Implementing Ensemble Method with stacking approach for Machine Learning and Deep Learning Algorithms for Credit Card Fraud Detection

MSc Research Project
Data Analytics

Charan Teja Marlabeedu
Student ID: X22161163

School of Computing
National College of Ireland

Supervisor: Vladimir Milosavljevic

National College of Ireland
Project Submission Sheet
School of Computing

Student Name:	Charan Teja Marlabeedu
Student ID:	X22161163
Programme:	Data Analytics
Year:	2024
Module:	MSc Research Project
Supervisor:	Vladimir Milosavljevic
Submission Due Date:	31/01/2024
Project Title:	Implementing Ensemble Method with stacking approach for Machine Learning and Deep Learning Algorithms for Credit Card Fraud Detection
Word Count:	7735
Page Count:	22

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own 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:	
Date:	31st January 2024

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):	

Implementing Ensemble Method with stacking approach for Machine Learning and Deep Learning Algorithms for Credit Card Fraud Detection

Charan Teja Marlabeedu
X22161163

Abstract

In the present world a use of online financial transactions has led to increase in risk of credit card fraud, which is major issue for consumers along with financial institutions alike. Improving approaches for detecting a fraud via combination/integration of deep learning & machine learning is major topic of our research.

This study aims to evaluate an efficacy of "stacked" approach to fraud detection by a combining several prediction models.Research is divided in 3 distinct case studies. As first one demonstrates, there is lot of promise into combining a different machine learning models, yet it can be rather difficult. Second research demonstrates that deep learning methods, namely CNNs & RNNs, are superior at detecting most typical fraud patterns. A hybrid model combining a stacked ML & stacked DL is tested into third trial. Its crucial to select & fine-tune primary model which incorporates all of models, as shown by extensive testing into thesis, even if combining multiple models might improve performance. Study represents the significant advancement into field of fraud detection, paving way for more robust & adaptable systems to ensure security of online financial transactions. The proposed models as per a research implementing the stacking approach for the machine learning and deep learning has shown the promising results with accuracy 92.472%, Precision 0.912, Recall 0.934 and f1score 0.923 than the individual models performance.

Keywords: machine learning and deep learning models, predictive analytics, fraud detections, stacking approach, meta models

1 Introduction

In many industries, digital transactions are growing at the fast pace these days. Ease of credit cards has led to their widespread usage for online & cashless purchases. The downsides and upsides of technology are the inherent to every system. As per study Rolfe (2022) IBM Financial Fraud Report, simplicity of online transactions also led to criminals taking advantage of system vulnerabilities to do fraudulent operations, especially with credit card transactions as shown in Figure 1.

Training Machine Learning & Deep Learning models utilizing current fraud transaction data is a main goal of research work in order to identify & recognize fraudulent credit card transactions. Considering huge amount of transactions happening every second, there are a several obstacles to overcome in fight against fraud without giving clients

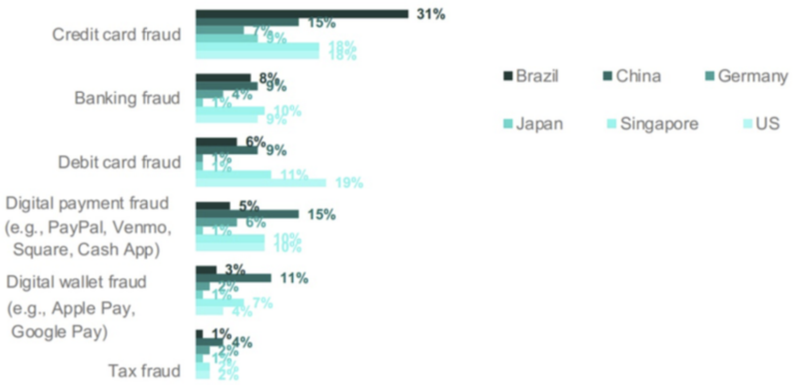


Figure 1: Financial fraud report 2022 by IBM

incorrect information. For reliable fraud detection, it's essential to train models using large and varied datasets.

Machine learning & deep learning models like RNN, Naive Bias, SVM, LightGBM, Logistic Regression, Random Forest, and Logistic Regression have been used for credit card fraud detection in previous study Najadat et al. (2020). Due to a specific restrictions, Random Forest, SVM, KNN, Logistic and LightGBM did not achieve the complete accuracy without any fraud transactions, even though they demonstrated better accuracy. There was still a chance of fraudulent transactions, even if Deep Learning algorithms like RNN and CNN were more accurate than ML algorithms Alarfaj et al. (2022).

Some studies have investigated using the ensemble approaches, including stacking, to improve accuracy of credit card fraud detection in order to overcome these limits and expand upon earlier work. Findings improved when **?** used ensemble technique in conjunction with strict voting strategy. Stacking method, in contrast, was a thought to be more sophisticated, providing more characteristics for efficient credit card fraud detection, taking into account both current and future research limits.

Compared on prior work in the field, we want to use the ensemble technique with a stacking strategy Khandelwal (2021) to choose and combine most effective models from a Machine Learning & Deep Learning algorithms.

1.1 Motivation

The main motivation for choosing the topic is that myself was one of the victim for the topic I have taken. It was two years back there was some fraud transactions happened in my credit and when I report it to the bank they raised a ticked and said will be refunded back. I got my money back but where the money has gone and how come they hacked it?? These questions raised in my mind with several thoughts because they could not find the victim and they had refunded me their money as per the terms and conditions as they failed from preventing over the security measurements. Banking sectors are the becoming the main victims because of these activities and loosing the potential as per the security. So These all questions made me to do research on regarding the credit card fraud transaction while doing those research I came across every day there are some or how the fraud transactions are happening and causing the loss to the financial institutions as shown in Figure 2 as per the Nilson report Nilson Report (2020) and loosing the trust in respective of their firms among the people. Several methods have been implementing on regular basis as per the advancement of the technologies but some how still the accuracy is not 100 %.

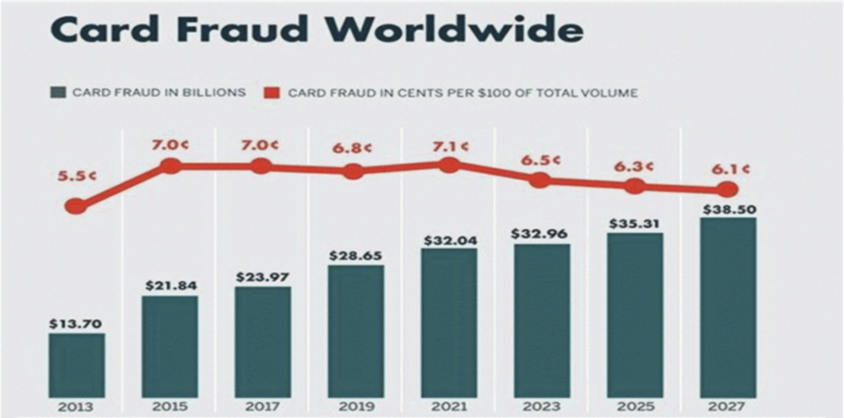


Figure 2: The loss due to credit card fraud transactions globally

1.2 Document Structure

All the required informations such as Related work has been explained in Section II, Methodology has been explained in Section III, Design Specification has been explained in Section IV, Implementation has been explained in Section V, Evaluation and Results has been explained in Section VI and the Conclusion and Future Work has been explained in Section VII.

2 Related Work

2.1 Research Question

How the ensemble method stacking approach can result in predicting credit card fraud transaction cases when the stacking method implemented between stacked machine learning models and stacked deep learning models ?

