Advancing Credit Card Fraud Detection: A Deep Learning Approach for Improved Risk Management

Sangbing Tsai*

International Engineering and Technology Institute, Hong Kong; klj0418@gmail.com

*Corresponding Author: klj0418@gmail.com

DOI: https://doi.org/10.30209/IJMO.202503.001

Submitted: Dec. 04, 2024 Accepted: Mar. 05, 2025

ABSTRACT

With increasing transaction volumes, traditional fraud detection methods are becoming inadequate against complex and sophisticated fraudulent tactics. Particularly excelling in processing large datasets and recognizing intricate patterns. This study proposes a risk detection to enhance fraud identification and prevention capabilities. The model integrates three technologies: BERT, CNN, and LSTM. BERT extracts deep features from transaction texts, CNN excels in pattern recognition, and LSTM effectively processes the time series aspects of transaction data. This combination enables the model to understand the complex semantics behind transactions and capture potential fraudulent patterns and temporal correlations. The model was compared with traditional methods and other deep learning models in experiments. The results show superior detection accuracy and speed, outperforming other approaches. By real-time identification of fraudulent transactions, the model improves efficiency and accuracy in risk detection and provides financial institutions with a powerful tool to protect customer assets and reputation. This highlights the immense potential and practical value of deep learning in financial risk management.

Keywords: Deep Learning; Credit Card Fraud; BERT; CNN; LSTM; Fraud Detection

1. Introduction

Credit cards have become integral to modern payment systems[1, 2], offering various functionalities such as consumption, savings, transfers, and withdrawals, significantly facilitating users' daily transactions [3]. However, the widespread adoption of credit cards has also introduced new challenges, particularly in fraud detection[4, 5]. Traditional risk control methods, such as rule-based systems and manual reviews, are often inefficient and struggle to scale with the increasing volume of data, rendering them inadequate in the face of evolving and increasingly sophisticated fraudulent activities. As a result, deep learning technologies have become essential in addressing these challenges. These technologies are particularly well-suited to detecting fraud patterns, significantly improving both the efficiency and accuracy of fraud detection[6, 7].

Several deep learning models and machine learning techniques, such as Neural Networks (NN), Support Vector Machines (SVM), Decision Trees, Random Forests (RF), and Long Short-Term Memory networks (LSTM), are commonly used in fraud detection [8]. Neural networks, renowned for their strong nonlinear modelling capabilities, excel at identifying complex relationships in large datasets but are often criticized for their lack of interpretability[9]. Support Vector Machines, effective with smaller datasets, show good generalization abilities but can be computationally inefficient when applied to larger datasets[10]. Decision Trees are helpful for initial data analysis and offer ease of interpretation, yet they are prone to overfitting. Random Forests, an ensemble learning method, perform well on high-dimensional data and exhibit strong resistance to noise, though they may underperform on some complex datasets compared to deep learning models[11]. LSTMs are especially suited for handling time-series data and capturing temporal dependencies and long-term patterns in transaction behaviour, although they require significant computational resources for training. Each model has strengths and limitations, playing an important role in credit card fraud detection practices[12]. With ongoing technological advancements and increasing data volumes, effectively combining the advantages of these models or developing new models for more effective identification and prevention of credit card fraud represents a key direction for future research.

Credit card fraud remains one of the most pressing challenges in the global financial sector, with fraudulent activities becoming increasingly sophisticated in both scale and complexity. As payment systems evolve and the volume of online transactions continues to rise, financial institutions face more significant risks. Industry reports estimate that credit card fraud costs the global economy billions of dollars annually, impacting consumers and financial institutions. Fraudulent activities, such as unauthorized transactions, card theft, and counterfeiting, are major contributors to this issue. As fraud tactics grow more advanced, traditional fraud detection methods are increasingly inadequate in effectively identifying and preventing these activities.

Historically, credit card fraud detection relied heavily on rule-based systems and simple machine-learning models. While these approaches laid the groundwork for detecting suspicious transactions, they were often limited in capturing complex and evolving fraud patterns, mainly when dealing with large-scale datasets. Traditional methods typically depend on predefined rules and feature engineering, which makes them less adaptable to new or emerging fraud strategies. Furthermore, the class imbalance in fraud detection datasets—where legitimate transactions vastly outnumber fraudulent ones—poses significant challenges for these models in terms of both accuracy and robustness.

In light of the rapid advancements in financial technology and the evolving nature of fraudulent tactics, we propose a hybrid deep learning model that integrates BERT, CNN, and LSTM technologies. This model combines each approach’s strengths to enhance detection accuracy and efficiency. BERT analyses transaction text data in-depth, extracting key semantic information. CNN is used to identify complex patterns indicative of fraudulent activities. Meanwhile, LSTM focuses on analyzing the time-series nature of transaction data, capturing long-term dependencies and dynamic changes in transaction behaviour.

The three main contributions of our study are as follows: First, by integrating BERT, CNN, and LSTM, our model achieves significant advancements in understanding and analyzing complex transaction data, mainly when dealing with transactions that contain rich textual information. Second, compared to existing single-model approaches or traditional methods, our model demonstrates higher accuracy and a lower false positive rate in experimental tests, indicating its superior ability to identify genuine fraudulent transactions. Lastly, our model proposes an innovative hybrid approach and demonstrates its practical application using real-world data. This provides financial institutions with a powerful tool to safeguard consumer assets and reputations, thereby enhancing the security and stability of the broader financial system.

