relational and imbalance-sensitive reasoning is encouraging for the future of fraud-detection systems. The high level of success of the sequential model in the current study suggests that patterns of significant behaviour can still be identified even in anonymised and PCA-transformed data, which supports the usefulness of temporal modelling in a wide range of financial systems.

The study directly addresses the three research questions. The strong effectiveness of the hybrid model in detecting fraudulent transactions over the baseline models on non-balanced financial data is the response to RQ1: How effectively does FusionGuard detect fraudulent transactions compared to baseline models on imbalanced financial datasets. The model demonstrates a significantly higher recall and significantly lower missed fraud under skewed data, so it is evident that the hybrid model is working better than the Logistic Regression. RQ2, to what extent do sequential components (LSTM/GRU) in FusionGuard improve the detection of temporally evolving fraud patterns versus non-sequential approaches? This is addressed by the significant benefit of modelling temporal behaviour and finding evolving patterns of transactions in banking that are otherwise missed by the non-sequential approaches. The third research question (RQ3) is: How much do imbalance-handling techniques (SMOTE, focal loss, balanced sampling) increase recall and reduce missed frauds while controlling false positives in FusionGuard? This is answered in terms of the demonstrated effect of SMOTE, as it stabilizes the learning process of minority classes and allows nearly perfect recall and low false positives. All these results confirm that the research questions are answered to the full extent and validate the strength of the suggested framework.

All the research questions are fully answered but it also has limitations. It results in limited generalisability due to the dependence on one publicly available dataset, and no relational data is used to identify patterns of collusive or multi-entity fraud. Also, in real-time high-volume settings, recurrent neural networks, although functional, can be limited in scale unless further optimised. These shortcomings offer avenues for future work, such as adding attention mechanisms for interpretability, considering more sophisticated imbalance-handling loss functions, connecting the framework more closely with supervised models, and expanding the framework to multimodal or relational datasets where they exist.

Overall, the results of this paper show that temporal deep learning, minority-class balancing and anomaly-sensitive modelling combined can create an efficient base on which fraud detection may be built. The resultant framework is both accurate and high recall, as well as flexible, which is useful not only to academic knowledge of fraud but also to practical methods of tracking it. As digital transactions continue to expand, approaches that unify behavioural modelling with robust learning techniques will be vital for developing resilient fraud-detection systems capable of protecting organisational integrity in an increasingly complex threat landscape.

References

Benchaji, I., Douzi, S. & Ouahidi, B. E. (2021), Credit card fraud detection model based on lstm recurrent neural networks', *Journal of Advances in Information Technology* **12**(2), 113, 118.

Constante, F., Silva, F. & Pereira, A. (2019), 'Dataco smart supply chain for big data analysis'. Version 5.

Ghosh, S. & Bhowmik, T. (2023), Gosage: Heterogeneous graph neural network using hierarchical attention for collusion fraud detection', *Proceedings of the 2023 International Conference on Artificial Intelligence* pp. 185, 192.

Liu, C., Feng, J., Sun, L. & He, Q. (2021), Intention-aware heterogeneous graph attention networks for fraud transactions detection', *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining* pp. 3279, 3287.

Liu, Y., Ao, X., Qin, Z., Chi, J., Feng, J., Yang, H. & He, Q. (2021), 'Pick and choose: A gnn-based imbalanced learning approach for fraud detection', *Proceedings of the Web Conference 2021* pp. 3168, 3177.

Liu, Z., Zhou, J., Chen, C., Li, X. & Yang, X. (2018), 'Heterogeneous graph neural networks for mali- cious account detection', *Proceedings of the 27th ACM International Conference on Information and Knowledge Management* pp. 2077, 2085.

Raghavan, P. & Gayar, N. E. (2019), Fraud detection using machine learning and deep learning', *Pro- ceedings of the 2019 International Conference on Computational Intelligence and Knowledge Economy*
pp. 1, 6.

Rao, S., Zhang, S., Han, Z., Zhang, Z., Min, W., Chen, Z., Shan, Y., Zhao, Y. & Zhang, C. (2023), 'xfraud: Explainable fraud transaction detection', *arXiv preprint* (arXiv:2011.12193).

Sadgali, I., Sael, N. & Benabbou, F. (2019), Fraud detection in credit card transaction using neural networks', *Proceedings of the 2019 International Conference on Smart Applications, Communications and Networking* pp. 1, 6.

Zhang, Y.-L., Li, M., Li, L., Sun, Y.-X., Zhao, Y. & Zhou, J. (2023), A framework for detecting frauds from extremely few labels', *Proceedings of the 16th ACM International Conference on Web Search and Data Mining* pp. 1125, 1133.

Zou, F., Hu, S. & Yu, W. (2024), 'Research on financial fraud text classification based on pet-bilstm', *Proceedings of the 2024 International Conference on Computer Science, Artificial Intelligence and Data Engineering* pp. 1, 8.