2.2 Research Background

In the article (Btoush et al.; 2021) titled ”A Survey on Credit Card Fraud Detection Techniques in Banking Industry for Cyber Security,” the authors thoroughly a review different methods used to detect credit card fraud. They cover the shift from older, rule-based systems to newer techniques involving machine learning (ML) and deep learning (DL) Priya and Saradha (2021). The paper examines a range of methods, including Hidden Markov Models (HMMs), basic machine learning algorithms such as a k-nearest neighbors (KNN), Naive Bayes (NB), Logistic Regression (LR), as well as more complex ones like Random Forest (RF) and Multilayer Perceptrons (MLPs). It looks at a how effective these techniques are in tackling the specific challenges of fraud detection, like the uneven distribution of fraud data and the need for quick processing.

Authors note that combination approaches, such as RF, are quite accurate, but they also note that they are a computationally intensive and where we analyse its difficult to grasp how models function (Al Smadi and Min; 2020). Although basic ML algorithms simplify fraud detection, they rely too much upon pre-labeled data where they can't identify novel kinds of fraud, which are big drawbacks. Unsupervised learning techniques, like GANs, are further addressed into article. These approaches are a effective in discovering novel fraud patterns but have lower accuracy & computational efficiency.

Research highlights a increasing popularity of hybrid models, that combine several ML approaches to enhance precision of predictions. According to Nancy et al. (2020) it is still a difficult to optimize such hybrid models for different datasets, despite fact that they are more effective that single-method ones. Despite advancements in ML & DL for fraud detection, research finds which existing systems still struggle to strike a balance amongst precision, computational effectiveness, including with the capacity for adaptation to novel fraud techniques.

For purpose of answering the following research question: This study examines several approaches in depth, with doing a analysis upon advantages of combinations of individual models along with hybrid models. It concludes that stacking strategy might be key to solving the problems that a single-method systems have. A stacked combination technique is being considered as an potential solution because to its high precision, adaptability to new fraud trends, & computing economy. This has potential to combine best features of different models whilst minimizing their drawbacks, which might be a great answer to problems with detecting credit card fraud. A system which is a both rapid and effective into real-world situations, as well as a very accurate, is the goal of this technique.

2.3 Data Imbalance – SMOTE Technique for Oversampling

This research studies detailed into a improved SMOTE method that tackles prevalent issue into machine learning: data that isn't uniformly distributed, particularly when dealing with massive datasets. Study contributes to broader conversation upon a methods for handling unevenly dispersed data, which might hinder performance of classification systems.A scenario which could compromise an accuracy of a categorization systems. In order to create more equitable data, original SMOTE approach is famous into this field for its ability to create false samples of underrepresented class. Research does point out 2 major problems using this technique, though: first, it tends to ignore manufactured samples. Second, parameters are chosen randomly, that could occasionally lead to a best findings Chen et al. (2021).

The authors propose a novel approach that replaces the uniform random number generation in SMOTE with a Normal distribution Pescim et al. (2010), aiming to center the synthetic samples around the minority class core, thereby reducing a marginalization and improving the classifier's sensitivity and a specificity. This is encapsulated in the proposed algorithm's use of the equation

$$p_{ij} = x_i + \text{randn}(0, 1) \times (x_{ij} - x_i), \tag{1}$$

where $\text{randn}(0, 1)$ is a random number from a Normal distribution with mean $\mu = 1$ and adjustable the standard deviation σ. The effectiveness of this method is demonstrated through improved classification outcomes on several medical datasets, with a algorithm's parameter selection maintaining the original data's distribution characteristics more an effectively than the original SMOTE.

The paper Wang et al. (2021) critically contributes to the academic conversation by a addressing the limitations of existing oversampling techniques and introducing an algorithm that is more sensitive to the distribution characteristics of the minority class. The mathematical underpinning provided by the Normal distribution ensures that the a newly generated samples are statistically representative of the minority class, with a higher probability of being closer to the class center, as an expressed in the distribution probabilities

$$P(\mu - \sigma \leq p \leq \mu + \sigma) \approx 0.6826, \tag{2}$$

$$P(\mu - 2\sigma \leq p \leq \mu + 2\sigma) \approx 0.9544, \tag{3}$$

$$P(\mu - 3\sigma \leq p \leq \mu + 3\sigma) \approx 0.9974. \tag{4}$$

They show that synthetic data points don't deviate too far by a minority cluster, which is feature of normal distribution. Bulk of data is located in a 3 standard deviations of mean.

Finally, this study greatly improves upon a original SMOTE approach by introducing updated version. It fixes its primary issues and provides a more robust answer to issue of unequal data into ML. Researchers have made important & worthwhile contribution to a discipline by basing their technique upon sound statistical concepts. This study implies there's has to be additional research into parameter adjustment in order to include more data. It additionally also expands our understanding of SMOTE additionally reveals new avenues for a better data preparation in ML.

2.4 Ensemble Methods

In order to have the better comprehensive evaluation of current fraud detection systems, Tomar et al. (2021) we need to examine a fast expanding field of digital transactions how they are processing on a daily basis. It keeps us to invlove into how data-focused approaches are going to be replacing older, and also the rule-based systems that are struggling to keep up with exponential growth in to data volumes along with the complexity. In 2019 the research study, as per the Kim et al. (2019) Class imbalance and along with the ideas drift are a primary known issues highlighted in to the article, which provide a lot range of the obstacles to traditional methods of detecting fraud in regular daily process.

Compared to using individual models, the ensemble technique as shown in Figure 3 yields better results, which is major conclusion of this work. Data balancing preliminary processing, a data partitioning into training & testing sets, & ensemble learning to merge model predictions are all part of method. Through this procedure, efficacy of ensemble approaches into handling the imbalanced data is shown. With an emphasis on enhanced predictive potential of these hybrid models, article highlights a transition from individual classifiers to ensemble approaches in its literature review Tomar et al. (2021).

Current solutions fail to tackle intricate and ever-evolving character of credit card theft, as according to paper's conclusion. To improve efficacy of fraud detection systems, further study in the ensemble approaches is required, especially those that use a stacking strategy. Importantly, this study's mainly working on the goal is to find solution to problems caused by existing methods by stacking ensembles of machine learning & deep learning models. Area is predicted to benefit from this technique since it provides a fraud detection system which is strong, flexible, & very accurate; it can also adapt to new fraud strategies & ensure security of a monetary transactions into digital age.

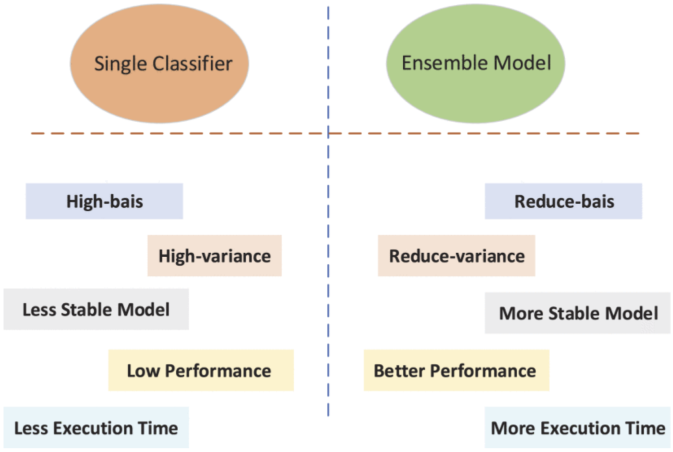


Figure 3: Ensemble method [Source: Google]

2.4.1 Stacking Approach

An important step forward into battle against credit card fraud utilizing machine learning is work of Nayyer et al. (2023). It presents ensemble learning, technique that combines predictions of many basic classifier using stacking model in an effort to improve precision of predictions. While the research does note stacking model is data & computationally intensive, additionally highlights its advantages, such as enhancing accuracy of predictions while decreasing likelihood as model would match training data overly closely.