2. Related work

Data processing and feature engineering are pivotal in credit card fraud detection. [13, 14] Their primary purpose is to ensure the extraction of the most valuable information from raw data while minimizing the interference of noise and irrelevant features, thereby directly impacting the performance and accuracy of the final model. The data preprocessing stage initially encompasses key steps such as data cleaning, handling missing values, normalization, and standardization. For instance, incomplete or erroneous records in transaction data must be cleaned or corrected to ensure the data quality for model training. Following this, the feature selection and extraction phase involves identifying and utilizing features most conducive to predicting fraudulent behaviour. This process may include deriving new features from the original data or using dimensionality reduction techniques, including Principal Component Analysis (PCA), to reduce the volume of data[15]. For example, analyzing transaction histories can reveal patterns in consumer behaviour, or scrutinizing the timing and location of transactions can identify unusual activity patterns. Appropriate encoding methods (like one-hot encoding) are necessary for categorical data, such as transaction types or country codes, to ensure that the model correctly interprets and utilizes these variables [16].

Despite the significant advantages of data processing and feature engineering in enhancing the model's learning efficiency and predictive accuracy, there are also challenges. The primary challenge is the risk of losing important information during the feature engineering process, especially during feature selection and dimensionality reduction [17]. Effective data processing and feature engineering require specialized knowledge, particularly in interpreting and understanding data, which can be time-consuming and complex. However, considering their importance in accurately identifying and preventing credit card fraud activities, these steps remain indispensable in the fraud detection system.

In summary, credit cards have made significant progress, particularly in handling class imbalance, improving model interpretability, and adapting to the rapidly changing nature of fraudulent tactics. Our research builds upon these existing approaches by combining advanced techniques such as BERT, CNN, and LSTM to offer a more comprehensive and effective solution for detecting credit card fraud. Additionally, we aim to address some of the limitations of previous models, particularly regarding class imbalance and model interpretability, through novel methodologies and experimental analysis.

Another important trend in fraud detection is ensemble learning and multi-model fusion. These approaches combine the strengths of multiple models to handle complex datasets and enhance prediction accuracy. For example, Kokkinaki [18] combined decision trees with cluster analysis to distinguish between normal and fraudulent transactions. Similarly, Wang et al. [19] integrated various classifications, CART and Ripper, with Naïve Bayes using a stacking model to improve performance. Random Forests, Boosting, and Bagging combine multiple decision trees to improve detection accuracy.

Additionally, different models can be combined to leverage each other's unique strengths, improving overall prediction power [20]. Furthermore, ensemble learning can be integrated with feature engineering techniques, such as using feature selection methods to reduce dimensionality before applying multiple models. The primary benefit of ensemble learning is its ability to increase accuracy and reduce bias, leading to more reliable fraud detection [21]. Moreover, due to the integration of multiple models, this method is generally more robust than single-model reliance, reducing the risk of overfitting and maintaining high performance across various datasets. However, ensemble learning also has its drawbacks. Primarily, this method can require more computational resources and time, especially when training and tuning multiple models. Furthermore, the integration of models can lead to increased complexity, reducing the interpretability of the model. In some cases, integrating too many models might lead to diminishing marginal returns in performance.

These models are well-suited for handling the complexity of credit card transaction data. For instance, Artificial Neural Networks (ANNs) have been successfully used to model complex nonlinear relationships in transaction data, improving fraud detection accuracy[22]. While Convolutional Neural Networks (CNNs) are traditionally applied to image processing, they have also demonstrated effectiveness in identifying patterns within data that have spatial relationships. Analyzing time-series data enables capturing temporal patterns in transaction behaviour-an essential feature for fraud detection.

The main advantages of these deep learning models lie in relevant features from tasks that are difficult to achieve through manual feature engineering. Neural networks excel at handling nonlinear relationships and high-dimensional data, making them highly effective at identifying complex fraud patterns. Models such as LSTMs are especially valuable in detecting emerging fraudulent tactics by adapting to dynamic transaction data. However, deep learning models come with certain drawbacks. They often require significant computational resources, especially during training [23].

3. Materials and Methods

3.1 Overview

This study proposes a deep learning-based model for credit card fraud detection, structured into three key components, as illustrated in Figure 1:

(1) BERT Feature Extraction Module: This module processes anonymized transaction text data as its input to extract deep features. Using BERT, this module captures the semantic complexity of transaction texts. The BERT pre-trained is used to derive rich feature representations that facilitate understanding the intricacies of transaction data.

(2) CNN Pattern Recognition Module: The second module focuses on identifying patterns within the data. It receives input from the features extracted by BERT and the original transaction data. Convolutional Neural Networks (CNNs) are employed due to their ability to detect local features and patterns. CNNs are particularly effective at identifying associations within transaction data, making them suitable for fraud detection by recognizing anomalous behaviours and potential fraud patterns.

(3) LSTM Temporal Processing Module processes time-series data, including transaction timestamps and sequence-related information.

These modules work to create an integrated model capable of comprehensive fraud detection. Combining feature extraction, pattern recognition, and temporal processing, this deep learning model provides a robust solution for identifying fraudulent transactions, offering financial institutions an advanced tool to combat fraud and improve risk management practices.

