Computers and Chemical Engineering 182 (2024) 108560

Contents lists available at ScienceDirect

Computers and Chemical Engineering

journal homepage: www.elsevier.com/locate/compchemeng

Time series forecasting and anomaly detection using deep learning

Amjad Iqbal [a], Rashid Amin [a,b,*]

[a] Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan
[b] Department of Computer Science and Information Technology, University of Chakwal, Chakwal, Pakistan

ARTICLE INFO

Keywords:
Anomaly detection
Credit card
LSTM
Time series
Deep learning

ABSTRACT

Recent advances in time series forecasting and anomaly detection have been attributed to the growing popularity of deep learning approaches. Traditional methods, such as rule-based systems and statistical techniques, have limitations when applied to complex and dynamic real-world data. This study investigates using various deep learning models for anomaly detection, recognising aberrant patterns in data, and time series forecasting. The performance of the proposed models is evaluated on benchmarks like the Numenta Anomaly Benchmark (NAB) corpus and credit card fraud detection, showing their ability to detect aberrant patterns in various scenarios. Preprocessing strategies, such as normalisation and feature scaling, play a significant role in both time series forecasting and anomaly detection. In addition, the paper proposes a statistical method for selecting different or more important features from a dataset to overcome the limitations of high-dimensional sequencing data. In many ways, the suggested feature selection technique outperforms previous solutions. It keeps the original meanings of the attributes while selecting those with statistical relevance. Furthermore, it is computationally efficient and successfully solves the problem of excessive dimensions. Overall, deep learning approaches for time series forecasting and anomaly detection are promising in banking, healthcare, and manufacturing industries.

1. Introduction

Time series data is a collection of data points over time commonly utilised in finance, healthcare, manufacturing, and cybersecurity industries. Time series forecasting estimates future values created on prior observations, whereas anomaly detection is the discovery of anomalous or unexpected patterns in data. These data points might be numerical, categorical, or a combination. A time series is a fixed or changing succession of observations varying from seconds to years, depending on the application. Time series data is studied for trends, patterns, and anomalies. It can estimate future values, detect changes in data distribution, and monitor and regulate operations. The time series data is time-vary and can be analysed using statistical and machine learning based techniques (Deb et al., 2017). Deep learning, a form of machine learning that trains multi-layer neural networks, is a powerful tool for time series forecasting and anomaly detection (Rafique et al., 2023). Deep learning models provide accurate predictions and better detect abnormalities by capturing complex and nonlinear patterns in data. Statistical and machine learning approaches (Choi et al., 2021) are popular for studying time series data. A statistical technique is a moving average, which computes the average of a set of data points over a specific period. The Autoregressive Integrated Moving Average (ARIMA) (Noor et al., 2022) model captures the underlying structure of a time series by merging autoregression, differencing, and moving average components. The Holt-Winters Method (Elkourchi et al., 2023) is an exponential smoothing strategy for time series data considering level and seasonality.

This study investigates recurrent neural networks (RNNs) and convolutional neural networks (CNNs) as deep learning models for time series forecasting and anomaly detection (Koundal et al., 2023). RNNs are well-suited to time series forecasting because they can manage sequential data by recalling previous observations and utilising them to guide future forecasts. This study investigates several RNN versions, such as LSTM, LSTM-Autoencoder, GAN and Transformer and evaluates their performance against various benchmark datasets (Shaukat et al., 2023). Although there are various datasets, those can be used to evaluate the performance of the proposed work. However, due to its severeness and complexity, this focuses on anomaly detection from the benchmark credit card dataset. Moreover, we have gone through several news reports showing daily credit card fraud events. Since credit cards are a common payment mechanism used for online and offline transactions (Beju and Făt, 2023).

* Corresponding author at: Department of Computer Science, University of Engineering and Technology, Taxila, Pakistan.
E-mail address: rashid.sdn1@gmail.com (R. Amin).

https://doi.org/10.1016/j.compchemeng.2023.108560
Received 30 June 2023; Received in revised form 30 November 2023; Accepted 17 December 2023
Available online 18 December 2023

When purchasing with a credit card in a physical store, the consumer must present a physical card to complete the transaction. An unauthorised person would need to receive the card from the user to perpetrate fraud, resulting in financial losses for both the consumer and the credit card company. The offenders were apprehended using victims' credit card information, which has the potential to minimise fraudulent transactions. Each credit user is given a unique code with evidence and data about the transaction, amount spent, and most recent purchase date. Any deviation from this pattern is considered fraudulent, with financial implications (Salazar et al., 2018). The usage of credit cards for online businesses has improved owing to the frequent use of the internet. However, this has also made credit card fraud more prevalent and targeted. Fraud is a type of criminal deception used to obtain financial gain, and credit cards are a common target for fraudsters (Gao et al., 2019). Credit card fraud is becoming increasingly common due to online events, for example, internet shopping, networked banking, and credit card transactions. Unfortunately, due to an absence of computer models and software capable of detecting such activity in real-time, fraudulent transactions are often identified after they have already occurred. This fraud jeopardises the assets of both users and cardholders, as well as the assets of the firms involved. Losses due to fraud climbed by 16.2 % between 2017 and 2018, according to the 2019 yearly report (Yee et al., 2018).