According to Khan et al. (Mar. 2022), procedure outlined into article entails training fundamental learning models (B[i]), combining their predictions (R), then feeding these combined projections into meta-model. This model subsequently produces final prediction (P). Complex & non-linear data interactions are common in fraud detection, technique shows how stacked models may handle them well.

Authors go over problems with previous approaches, emphasizing fact that single classifiers have hard time dealing with things like imbalanced classes with evolving fraud detection trends. They propose ensemble approach combining several ML models—including Decision Trees, Logistic Regression, & Naive Bayes—to tackle these issues more effectively. Findings from this study provide strategy to enhance precision of credit card fraud detection by using meta-model to improve predictions. Given dynamic nature of financial fraud, comprehensive & adaptable solution is required; here is where ensemble technique using stacking approach shines Khan et al. (Mar. 2022).

Finally, article argues that suggested stacking paradigm is far better than alternatives. More study upon ensemble approaches for credit card fraud detection is needed. As shown in Figure 4 describes stacking technique as comprehensive & viable solution requiring more research & implementation in realm of financial security *Stacking in Machine Learning* (2023).

2.5 Machine Learning and Deep Learning Models

In order to identify instances of credit card fraud, Varmedja et al. (2019) used variety of ML techniques, including Naive Bayes, Logistic Regression, & RF. In order for eliminating

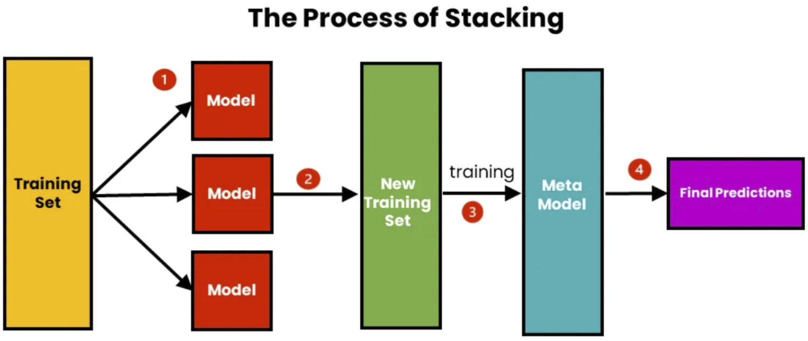


Figure 4: The Stacking process

data imbalances during model train & test, they utilised SMOTE approach. According to their results, RF model had highest accuracy rate at 99.96%, then Naive Bayes at 99.23%, & Logistic Regression at 97.46%. The outcomes demonstrate Random Forest excels at detecting credit card fraud.

Credit card fraud detection features were selected using Genetic Algorithm, according to paper Saheed et al. (2020). After sorting data characteristics into two categories, they trained and tested using RF, SVM, & Naive Bayes. Having success rate of 96.40%, RF algorithm was shown to be most accurate, surpassing both Naive Bayes (94.3%) & SVM (96.3%).

In separate research, Ge et al. (2020) investigated possibility of identifying credit card fraud using algorithms such as Logistic Regression, & LightGBM. LightGBM method, that used improvements using gradient-based One-side Sampling & exclusive feature creation, was shown to provide most benefits into this study. Achieving 98.2% precision, LightGBM topped list, followed by Xgboost (97.1%), SVM (95.2%), & Logistic Regression (92.6%). Report proposed that by using new approaches, more gains into accuracy may be accomplished.

In response to previous studies shortcomings, authors of present study suggest machine learning and deep learning models into Ensemble using Stacking methodology. By capitalizing on capabilities of each machine learning model and deep learning, this method is anticipated to improve the precision of detecting credit card fraud.

Credit card theft was detected using CNN, according to research by Gambo et al. (2022). CNNs demonstrate encouraging results in several domains, especially when it comes to feature selection & avoiding overfitting. Completely linked layers, pooling, downsampling, are all components of CNN. Researchers improve CNN model's precision, recall, precision, & F1-score by combining it with approaches such as SMOTE, LSTM, & ADASYN. CNN+ADASYN model outperformed others, with an accuracy rate of 99.82%. In light of these findings & caveats highlighted by their literature evaluation, researchers opted to use the CNN method into their investigation. In addition to other deep learning methods, they want to include this model in ensemble technique using stacking strategy.

Throughout COVID era, there was a dramatic increase in both online transactions as well as fraud instances, hence another article by Dutta and Bandyopadhyay (2020) addressed topic of identifying fraudulent transactions. Minimizing fraud & safeguarding clients' financial interests were their primary objectives. In order to detect credit card fraud, they used hyperparameter-tuned stacked-RNN model. The 99.87% accuracy

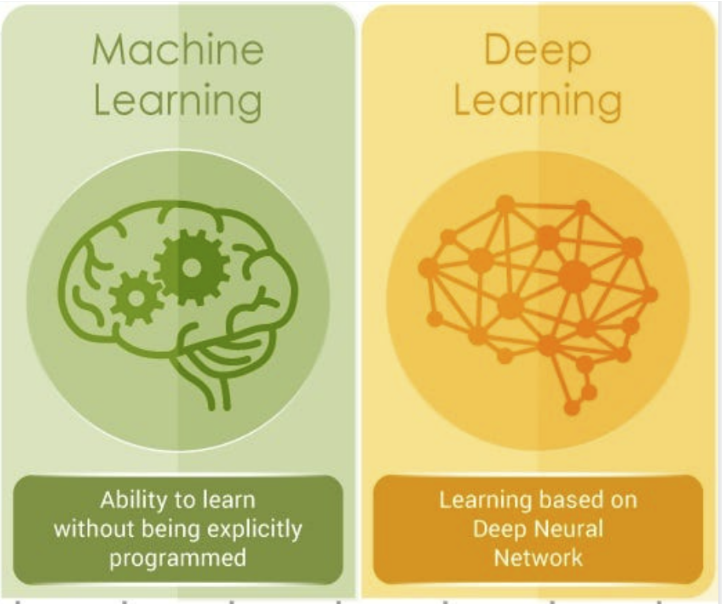

Machine Learning
Ability to learn without being explicitly programmed
Deep Learning
Learning based on Deep Neural Network

Figure 5: Machine learning and Deep Learning [Source: Google]

percentage shown by this model is quite remarkable. These results, together with gaps found in literature, led the researchers to decide to use stacked RNN approach in their investigation. Their goal is to enhance detection of credit card fraud even further by combining this technology with CNN deep learning algorithm into ensemble way using stacking strategy.

3 Methodology

3.1 Data Collection and Preprocessing

First stage of study was to gathering an extensive database of credit card transactions and split it in sets for testing and training. Non-essential an index column was removed as going to be a of data evaluation executed with a Python Pandas package. Every transaction's the authenticity was safeguarded by this a exhaustive pre-processing, which sought to the preserve data quality by removing duplicate entries and a missing values.

Algorithms that are the sensitive to feature scales must undergo data preparation that can going to involved standardizing numerical characteristics to guarantee consistency as according to Alam et al. (2020). Following is a algebraic illustration of this below mentioned standardization:

$$X' = \frac{X - \mu}{\sigma} \tag{5}$$

where X is an one that is taken from a original data, μ is the mean taken, and σ is known as a standard deviation.