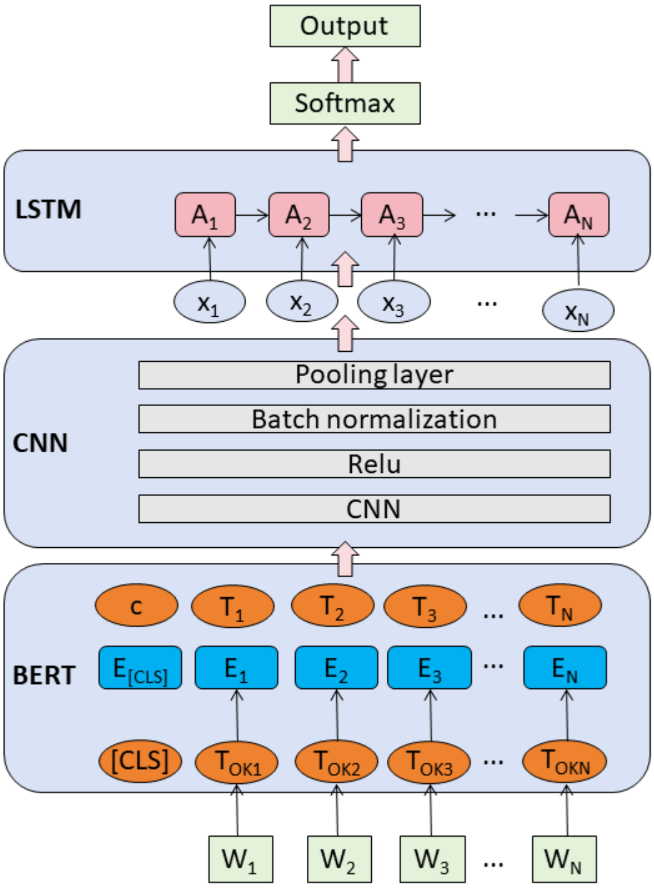

Fig. 1 Overview of our model

3.2 Bidirectional from Transformers Encoder Representations

BERT is a key component in the proposed fraud detection model for credit card transactions[24]. Leveraging its exceptional natural language processing (NLP) capabilities, BERT is critical in extracting deep semantic features from transaction text data. Built on the Transformer architecture, BERT is particularly effective in sequence-to-sequence tasks, where understanding context is essential[25]. Its bidirectional nature allows it to consider both the left and right contexts of each word, making it highly suitable for capturing complex semantics in transaction descriptions (Fig 2).

BERT processes textual transaction data, which may contain crucial information about fraudulent activities. By pre-training on vast text corpora, BERT learns general language semantics and relationships between words, enhancing its ability to extract meaningful features from transaction descriptions. The output of BERT consists of high-dimensional vector embeddings for each word, representing the semantic meaning of the transaction details. These embeddings are crucial for identifying potential fraud patterns and forming essential input features for the subsequent modules.

The BERT model uses a word based on their relevance in context. Mathematically, the attention mechanism is computed by comparing, with the attention scores being normalized using a softmax function to highlight the most relevant words. This mechanism helps BERT focus on the important elements of transaction texts for fraud detection.

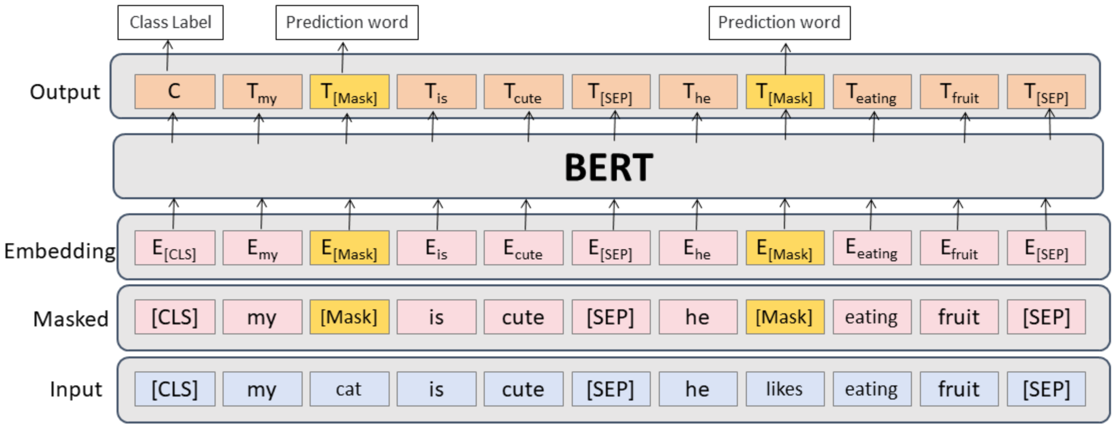

Fig 2. BERT model structure

Mathematically, BERT employs a self-attention mechanism, which calculates weighted sums of values based on the relevance between queries, keys, and values. This mechanism can be represented as follows:

$$\text{Attention}(Q, V, K) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad \text{[Formular 1]}$$

Here, Q, K, and V are derived from BERT, where Q corresponds to the query embedding for a specific word in the text, while K and K encompass the key and value embeddings for all words. The contextual embeddings generated by BERT serve as crucial input features for subsequent modules within the model. These modules further analyze and identify fraud patterns within transaction data, leveraging the enriched semantic information provided by BERT.