The advent of e-commerce has led to an increase in non-physical credit card purchases, which are especially prone to fraud due to lack of physical authentication mechanisms. Anti-fraud systems must adapt to these shifting techniques (Roy et al., 2018). A two-part technique is commonly used to identify credit card fraud: dubious management activities are discovered, and behaviours are analysed to establish whether fraud has happened. A computer programmer can complete the first part of this task, but a human element must be involved in fraud detection and prevention. Fraud detection has become increasingly important due to the number of false positives created by automated systems. Machine learning models have become useful in detecting anomalies and probable fraud, as they can train to discover trends that may signal fraudulent conduct. This can help prevent financial losses by early identification and avoiding financial losses. Autoencoders are in high demand due to their capacity to create typical behaviour data and discover deviations from the norm. An event-based or sequence-based strategy can search for anomalies in transaction data. This study's technique is based on an artificial intelligence model that has shown good results while creating picture data. A fresh study looked at the usage of LSTM in an oversampling technique and showed that it has a high sensitivity for detecting unbalanced data (Hasib et al., 2023).

While the LSTM method has demonstrated promise, it also has significant drawbacks (Ahmed et al., 2023). One concern is that during the early training of the LSTM, the discriminator and generator may fail to establish theoretic likeness. Another issue is that the model may have difficulties converging, resulting in case replication owing to a lack of diversity in the input data. These issues must be addressed for the LSTM approach to detect fraud to be effective. The loss function of the LSTM and the structure of text input render it unsuitable for processing discrete data, which is a significant difficulty (Fiore et al., 2019). To remedy this, a new classifier may be required. Experiment results show that training a classifier on a larger dataset is far superior to training on genuine data for detecting credit card fraud. Even though this structure is described in the background of credit card transactions, it is versatile and easily adaptable to various application domains. The paper's framework is intended to manage credit card transactions systematically and efficiently (Dal Pozzolo et al., 2017). The experimental material is described, including data preparation comparisons and experimental outcomes.

Finally, the research investigates the importance of preprocessing strategies in time series forecasting and anomaly detection. The authors investigated the effects of various preprocessing procedures, such as normalisation and feature scaling, on the performance of deep learning models and introduced a new method. The novelty of this paper is a statistical method for selecting different or more important features from a dataset to overcome the limitations of high-dimensional sequencing data. The main contribution of this paper is the following.

- A statistical method is proposed for selecting more important features from a dataset and overcoming the limitations of high dimensionality in data.
- We investigated several deep-learning models after their parameters are fine tuned.
- Transformer-based network architecure is proposed for anomaly detection in time series data, which gives significant results.
- An ensemble of different CNN models improves the system's performance.
- We provided thorough assessments of popular time series anomaly detection techniques in various contexts on various data sets.

The paper is ordered as follows. Section 2 presents the review of the related literature on anomaly detection. Section 3 provides comprehensive details of the proposed approach for detecting anomalies. Section 4 shows the detailed experimental results and findings of several models established by modifying the network architecture and parameters. The final part summarises the research and emphasises on the important findings.

2. Related work

Time series anomaly detection and forecasting have recently gained significant attention from researchers. Various techniques have been proposed recently. A comprehensive literature review of recently published work on time series anomaly detection and forecasting using deep learning is presented next. The authors in Ketepalli et al. (2022) used deep learning methods to detect anomalies in credit card transactions. They analysed the performance of several deep learning models, such as autoencoders and LSTM networks, and demonstrated the worth of deep learning approaches for detecting anomalies in credit card transactions. The study offered useful insights into using deep learning algorithms for anomaly identification in credit card transactions. This work presented two data-driven strategies for improving supply chain management choices in their study (Nguyen et al., 2021). The authors created a method for predicting multivariate time series data by combining LSTM networks and an LSTM-Autoencoder network-based approach utilising a one-class support vector machine technique. Results showed that the LSTM-Autoencoder-based technique beat the prior study's LSTM-based method for anomaly identification.