It's critical step since the a accuracy of a prediction models can only be as good as data used to create them. In order to a develop more the reliable models & ensure trustworthiness of future predictive studies, it's necessary to filter dataset by removing a redundant information & irregularities.

3.2 Feature Analysis

Research conducted comprehensive feature analysis following data cleaning. To better get a understand connections among numerical characteristics, we used correlation matrix that was a computed utilizing Pearson's correlation coefficient:

$$r_{xy} = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum(X - \overline{X})^2 \sum(Y - \overline{Y})^2}} \tag{6}$$

This coefficient indicates the strength & direction of the linear relationship among a two variables; it ranges from -1 to 1 Sailusha et al. (2020). In order avoid overfitting, we only included a characteristics which had high correlation by 'is fraud' result into a model and the eliminated elements which were too comparable.

We additionally developed a graphics for category data to help spot trends that might indicate fraud. Key characteristics for a predictions were identified by doing careful examination of data utilizing count & and a scatter plots. Feature selection procedure was guided by a information gathered from this study, making sure most a important variables were brought through to model train.

3.3 Model Training

When training the models, number of different machine learning methods were used. Logistic Regression, SVM having polynomial kernel, KNN, & Random Forest were all selected because to their shown ability to a recognize and classify patterns. To guarantee thorough a learning by characteristics of dataset, training was carried out upon varied data samples.

To further aid in detection of non-linear, complicated patterns into a data, deep learning algorithms like RNNs & CNNs have been used. Models were built utilizing TensorFlow & Keras, and they're a great at identifying complex patterns seen in fraud detection data because of their ability to recognize time-based & multi-level data structures.

During ensemble stage, stacking strategy was the used to train meta-model using predictions by basic models. Following is mathematical training of meta-model M, where $B_1, B_2, \ldots, B_n$ are a predictions by basis models:

$$M.fit([B_1, B_2, \ldots, B_n], Y) \tag{7}$$

Wherein Y signifies true target results. Stack method is made for improving predictive precision with collective strengths of numerous models.

3.4 SMOTE for Class Imbalance

An important component of research's approach consisted SMOTE algorithm as shown in Figure 6, which dealt with a problem of unbalanced classes by generating synthetic cases within minority class:

$$s = x + \lambda \times (x' - x) \tag{8}$$

Where mentioned x is sample by minority class, x' is close neighbor by same class, λ is random number amongst 0-1 Chen et al. (2021). By increasing representation of

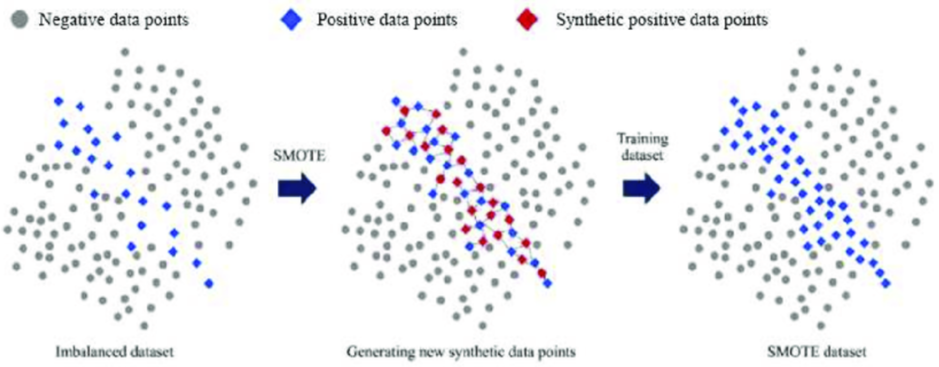


Figure 6: SMOTE Technique Process

underrepresented groups into training data, our technique contributed to more equitable dataset.

Without bias of uneven distribution of classes, models were able to pick up upon finer points of fraud into this balanced sample. Thus, models' accuracy & generalizability to new data were enhanced by using SMOTE.

3.5 Evaluation Metrics

Several measures which shed light on models' predictive power were used to evaluate their efficacy. A comprehensive assessment of models' efficiency was achieved by evaluating their recall, precision, & F1score. We further took precision & recall into account since precision alone could be deceiving into datasets where the classes are not evenly distributed. Here are the formulas for calculating for each of these measures:

The formulas for these metrics are:

$$\text{Precision} = \frac{TP}{TP + FP},$$

$$\text{Recall} = \frac{TP}{TP + FN},$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}.$$

Due to its unique ability to combine accuracy & recall in single indication, the F1-score received special attention. This score reflects balance among accurately detecting number of fraud instances & minimizing false positives.

3.6 Stacking Ensemble Method

An essential part of research strategy stacking ensemble technique, which included training meta-model by merging predictions of starting models. Improving precision of predictions, each higher-level model drew upon collective wisdom of its forerunners to function as sophisticated classifier Khan et al. (Mar. 2022).

```
M.fit([B_1, B_2, ..., B_n], Y)
```

To provide strong barrier against credit card fraud, our meta-learning strategy inorder to compensate for shortcomings of individual models while capitalizing on their combined capabilities.

3.7 Combined Stacked Model

A combination stacking model utilizing machine & deep learning outputs was presented as advanced step into research. Last meta-model was Logistic Regression, which combined predictions of underlying models in collection of features. A powerful fraud detector was created by combining distinct viewpoints of two models.

To determine model efficacy formula can be given as:

```
Stacked Accuracy = 1/n * sum(I(y_i = y_hat_i))
```

where I is used as an indicator function.

It was anticipated as cumulative stacked model's precision would surpass ability of any individual contributing model, demonstrating effectiveness of ensemble stacking method into field of fraud detection.

4 Design Specification

Credit card fraud detection system that was built into this study follows the design requirements. It describes system's architecture in depth, covering parts & how they work together, as well as reasoning behind certain design choices.

4.1 System Architecture

Data collection, preparation, analysis of characteristics, training of models, evaluation of those models, & prediction are all critical steps into system's architecture as shown in Figure 7. To improve precision of fraud detection forecasts, it uses ensemble approach that combines deep learning & machine learning models.

At outset, transaction data is input into system via data manual uploading into repository and then to module by python packages. In next step of data preparation, known as preprocessing, data is filtered by correcting or deleting any incorrect, insufficient, or unnecessary portions of data. Data is then normalized by converting numbers from several scales to a single, conceptual one. At this point, we additionally scale the features. In order to choose what information characteristics to utilize, feature analysis module utilize methods to analyze data & uncover patterns & correlations.

The next step is to train variety of models based on machine learning, such as Random Forest, SVM, KNN, & Logistic Regression. Predictive power of system rests on these models. We train deep learning models like CNNs & RNNs to find more intricate patterns into data. In ensemble stacking architecture, such models are stacked one on top of other.

Metrics like F1-score, recall, accuracy, & precision assess performance of every model throughout assessment step. Next, ensemble framework's meta-model takes forecasts from those models and feeds them into prediction stage. Ultimate fraud prediction is made by merging insights by several base models using this meta-model, which is frequently Logistic Regression classifier.