3.3 Convolutional Neural Networks

CNNs are integral for pattern recognition in the fraud detection model, especially for handling both image-like data and sequential data[26].In this context, the primary function of CNNs is feature extraction through convolution operations. A typical CNN consists of several convolutional layers, which detect local features in the data, and pooling layers, which reduce dimensionality while preserving key information. The CNN's ability to learn the optimal convolutional kernel weights enables it to detect relevant features in transaction data, which are essential for identifying fraud patterns [27-29].

The core operation of a CNN is the convolution operation, mathematically represented as follows:

$$(S * K)(i, j) = \sum_m \sum_n S(i + m, j + n)K(m, n) \qquad \text{[Formular 2]}$$

S represents the input data, K represents the convolutional kernel, (i, j) denotes the coordinates of the output position, and (m, n) represents the coordinates within the convolutional kernel.

S: Input data, typically a two-dimensional matrix representing an image or a feature map.

K : Convolutional kernel, also a two-dimensional matrix that contains learnable weight parameters; (i, j): Coordinates of the output position, indicating the location of the convolution operation's result on the feature map.(m, n): Coordinates within the convolutional kernel, used for sliding the kernel over the input data.

3.4 Long Short-Term Memory

LSTM networks are a specialized form of Recurrent Neural Networks (RNNs) designed to handle sequential data, particularly useful for modeling time series data like financial transactions[30]. Regulate information flow through the network. These gates enable LSTMs to retain important information. The LSTM module processes time-dependent transaction data, capturing patterns and correlations across multiple transactions. This temporal aspect is crucial for detecting fraudulent behaviour that may evolve. The LSTM's memory cells maintain a state that reflects relevant information from past transactions, allowing the model to make informed predictions about future transactions. This helps identify potential fraud activities, which may unfold progressively across a series of transactions (Fig 3).

1)Input Gate (i_t):

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad \text{[Formular 3]}$$

- i_t Is the input gate's activation.
- W_i and b_i are weight parameters and biases for the input gate.
- σ is the sigmoid activation function.

2) Forget Gate (f_t):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad \text{[Formular 4]}$$

- f_t is the forget gate's activation.
- W_f and b_f are weight parameters and biases for the forget gate.

3) Update Memory Cell (C_t):

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \quad \text{[Formular 5]}$$

- $\tilde{C}_t$ is the candidate cell state.
- W_C and b_C are weight parameters and biases for the candidate cell state.

4) Memory Cell (C_t) Update:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t \quad \text{[Formular 6]}$$

- C_t is the updated cell state.

5) Output Gate (o_t):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad \text{[Formular 7]}$$

- o_t is the output gate's activation.
- W_o and b_o are weight parameters and biases for the output gate.

6) Hidden State (h_t) Update:

$$h_t = o_t \cdot \tanh(C_t) \quad \text{[Formular 8]}$$

h_t is the hidden state.i_t : Input gate activation, f_t: Forget gate's activation, controls what information should be discarded from the previous cell state.C_t: Cell state, stores information over time.$\tilde{C}_t$: Candidate cell state, a proposed update to the cell state.o_t: Output gate activation, determines what information should be output as the hidden state.

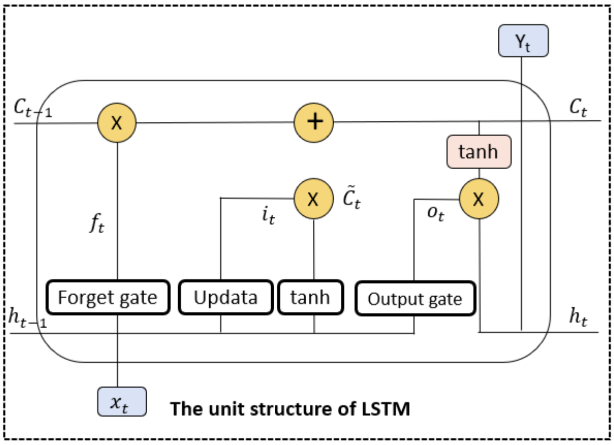

Fig 3. LSTM model structure

4. Experiment

In this study, the operating system used in our laboratory is Ubuntu 20.04, a widely adopted operating system known for its rich development and computing resources. Python 3.8 was chosen as the primary programming language. On the hardware front, we employed an Intel(R) Xeon(R) CPU E5-2680 v4 @ 2.40GHz as the primary Central Processing Unit (CPU). Additionally, we configured an NVIDIA Tesla M60 as the Graphics Processing Unit (GPU). To handle large-scale datasets and complex computational tasks, we equipped our setup with a substantial amount of memory, totalling 284GB and 8GB of graphics memory (VRAM).

4.1 Dataset

This dataset was provided through a collaboration between Université Libre de Bruxelles and a machine learning research team. To preserve the confidentiality of sensitive credit card transaction data, the dataset has undergone desensitization and dimensionality reduction using Principal Component Analysis (PCA). It contains transaction data from European cardholders recorded over two days in September 2013. This dataset is particularly significant for credit card fraud detection, given the global financial impact of fraud and the need for effective detection methods to minimize losses.