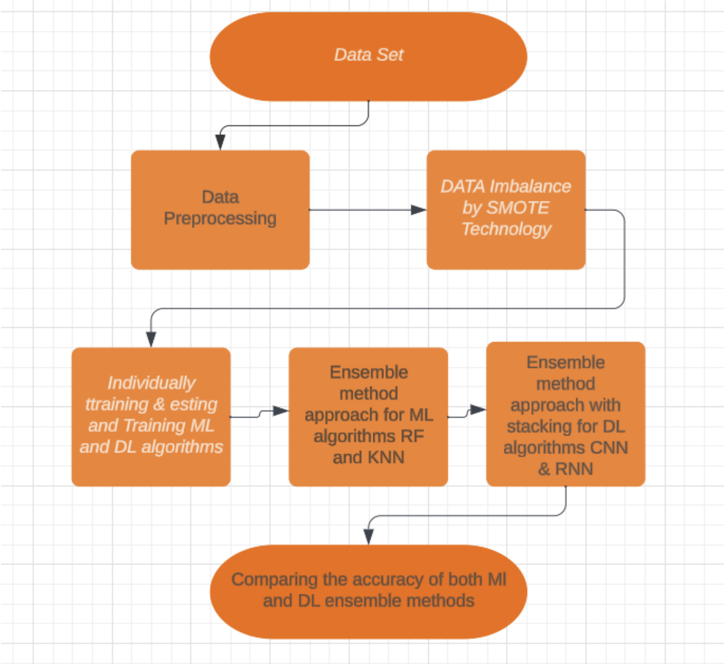

Data Set
Data Preprocessing
DATA Imbalance by SMOTE Technology
Individually ttraining & esting and Training ML and DL algorithms
Ensemble method approach for ML algorithms RF and KNN
Ensemble method approach with stacking for DL algorithms CNN & RNN
Comparing the accuracy of both Ml and DL ensemble methods

Figure 7: Design Architecture the flow how it has been build

4.2 Framework and Tools

Due to its extensive library supporting the active user community, Python has been selected as primary language for development. Pandas & NumPy are two libraries that might be helpful when handling data in-order for doing mathematical computations. To display data graphically, we utilize Matplotlib & Seaborn. Building and as well as train deep learning CNN & RNN is done using TensorFlow & Keras, whereas tools for classic artificial intelligence models along with metrics are provided via scikit-learn package.

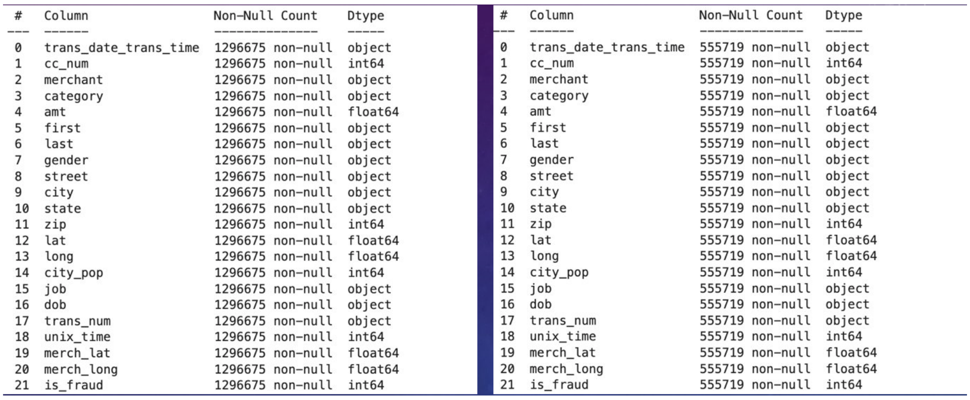
A common problem in datasets is that one class is over-represented compared to others. To address this, imbalanced-learn library's SMOTE approach is utilized. As a result, models are less likely to favor more prevalent category.

5 Implementation

A thorough synthesis of careful data analysis & complex mathematical design went into creation of credit card fraud detection system. In order to improve precision of fraud detection, ensemble architecture is used during execution phase to draw on many prediction models. Last step of project is described into this account along with the outcomes, tools used, & programming languages.

5.1 Data Loading and Preprocessing

Loading fraudTrain.csv & fraudTest.csv datasets in Pandas dataframes was first step into implementation as shown in Figure 8 . This is essential step for handling data & analysis. The mathematical reliability of these datasets was confirmed upon first investigation. Due to its lack of relevance to investigation, first column of datasets comprising timestamps, amounts, client as well as merchant information, & fraud classifications was classified.



```
 #   Column                 Non-Null Count    Dtype
---  ------                 --------------    -----
 0   trans_date_trans_time  1296675 non-null  object
 1   cc_num                 1296675 non-null  int64
 2   merchant               1296675 non-null  object
 3   category               1296675 non-null  object
 4   amt                    1296675 non-null  float64
 5   first                  1296675 non-null  object
 6   last                   1296675 non-null  object
 7   gender                 1296675 non-null  object
 8   street                 1296675 non-null  object
 9   city                   1296675 non-null  object
 10  state                  1296675 non-null  object
 11  zip                    1296675 non-null  int64
 12  lat                    1296675 non-null  float64
 13  long                   1296675 non-null  float64
 14  city_pop               1296675 non-null  int64
 15  job                    1296675 non-null  object
 16  dob                    1296675 non-null  object
 17  trans_num              1296675 non-null  object
 18  unix_time              1296675 non-null  int64
 19  merch_lat              1296675 non-null  float64
 20  merch_long             1296675 non-null  float64
 21  is_fraud               1296675 non-null  int64
```

```
 #   Column                 Non-Null Count   Dtype
---  ------                 --------------   -----
 0   trans_date_trans_time  555719 non-null  object
 1   cc_num                 555719 non-null  int64
 2   merchant               555719 non-null  object
 3   category               555719 non-null  object
 4   amt                    555719 non-null  float64
 5   first                  555719 non-null  object
 6   last                   555719 non-null  object
 7   gender                 555719 non-null  object
 8   street                 555719 non-null  object
 9   city                   555719 non-null  object
 10  state                  555719 non-null  object
 11  zip                    555719 non-null  int64
 12  lat                    555719 non-null  float64
 13  long                   555719 non-null  float64
 14  city_pop               555719 non-null  int64
 15  job                    555719 non-null  object
 16  dob                    555719 non-null  object
 17  trans_num              555719 non-null  object
 18  unix_time              555719 non-null  int64
 19  merch_lat              555719 non-null  float64
 20  merch_long             555719 non-null  float64
 21  is_fraud               555719 non-null  int64
```

Figure 8: fraud_Train Data and fraud_Test Data

Resolving missing values & eliminating duplicates were part of comprehensive cleaning procedure that datasets underwent to guarantee their reliability. Making sure data had been arranged and polished at this first step was crucial for accuracy of following model training.

5.2 Data Visualization and Analysis

Several methods of data visualization have been employed to reveal previously unseen relationships and the related patterns. In order to understand data's framework, certain methods are essential, including:

One way to understand disparity between classes is to look at distribution of "is fraud" variable as shown in Figure 9.

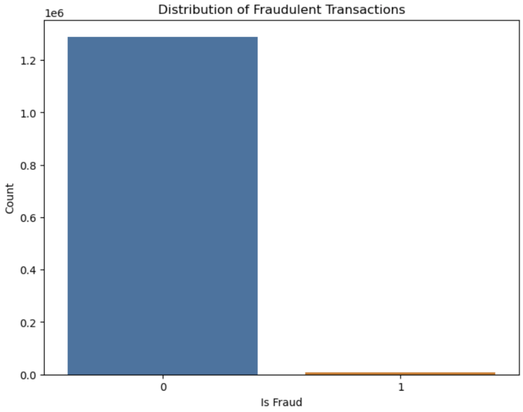


Figure 9: The visualisation of the fraud transactions

Building correlation matrix to identify correlations between variables using Seaborn as shown in Figure 10.

Sorting and also plotting the frequency of fraud to see how it varies across different kinds of transactions. Creating map of every transaction to find patterns of fraud into certain areas as shown in Figure 11.

Investigating potential trends in fraudulent activity over time by graphing the frequency of transactions as shown in Figure 12.

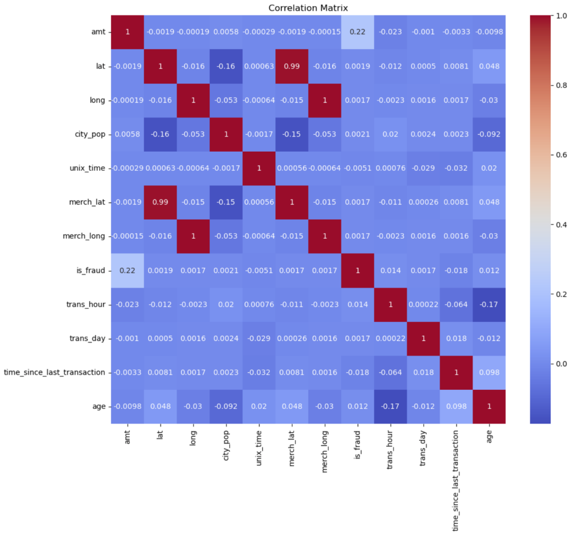


Figure 10: The visualisation of the correlation matrix between variables

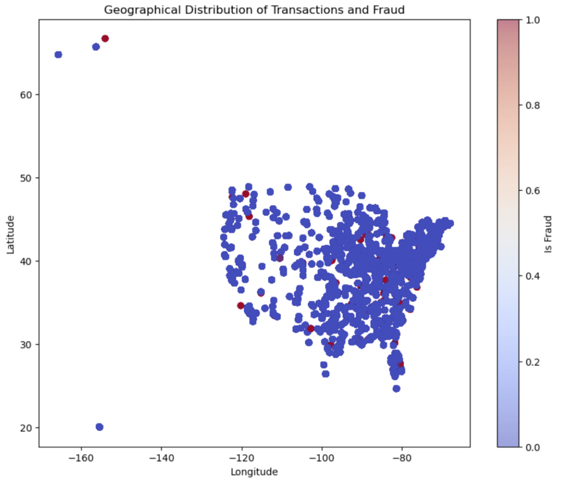


Figure 11: The visualisation of the fraud over areas

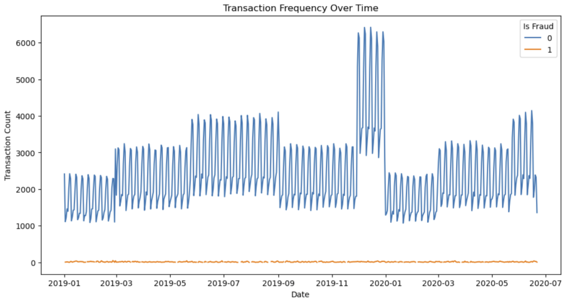


Figure 12: The visualisation fraud transactions over period of time

Such visual investigations helped direct selection of features and in-order for the refinement of models for certain data variables.

5.3 Model Development and Training

Creation and training of a set of models formed backbone of implementation. Among these models, Logistic Regression, SVM, KNN, & Random Forest were selected for their unique categorization methodologies. We used F1-score, recall, accuracy, & precision as our standard measures of effectiveness. In order to capture complex data patterns, sophisticated deep learning models were created utilizing TensorFlow & Keras. These models include CNNs & RNNs. Intricate nature of transaction sequences makes it difficult for these algorithms to pick up on small indications of fraud. To address issue of class imbalance, SMOTE approach was used to synthesize fresh samples from minority class in order to train more balanced model. As shown in Figure 13 balanced the data by applying SMOTE approach.

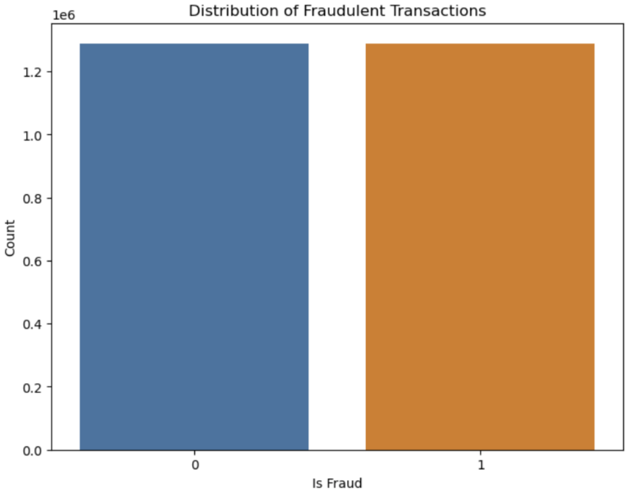


Figure 13: Data balancing with SMOTE Approach

5.3.1 Stacking method

The first stacking layer combined the outputs from the Random Forest and KNN models, with Logistic Regression acting as the meta-model to assimilate these predictions. A second layer integrated forecasts from the CNN and RNN models, again employing Logistic Regression as the meta-model. The combined stacked model merged the features from the first and second stacking layers into a comprehensive set, which was then used to train a final Logistic Regression model.

5.4 Outputs and Evaluation Metrics

A strong ensemble model that could accurately forecast fraudulent transactions was end result of implementation. A number of criteria were used to assess not just ultimate ensemble model but also every model that has been utilised:

Precision, indicating how well forecasts hold up in general. Accuracy, measuring how well model detects actual instances of fraud. As you may remember, this metric is based on number of observed cases of fraud. The F1-score, which combines recall & accuracy in one metric. Impressive results were achieved by ensemble models, particularly combined stacked model, which proved efficacy of ensemble approach Hisham et al. (2022).

6 Evaluation

The evaluation mainly presents on how far the implementation I did is actually predicting the credit card fraud transactions. Have gone through so many modification during the implementations for finding out how these machine learning, deep learning and ensemble methods are going to perform for detecting credit card fraud transactions.

6.1 Machine Learning Models

6.1.1 Logestic Regression

In assessing the Logistic Regression model, the results highlight its capability using several important measures. The model has an accuracy rate of approximately 81.240%, which translates to it accurately identifying fraudulent transactions roughly 82 times out of 100. The precision value of 0.852 suggests that the model is quite dependable when it determines a transaction is fraudulent, and its recall rate of 0.755 signifies that it is effectively identifying a significant portion of actual fraud cases. With an F1 score of 0.801, there's a healthy equilibrium between precision and recall, indicating the model's robust performance.

An outstanding precision of 0.983 was determined for SVM model during evaluation, indicating that it is very effective in detecting fraud as it forecasts it at accuracy of 86.126%. Contrarily, its recall of 0.735 indicates that it misses a few fraudulent transactions. F1-score of 0.841 strikes a balance amongst recall and precision, suggesting model might be fine-tuned to detect more fraud cases, despite its excellent accuracy.

Our examination of nonparametric KNN model showed encouraging outcomes on validation set. Program correctly identified fraudulent transactions in approximately 88% of cases, with an accuracy rate of 87.581%. In addition, model was quite good at identifying most real fraudulent transactions. Model's efficacy was highlighted by F1 score of 0.879, a crucial statistic combining recall with accuracy.

With score of 88.702%, RF model demonstrates an impressive equilibrium in its forecast accuracy. With a precision of 0.970, model proves it correctly predicts fraud most of time. A good overall performance is indicated by F1 score of 0.876, which is the average of recall & accuracy. Recall score of 0.799 suggests as model has room to grow, even if accuracy is good; about 21% of real fraud cases were missed.