Credit card fraud takes many forms, including card theft, identity theft during applications, and card forgery. Card forgery accounts for most fraud cases, typically carried out by organized groups involved in stealing card data, producing counterfeit cards, and using them for illicit transactions. Fraud detection is vital in helping financial institutions reduce financial damage. The dataset consists of 284,807 transactions, with 492 labelled as fraudulent, making up only 0.172% of the total dataset (Fig. 4). This imbalance presents a significant challenge for fraud detection. The dataset includes the following attributes: “Time” represents the time elapsed from the first transaction, “V1-V28” are principal components from PCA used for dimensionality reduction, “Amount” denotes the transaction amount, and “Class” is a binary variable indicating whether a transaction is normal (0) or fraudulent (1). The dataset has no missing values, but due to the low number of fraudulent transactions, the class imbalance may affect model performance.

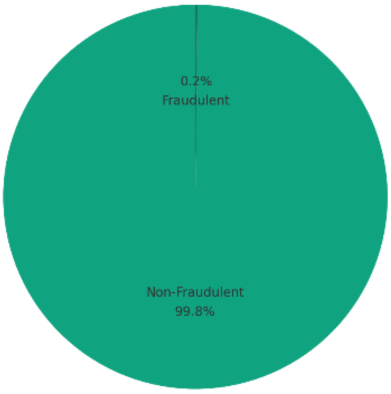

Fig 4. Distribution of Credit Card Transactions

4.2 Experimental Design

4.2.1 Data preparation and preprocessing

(1) Handling Missing Values: We first check for missing values in the dataset. Fortunately, this credit card fraud detection dataset has no missing values to deal with. This means that we can proceed with modelling using complete data without the need for imputation or removing missing values.

(2) Data Standardization: In credit card fraud detection, it is common practice to standardize the data to ensure the model's consistent performance. Standardization involves transforming the data with a mean of zero and a standard deviation of one. This helps to bring features to the same scale and aids in model convergence and performance. We applied standardization to the "Amount" feature to maintain a consistent numerical range.

(3) Feature Analysis and Processing: During data preprocessing, we conducted feature analysis, including statistical analysis and visual exploration of each feature. This helped us gain insights into the data distribution and relationships between features. We had already performed PCA dimensionality reduction for the principal component data ("V1-V28"), with these principal components representing the key features of the original data. These principal components were used as input features during modelling.

(4) Data Format Transformation: Due to the significant, we applied an oversampling strategy to address this issue. Oversampling is a technique that balances the number of fraudulent and non-fraudulent transaction samples by increasing the number of fraudulent samples. Specifically, we used a technique that generates synthetic fraudulent transaction samples to augment the number of fraudulent transactions, thereby enabling the model to detect fraudulent transactions, ensuring that the model is not biased towards the majority class of normal transactions. Data preprocessing also involves converting the data into a format suitable for modelling. In this dataset, besides numerical features, we have a binary classification variable ("Class") indicating whether a transaction is fraudulent. We ensured this variable was correctly encoded, typically using one-hot or binary encoding (0 and 1).

Through these data preprocessing steps, we ensured data integrity, consistency, and suitability, laying a solid foundation for the subsequent development of the credit card fraud detection model. These steps improve model accuracy and interpretability, enabling better capture of potential fraudulent activities.

4.2.2 Model building and prediction

1) Data Preprocessing: First, the credit card transaction data undergoes data preprocessing. This includes data splitting, data augmentation, and format conversion. Data splitting divides the dataset into training and testing sets, typically using an 8:2 ratio. Data augmentation employs SMOTE and ENN methods to address data imbalance, increasing the number of fraudulent transaction samples. Format conversion transforms the text data into a format acceptable to the model.
2) BERT Encoder: The first part of the model is the BERT encoder, which is used to transform the input text sequences into vector representations. The BERT model employs self-attention mechanisms, enabling it to capture crucial information and semantics within the text. This step converts each word and subword in the text sequences into vector form.
3) CNN Layer: Next, convolutional (CNN) and pooling operations are applied to the BERT outputs to concatenate local feature vectors into a global feature vector. Convolution operations help identify local patterns and features within the text, while pooling operations combine these local features into a worldwide feature vector, enhancing the model's representational capacity.
4) LSTM Layer: The global feature vector is input into an LSTM model. LSTM models excel in handling time-series data and help capture temporal correlations in credit card transactions. They can remember and utilize previous information to predict future events.
5) Fully Connected Layer: A fully connected layer maps the LSTM's output to a vector of length 2. Each element in this vector represents the probability of fraud and non-fraud. For example, the first element represents the probability of fraud, and the second represents the probability of non-fraud. The output from this layer is used for the final classification decision.

Table 1. Model Parameter Settings.

Parameter	**Description**	**Setting**
Data Split Ratio	Proportion of data split into training and testing	Training: 80%, Testing: 20%
Data Augmentation Method	Data augmentation techniques used	SMOTE and ENN
BERT Model	Configuration and parameters of the BERT model	Pretrained Model: BERT-Base Learning Rate: 0.0001 Batch Size: 32 Number classes: 2
CNN Layer	Configuration and parameters of the Convolutional Neural Network layer	Kernel Size: 3x3 Pooling Size: 2x2 Activation Function: ReLU
LSTM Layer	Configuration and parameters of the Long Short-Term Memory layer	Hidden Units: 128 Activation Function: tanh
Fully Connected Layer	Configuration and parameters of the fully connected output layer	Output Units: 2 Activation Function: Softmax
Optimizer	Optimizer type used for model training	Adam Optimizer
Learning Rate	Initial learning rate	0.0001
Loss Function	Loss function used for calculating model loss	Cross-Entropy Loss