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1 Score**	**ROC**
Logistic	82.240	0.852	0.755	0.801	0.91
SVM	86.126	0.983	0.735	0.841	0.89
K-Nearest(KNN)	87.581	0.854	0.907	0.879	0.95
RF	88.702	0.970	0.799	0.876	0.98

Table 1: Performance Evaluation Metrics of Machine Learning Models

6.2 Deep Learning Models

When it comes to data classification, CNN model appears to be quite effective. Its 86.933% accuracy shows it is able to detect fraud in around 87 out of 100 instances. Its precission in predicting fraud is 0.951 percent, but its recall is 0.778 percent, therefore it captures a decent percentage of real fraud incidents. A well-balanced F1 score of 0.856 indicates as model efficiently detects large number of actual instances of fraud while simultaneously minimizing the occurrence of false alarms. Model trained well as its loss dropped throughout training, & its high ROC AUC score of 0.95 indicates that it does a great job of differentiating between legitimate and fraudulent transactions. According to confusion matrix, model makes very few mistakes and has high percentage of accurate predictions. To sum up, CNN has proved successful in identifying complicated patterns, that is critical in real-world scenarios, thanks to its comprehensive layers & utilization of dropout methods to prevent overfitting.

Renowned for its ability to handle sequences and remember previous data, RNN model has shown remarkable performance. With a documented accuracy of 92.212%, it proved to be very accurate in identifying right outcomes. An important attribute in domains wherein mistakes may be more costly, model generates minimal false alarms having precission rate of 0.912.Also, model has a high recall rate of 0.934, indicating that it will recognize most occurrences which need to be warned. This is especially important in fields like fraud detection where missing a real case may have serious consequences. An F1 score of 0.923 indicates that model has successfully tuned 2 crucial metrics, recall & accuracy, suggesting a balanced trade-off.

Model	Accuracy (%)	Precision	Recall	F1 Score	ROC
CNN	86.933	0.951	0.778	0.856	0.95
RNN	92.212	0.912	0.934	0.923	0.98

Table 2: Performance Evaluation Metrics of CNN and RNN Models

6.3 Case Study 1 : Stacked Machine Learning Models for Fraud Detection

In order to improve precision of banking transaction fraud detection, this review examines efficacy of layered machine learning strategy. This method trained two base models—Random Forest & KNN—and then used Logistic Regression meta-model to combine their findings before reaching a final decision on transactions. With accuracy of 88.702%, Random Forest model demonstrated excellent performance. With a recall of 0.799 and an precission rate of 0.970, model was clearly adept at detecting significant percentage of actual fraud incidents.

KNN model, on other hand, came out on top with reliability of 87.581%, a high precision rate of 0.854, with even more impressive recall of 0.907, all of which indicate its exceptional ability to identify fraudulent activity. When looking at combined model that included both Random Forest & KNN predictions, there was no change in efficiency to 88.702%. A good balance was shown by model's F1 score of 0.876, which showed that the precission remained high at 0.970 and the recall at 0.909. Ensemble meta-model's AUC upon ROC curve was 0.94, & model's dependability was validated by Precision-Recall Curve; these illustrative tools demonstrated model's outstanding predictive capacity.

Combined model's confusion matrix, which displayed substantial number of true positives & negatives, provided a clear picture of its predictive power. Model's efficacy was confirmed by small number of false positive & negative occurrences compared to number of true predictions. In conclusion, detecting fraudulent transactions was made easy by stacking technique that used machine learning.

Model	Accuracy (%)	Precision	Recall	F1 Score	ROC
Random Forest	88.702	0.970	0.799	0.876	0.98
KNN	87.581	0.854	0.907	0.879	0.95
Stacked Approach (RF & KNN with Logistic Regression)	88.702	0.970	0.799	0.876	0.94

Table 3: Performance Evaluation Metrics of Stacked Machine Learning Models

6.4 Case Study 2 : Stacking Deep Learning Models for Enhanced Predictive Analysis

To enhance predictive analysis, we investigated efficacy of stacked deep learning strategy that included CNNs & RNNs into this case study. In order to improve final predictions, CNN & RNN base models' outputs were combined utilizing logistic regression as metamodel.

Predictions were of good quality from both CNN & RNN. CNN is well-known for its spatial data understanding capabilities, while RNN excels at processing sequential data. After that, logistic regression meta-model took advantage of every neural network's unique strengths by using them as input.

Assessment metrics highlighted ensemble's impressive efficiency: a 92.212% accuracy rate showed that model was generally good at delivering accurate predictions. With precision of 0.912 and recall of 0.934, it was clear that it was good at detecting the majority of real cases & had no trouble producing false positives. For tasks like fraud detection, a well-balanced blend of recall & accuracy is required, & ROC score of 0.95 demonstrated this.

Model's capacity to distinguish amongst classes was shown by remarkable AUC values provided by ROC & Precision-Recall curves, which are visual evaluation tools. Confusion matrices confirmed low error rates and high rates of correct classifications, and also offered comprehensive analysis of model's predictions.

When it came to complicated prediction tasks, this layered deep learning system showed a lot of potential, as every statistic revealed its capabilities. Improving these models further or adding other kinds of neural networks to make more accurate predictions might be subject of future studies.

Model	Accuracy (%)	Precision	Recall	F1 Score	ROC
CNN	86.933	0.951	0.778	0.856	0.95
RNN	92.212	0.912	0.934	0.923	0.98
Stacked Approach (CNN & RNN with Logistic Regression)	92.212	0.912	0.934	0.923	0.95

Table 4: Performance Evaluation Metrics of Stacked Deep Learning Models

6.5 Case Study 3: Evaluation of Hybrid Stacked Model Combining Machine Learning and Deep Learning Approaches

The purpose of this particular case study was to assess the efficacy of stacked hybrid model for fraud detection, which incorporated ML and DL to improve the precision of predictions. To take use of best features in both cases, this model combines machine learning and deep learning ensembles by first two studies.

In this hybrid model, a logistic regression meta-model combines predictions of deep learning models with those of machine learning algorithms. The goal of this approach is to combine the superior feature extraction capabilities of deep learning by precise classification capabilities of machine learning. A accuracy of 92.472 was attained using hybrid model.The model successfully reduced false positives, an important part of fraud detection to avoid needless alerts, by precision of 0.912. Considering difficulty of combining several kinds of models, this accuracy is outstanding as we can see the increase in the hybrib stacked model compared to individual stacked model.

With recall rate of 0.934, the model clearly identified majority of fraudulent instances. This agreement with recall of deep learning model implies as hybrid model kept its sensitivity to fraud intact even after incorporating diverse methods.In fraud detection, wherein both false positives & negatives may have serious consequences, F1 score of 0.923 demonstrated an adequate equilibrium amongst recall and accuracy.

Model's accuracy and resilience were highlighted by ROC curve & Precision-Recall curve, which both had AUC of 0.97. A large number of accurate predictions with few incorrect classifications were shown by confusion matrix, demonstrating efficacy of model. Combining different analytical approaches yields better results than using either machine learning or deep learning alone; this is shown by hybrid model, which achieves marginally better accuracy. Overall performance indicates a good combination of deep learning & machine learning.