Number of Iterations	Total number of iterations for model training	1000 iterations
Batch Size	Number of samples per batch during training	32
Data Format Conversion	Conversion of text data to a format suitable for the model	Text Vectorization

4.3 Comparison of Study Results and Analysis

To evaluate the performance of the BERT model in credit card fraud detection, a set of controlled experiments was conducted. Traditional machine learning models, including Random Forest (RF), Decision Tree (DT), Logistics Regression (LR), Gradient Boosting Decision Tree (GBDT), and Extreme Gradient Boosting (XGBoost), were selected for comparison experiments. The controlled variable method was employed to ensure the experiments' accuracy and fairness. This means the same hardware environment, dataset, and data preprocessing methods were used to train and test all five models.

Table 2. Experimental Results.

Model	Average Accuracy	Average Precision	Average Recall	Average F1 Score	AUC Value
BERT-CNN-LSTM	**0.9791**	**0.9755**	**0.9816**	**0.9784**	**0.9884**
RF	0.9222	0.9068	0.9376	0.9216	0.9275
DT	0.9101	0.8946	0.9264	0.9105	0.9166
LR	0.9586	0.9523	0.9633	0.9577	0.9726
GBDT	0.9179	0.9046	0.9304	0.9148	0.9224
XGBoost	0.9190	0.9072	0.9325	0.9196	0.9263

In Table 2, we can clearly observe that the BERT-CNN-LSTM model excels in all evaluation metrics.

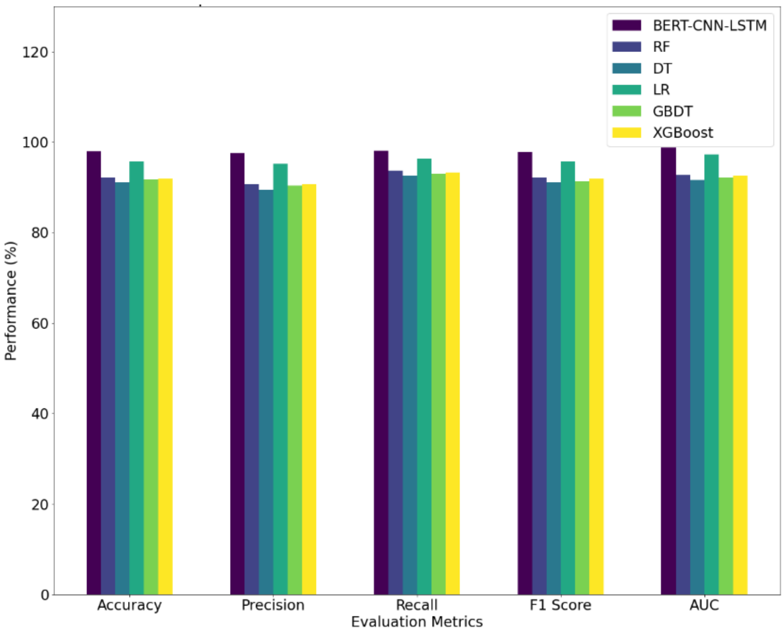

Fig 5. Comparison between different models

Compared to traditional machine learning models, the BERT-CNN-LSTM model does not require complex parameter tuning or expert-driven feature selection. It possesses better self-learning capabilities, enabling it to capture variable-target relationships accurately. Notably, the superiority of the BERT-CNN-LSTM model becomes even more evident when dealing with imbalanced data and high-dimensional samples. While the traditional Logistic Regression (LR) model performs well in some metrics, this may be attributed to introducing an l2 penalty term, which enhances the model's generalization ability. However, overall, the BERT-CNN-LSTM model outperforms all other models regarding performance metrics. For instance, in terms of accuracy, the BERT-CNN-LSTM model achieves an astounding 97.91%, ranking first among all models.

Additionally, the model exhibits impressive precision, recall, and F1-score, reaching 97.55%, 98.16%, and 97.84%, respectively.

In summary, the BERT-CNN-LSTM model demonstrates outstanding performance in credit card fraud detection, providing financial institutions with a more reliable safeguard against fraudulent activities. It efficiently detects fraud and reduces risks and losses. Compared to traditional machine learning methods, this model exhibits significant advantages in various aspects (Fig. 5).

4.4 Ablation Study Results and Analysis

Table 3. Comparative Performance Metrics between different models

model	Accuracy	Precision	Recall	F1	AUC
BERT	0.9075	0.8652	0.8213	0.8425	0.9243
BERT-CNN	0.9321	0.8974	0.8469	0.8713	0.9402
BERT-LSTM	0.9496	0.9137	0.8326	0.8702	0.9468
BERT-CNN-LSTM	**0.9791**	**0.9755**	**0.9816**	**0.9784**	**0.9884**

A series of ablation experiments were conducted to evaluate the impact of different models on the performance of credit card fraud detection. These experiments included training individual models such as BERT, BERT-CNN, BERT-LSTM, and a combined BERT-CNN-LSTM model. It was observed that using the BERT model alone yielded subpar results across all performance metrics, particularly in terms of recall, which exhibited the poorest performance (Table 3). This suggests that the BERT model tends to misclassify genuine fraud transaction samples as normal transactions, leading to a high false-negative rate. This finding underscores that a single natural language processing model may not suffice to capture essential features effectively for credit card fraud detection.