Model	**Accuracy (%)**	**Precision**	**Recall**	**F1 Score**	**AUC**
Machine Learning Ensemble	88.702	0.970	0.799	0.876	0.94
Deep Learning Ensemble	92.212	0.912	0.934	0.923	0.95
Hybrid Stacked Model (ML & DL) with Logistic Regression	92.472	0.912	0.934	0.923	0.97

Table 5: Performance Evaluation Metrics of Hybrid Stacked Model

6.6 Discussion

This study of combining machine & deep learning models for fraud prediction is shown by this comparative study across 3 case studies. Although first case study demonstrated some promising results, it also highlighted how combining several machine learning models can at times make things more complicated and impact overall accuracy. Higher precision was achieved in the second case study by using deep learning models, such as RNNs & CNNs, which proved their capacity to comprehend intricate patterns in data. 3rd case study demonstrated that there is accuracy increased, there was a no loss of precision when using a hybrid model that included deep learning and machine learning. Logistic regression meta-model, combining the stacked machine learning and the stacked deep learning has been choosen to be best model for credit card fraud detection.

It is crucial to choose & configure meta-model with care, as these insights show, even if combining models might enhance outcomes. Consistent with the prevailing consensus in field, this suggests that using many models simultaneously may provide better results than depending on single one. If we want to take predictive modeling to next level, we should probably use methods like crossvalidation to make the models even better, and then try other combinations to see what works best.

7 Conclusion and Future Work

This study mainly aimed working on the combining ML & DL models for banking fraud detection was primary goal of this work. For purpose of predictive analysis, 3 case studies were carried out, every one of which tested various combinations of these models.An ensemble of a ML models was center of attention in a first instance. It was encouraging, but it also pointed out the combining models might reduce ROC. In second instance, we see how CNNs & RNNs perform a well when faced with complicated patterns; by using deep learning approach. In third scenario, accuracy was increased but a precision was constant. This result indicates that a current meta-model, logistic regression, is useful for predicting the credit card fraud transactions. The study's authors stressed importance of selecting and setting up correct meta-model, noting that with predictive analytics, sum of many models' strengths is often greater than their individual ones.

In order to improve the model's parameters, future research might look into more a sophisticated optimization techniques, such as cross-validation. To improve prediction skills, try out different combos of deep learning & machine learning models, maybe with more advanced meta-models or fresher neural network designs. These sophisticated fraud detection systems have a commercialization potential because of their adaptability to the many types of a financial environments. Improving safety and effectiveness of monetary transactions, this study paves way for more advanced prediction models.

References

Al Smadi, B. and Min, M. (2020). A critical review of credit card fraud detection techniques, *2020 11th IEEE Annual Ubiquitous Computing, Electronics Mobile Communication Conference (UEMCON)*, pp. 0732–0736.

Alam, T. M., Shaukat, K., Hameed, I. A., Luo, S., Sarwar, M. U., Shabbir, S., Li, J. and Khushi, M. (2020). An investigation of credit card default prediction in the imbalanced datasets, *IEEE Access* **8**: 201173–201198.

Alarfaj, F. K., Malik, I., Khan, H. U., Almusallam, N., Ramzan, M. and Ahmed, M. (2022). Credit card fraud detection using state-of-the-art machine learning and deep learning algorithms, *IEEE Access* **10**: 39700–39715.

Btoush, E., Zhou, X., Gururaian, R., Chan, K. and 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)*, pp. 1–7.

Chen, B., Xia, S., Chen, Z., Wang, B. and Wang, G. (2021). Rsmote: A self-adaptive robust smote for imbalanced problems with label noise, *Information Sciences* **553**: 397–428.

Dutta, S. and Bandyopadhyay, S. K. (2020). Detection of fraud transactions using recurrent neural network during covid-19, *Journal of Advanced Research in Medical Science & Technology* **7**: 16–21.

Gambo, M. L., Zainal, A. and Kassim, M. N. (2022). A convolutional neural network model for credit card fraud detection, *2022 International Conference on Data Science and Its Applications (ICoDSA)* pp. 198–202.

Ge, D., Gu, J., Chang, S. and Cai, J. (2020). Credit card fraud detection using lightgbm model, *2020 International Conference on E-Commerce and Internet Technology (ECIT)* pp. 232–236.

Hisham, S., Makhtar, M. and Aziz, A. A. (2022). Combining multiple classifiers using ensemble method for anomaly detection in blockchain networks: A comprehensive review, *Int. J. Adv. Comput. Sci. Appl.* **13**(8): 404–422.

Khan, I. U., Javeid, N., Taylor, C. J., Gamage, K. A. A. and Ma, X. (Mar. 2022). A stacked machine and deep learning-based approach for analysing electricity theft in smart grids, *IEEE Trans. Smart Grid* **13**(2): 1633–1644.

Khandelwal, Y. (2021). Ensemble stacking for machine learning and deep learning, Data Science Blogathon.
URL: *https://www.analyticsvidhya.com/blog/2021/08/ensemble-stacking-for-machine-learning-and-deep-learning/*

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., kwan Nam, S., Song, Y., a Yoon, J. and il Kim, J. (2019). Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning, *Expert Systems with Applications* **128**: 214–224.
URL: *https://doi.org/10.1016/j.eswa.2019.03.042*

Najadat, H., Altiti, O., Aqouleh, A. A. and Younes, M. (2020). Credit card fraud detection based on machine and deep learning, *2020 11th International Conference on Information and Communication Systems (ICICS)* pp. 204–208.

Nancy, A. M., Kumar, G. S., Veena, S., Vinoth, N. A. S. and Bandyopadhyay, M. (2020). Fraud detection in credit card transaction using hybrid model, *AIP Conference Proceedings* **2277**: 130010.

Nayyer, N., Javaid, N., Akbar, M., Aldegheishem, A., Alrajeh, N. and Jamil, M. (2023). A new framework for fraud detection in bitcoin transactions through ensemble stacking model in smart cities, *IEEE Access* **11**: 90916–90938.

Nilson Report (2020). Newsletter 1187, `https://nilsonreport.com/newsletters/1187/`.

Pescim, R. R., Demétrio, C. G., Cordeiro, G. M., Ortega, E. M. and Urbano, M. R. (2010). The beta generalized half-normal distribution, *Computational Statistics Data Analysis* **54**(4): 945–957.
URL: *https://www.sciencedirect.com/science/article/pii/S0167947309003776*

Priya, G. J. and Saradha, S. (2021). Fraud detection and prevention using machine learning algorithms: A review, *2021 7th International Conference on Electrical Energy Systems (ICEES)*, pp. 564–568.

Rolfe, A. (2022). 2022 ibm global financial fraud impact report, *Payments Card Mobile* .

Saheed, Y. K., Hambali, M. A., Arowolo, M. O. and Olasupo, Y. A. (2020). Application of ga feature selection on naive bayes, random forest, and svm for credit card fraud detection, *International Conference on Decision Aid Sciences and Application (DASA)* pp. 1091–1097.

Sailusha, R., Gnaneswar, V., Ramesh, R. and Rao, G. R. (2020). Credit card fraud detection using machine learning, *2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS)*, pp. 1264–1270.

Stacking in Machine Learning (2023).
URL: *https://www.javatpoint.com/stacking-in-machine-learning*

Tomar, P., Shrivastava, S. and Thakar, U. (2021). Ensemble learning based credit card fraud detection system, *2021 5th Conference on Information and Communication Technology (CICT)*, pp. 1–5.

Varmedja, D., Karanovic, M., Sladojevic, S., Arsenovic, M. and Anderla, A. (2019). Credit card fraud detection - machine learning methods, *2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH)* pp. 1–5.

Wang, S., Dai, Y., Shen, J. et al. (2021). Research on expansion and classification of imbalanced data based on smote algorithm, *Scientific Reports* **11**: 24039.