We introduced two different deep learning components, CNNs and LSTM. Results revealed that the BERT-CNN model demonstrated improvements in all metrics over the BERT model, with notable gains in precision and AUC values by approximately 2% and 1.5%, respectively. Similarly, the BERT-LSTM model exhibited performance enhancements over BERT, particularly in accuracy, precision, and AUC values, with precision showing an improvement of approximately 3%. These outcomes highlight the beneficial contributions of CNN and LSTM in better feature extraction and enhanced classification performance. However, the most impressive results were observed with the BERT-CNN-LSTM model, which integrated both CNN and LSTM. It outperformed all other models in all metrics, achieving accuracy, precision, recall, F1-score, and AUC values of 0.979, 0.975, 0.981, 0.978, and 0.988, respectively. These metrics substantially exceeded those of the other models. This underscores that the BERT-CNN-LSTM model excels at feature extraction and exhibits superior integration capabilities within the classifier, resulting in exceptional performance in credit card fraud detection (Fig. 6).

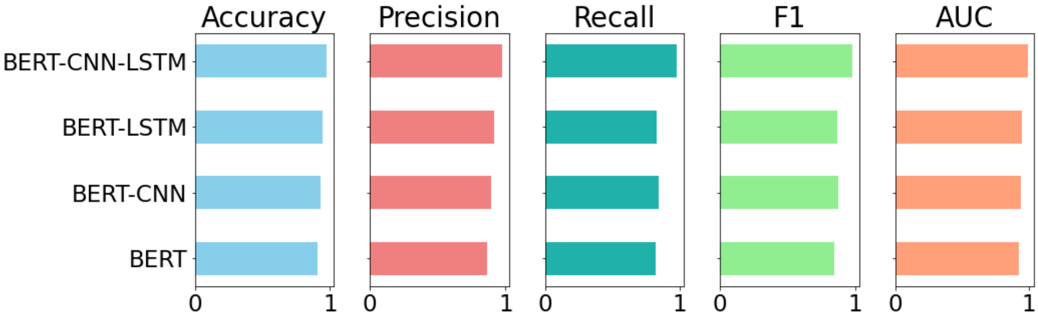

Fig 6. Comparison between different models

4 Conclusion

Traditional fraud detection methods have proven inadequate in addressing the evolving nature of fraudulent activities. To tackle this issue, we propose a deep learning-based credit card fraud detection model designed to enhance fraud detection capabilities, especially when dealing with large-scale transaction data and complex fraud patterns. Traditional methods often struggle with these challenges, which led to the development of a deep learning solution that integrates BERT, CNN, and LSTM to improve performance. BERT is used to extract deep features from transaction texts, CNN is used for pattern recognition, and LSTM handles the temporal aspects of transaction data, allowing the model to capture complex fraud patterns and temporal dependencies.

We compare our model with traditional fraud detection methods and other deep learning models through a series of experiments. The results show that our model outperforms others in accuracy and processing speed. It is capable of real-time fraud detection, enhancing the efficiency and accuracy of risk detection while providing financial institutions with a powerful tool to safeguard customer assets and reputation. This study highlights the potential of deep learning in financial risk management, particularly in credit card fraud detection. We offer an innovative solution to this critical issue by combining multiple deep learning techniques.

However, one limitation of this study is the failure to address the class imbalance in credit card fraud detection, where legitimate transactions vastly outnumber fraudulent ones. This imbalance can impact model training. Future work will focus on advanced techniques for managing class imbalance, such as oversampling, undersampling, or synthetic data generation, to improve model performance. Additionally, deep learning models, including the one presented here, often lack interpretability. Understanding the decision-making process and identifying key features influencing predictions remains a challenge. Future research should aim to improve model interpretability, providing actionable insights to financial institutions and ensuring compliance with regulatory requirements.

In conclusion, this study presents a deep learning-based credit card fraud detection model that utilizes BERT, CNN, and LSTM to address increasingly complex fraud issues. Through experimentation, we have demonstrated that our model outperforms traditional and other deep learning methods regarding accuracy and speed. This research provides an innovative and effective solution to a pressing issue in the global financial sector, making significant contributions to financial risk management. This approach is expected to enhance fraud detection capabilities, protect customer assets, safeguard institutional reputations, and underscore the transformative potential of deep learning in combating complex fraud schemes.

References

[1] Liu, J., Kauffman, R.J. and Ma, D. Competition, cooperation, and regulation: Understanding the evolution of the mobile payments technology ecosystem. Electronic Commerce Research and Applications, 2015, 14(5), 372-391.

[2] Sumanjeet. Emergence of payment systems in the age of electronic commerce: The state of art. In: 2009 First Asian Himalayas International Conference on Internet, 2009.

[3] Maurer, B. Mobile money: Communication, consumption and change in the payments space. Journal of Development Studies, 2012, 48(5), 589-604.

[4] Zojaji, Z., Atani, R.E. and Monadjemi, A.H. A survey of credit card fraud detection techniques: Data and technique oriented perspective. arXiv preprint arXiv:1611.06439, 2016.

[5] Lebichot, B., Le Borgne, Y.-A., He-Guelton, L., Oblé, F. and Bontempi, G. Deep-learning domain adaptation techniques for credit cards fraud detection. In: Recent Advances in Big Data and Deep Learning: Proceedings of the INNS Big Data and Deep Learning Conference INNSBDDL2019, held at Sestri Levante, Genova, Italy, 16–18 April 2019.

[6] Zioviris, G., Kolomvatsos, K. and Stamoulis, G. Credit card fraud detection using a deep learning multistage model. The Journal of Supercomputing, 2022, 78(12), 14571-14596.

[7] Ashtiani, M.N. and Raahemi, B. Intelligent fraud detection in financial statements using machine learning and data mining: a systematic literature review. IEEE Access, 2021, 10, 72504-72525.

[8] Huang, F., Xiong, H., Chen, S., Lv, Z., Huang, J., Chang, Z. and Catani, F. Slope stability prediction based on a long short-term memory neural network: Comparisons with convolutional neural networks, support vector machines and random forest models. International Journal of Coal Science & Technology, 2023, 10(1), 18.

[9] Abdi, H., Valentin, D. and Edelman, B. Neural networks. Sage, 1999.

[10] Kremer, J., Steenstrup Pedersen, K. and Igel, C. Active learning with support vector machines. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2014, 4(4), 313-326.

[11] Papoutsoglou, G., Tarazona, S., Lopes, M.B., Klammsteiner, T., Ibrahimi, E., Eckenberger, J., Novielli, P., Tonda, A., Simeon, A. and Shigdel, R. Machine learning approaches in microbiome research: challenges and best practices. Frontiers in Microbiology, 2023, 14.

[12] Sagheer, A. and Kotb, M. Unsupervised pre-training of a deep LSTM-based stacked autoencoder for multivariate time series forecasting problems. Scientific Reports, 2019, 9(1), 19038.

[13] Saia, R. and Carta, S. Evaluating the benefits of using proactive transformed-domain-based techniques in fraud detection tasks. Future Generation Computer Systems, 2019, 93, 18-32.

[14] Al-Hashedi, K.G. and Magalingam, P. Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. Computer Science Review, 2021, 40, 100402.

[15] Kambhatla, N. and Leen, T.K. Dimension reduction by local principal component analysis. Neural Computation, 1997, 9(7), 1493-1516.

[16] Dahouda, M.K. and Joe, I. A deep-learned embedding technique for categorical features encoding. IEEE Access, 2021, 9, 114381-114391.

[17] Zhang, X., Han, Y., Xu, W. and Wang, Q. HOBA: A novel feature engineering methodology for credit card fraud detection with a deep learning architecture. Information Sciences, 2021, 557, 302-316.

[18] Kokkinaki, A.I. On atypical database transactions: identification of probable frauds using machine learning for user profiling. In: Proceedings 1997 IEEE Knowledge and Data Engineering Exchange Workshop, 1997.

[19] Wang, H., Fan, W., Yu, P.S. and Han, J. Mining concept-drifting data streams using ensemble classifiers. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003.

[20] Zekić-Sušac, M., Has, A. and Knežević, M. Predicting energy cost of public buildings by artificial neural networks, CART, and random forest. Neurocomputing, 2021, 439, 223-233.

[21] Ganaie, M.A., Hu, M., Malik, A., Tanveer, M. and Suganthan, P. Ensemble deep learning: A review. Engineering Applications of Artificial Intelligence, 2022, 115, 105151.

[22] Osegi, E. and Jumbo, E. Comparative analysis of credit card fraud detection in Simulated Annealing trained Artificial Neural Network and Hierarchical Temporal Memory. Machine Learning with Applications, 2021, 6, 100080.

[23] Hassija, V., Chamola, V., Mahapatra, A., Singal, A., Goel, D., Huang, K., Scardapane, S., Spinelli, I., Mahmud, M. and Hussain, A. Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation, 2023, 1-30.

[24] Cesar, L.B., Manso-Callejo, M.-Á. and Cira, C.-I. BERT (Bidirectional Encoder Representations from Transformers) for Missing Data Imputation in Solar Irradiance Time Series. Engineering Proceedings, 2023, 39(1), 26.

[25] Bao, T., Ren, N., Luo, R., Wang, B., Shen, G. and Guo, T. A BERT-based hybrid short text classification model incorporating CNN and attention-based BiGRU. Journal of Organizational and End User Computing (JOEUC), 2021, 33(6), 1-21.

[26] Karthika, J. and Senthilselvi, A. Smart credit card fraud detection system based on dilated convolutional neural network with sampling technique. Multimedia Tools and Applications, 2023, 1-18.

[27] Liang, F., Zhang, H. and Fang, Y. The analysis of global RMB exchange rate forecasting and risk early warning using ARIMA and CNN model. Journal of Organizational and End User Computing (JOEUC), 2022, 34(8), 1-25.

[28] Zhang, Z. Deep analysis of time series data for smart grid startup strategies: A Transformer-LSTM-PSO model approach. Journal of Management Science and Operations, 2024, 2(3), 16-43.

[29] Wan, Q., Zhang, Z., Jiang, L., Wang, Z. and Zhou, Y. Image anomaly detection and prediction scheme based on SSA optimized ResNet50-BiGRU model. Journal of Intelligence Technology and Innovation, 2024, 2(2), 35-52.

[30] Sherstinsky, A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network. Physica D: Nonlinear Phenomena, 2020, 404, 132306.