This study proposed a VAE-based oversampling approach for detecting credit card fraud (Tingfei et al., 2020). Experiment findings presented that the VAE-based oversampling strategy improved precision, F-measure, accuracy, and specificity metrics. Overall, the VAE-based oversampling approach can potentially improve classification network performance in detecting credit card fraud. The proposed AERFAD method combines autoencoders and random forests to detect credit card fraud (Lin and Jiang, 2020). This work showed improved results in accuracy and metrics such as true positive rate, true negative rate, and Matthew's correlation coefficient. The work focuses on the progress of Autoencoder for credit card fraud detection (Kurien and Chikkamannur, 2019). The authors in (Raza and Qayyum, 2019) used the credit card transaction anomaly dataset to create a 10-layer deep Variational Auto-Encoder (VAE) neural network technique and compare its performance to decision trees, support vector machines, and ensemble classifiers. The researchers developed a fraud detection model using two deep neural networks, a fully connected network and an Autoencoder (Wu and Wang, 2021). Three white-box explainers in the explanation module can understand the Autoencoder, discriminator, and overall detection model.

The study in Liang et al. (2022) used deep neural networks to test a

self-supervised learning technique called Tab-iForest for anomaly identification on tabular data. Using a sequential attention mechanism, the TabNet model chose the most significant characteristics, resulting in more efficient and interpretable learning. The chosen features were then employed for unsupervised anomaly detection tasks using the isolated forest model, which used random down-sampling to produce sub-samples of tiny data volumes. The authors in Silva et al. (2021) proposed a strategy for objective categorisation of three types of transactions: normal, local, and global irregularity, using adversarial autoencoders and machine learning algorithms. The researchers looked at dimensionality and clusters formed by a former Gaussian mixture in latent vector space. Results showed that some classifiers could employ latent features efficiently, resulting in higher or equivalent performance when all the original characteristics were included.

The study in Lesot and Revault d'Allonnes (2012) aimed to generate fraud profiles by identifying patterns in credit-card scams. To accomplish this work, the researchers designed a clustering method that blended an incremental variant of the linearised fuzzy c-medoids with hierarchical clustering. This approach could handle large collections of heterogeneous data with both categorical and numerical variables, but the time required to generate profiles for each consumer must be considered. The study Zamini and Montazer (2018) provided an autoencoder-based unsupervised fraud detection system. The system was tested on fraud transactions from European banks using an autoencoder with three hidden layers and k-means clustering. This strategy beat earlier approaches with an accuracy of 98.9 % and a true positive rate (TPR) of 81 %. This work reported in Ahmed et al. (2019) sought to create and test a real-time approach for detecting data exfiltration and tunnelling using DNS. The authors created a machine learning system to identify abnormalities in DNS requests and tested it on 10 Gbps data streams between two organisations.

Several approaches have been proposed to solve the issue of uneven data. These include the increasingly common oversampling, undersampling, and SMOTE. Other sampling approaches, like GANs, were also used to boost the amount of minority class samples in Fiore et al. (2019). Authors in Zhang et al. (2021) proposed a method for detecting data anomalies based on CNNs and the CNN-LSTMED Encoder-Decoder architecture. They encoded the time series data with a CNN and used recovered features as input to an LSTM to decode and output the order. The authors proposed Autoencoders and Constrained Boltzmann Machines (CBM) to identify fraud using unsupervised learning techniques (Pumsirirat and Liu, 2018). They assessed the system using measures such as the ROC-AUC curve and the confusion matrix on three credit card datasets from Australia, Europe, and Germany. The authors (Carcillo et al., 2018) explored credit card fraud detection using two massive parallel anomaly detection approaches: artificial neural networks (ANN) and Bayesian belief networks. The two important elements in the examination were ROC and credit card transaction data. The findings demonstrated that ANN and BBN were successful in detecting credit card fraud.

According to Santos and Ocampo (2018), ANNs have limitations in terms of scalability, but Shen and Bhattacharyya found that neural networks outperformed SVM and LR in identifying credit card fraud. Stewart proposed an innovative solution for using SVM with categorical data by employing a density measurement. In this work (Panchal and Verma, 2019), the authors employed a self-organising map (SOM), an unsupervised deep learning approach, to forecast potential future defaulters. SOM is a widely used method in various applications, including anomaly detection and dimensionality reduction. It can also represent and visualise high-dimensional data in two dimensions effectively. This article discusses a strategy created by Srivastava that uses Hidden Markov Models and produces great results while scaling well (Liu et al., 2017). Hawkins proposed anomaly detection using autoencoders in 2002. Autoencoders have been utilised for several purposes, including anomaly detection and spot irregularities in satellite telemetry data.

Autoencoders have also been shown to be capable of identifying faults in picture and video data. LSTM was developed to improve the performance of recurrent neural networks and has proven useful in sequence prediction (Ghosh and Reilly, 1994). It has a higher learning rate and can resolve complex artificial problems. Graves and colleagues demonstrated how LSTMs may be used to predict whole data sequences using an LSTM architecture with several layers (Brause et al., 1999). Because of their capacity to model sequences and capture long-term correlations, LSTMs are ideal for anomaly detection applications. Subsieve and colleagues' study (Seyedhossein and Hashemi, 2010) has indicated that including additional inputs into the LSTM model, such as news and economic indicators, can improve the accuracy of forecasting stock prices. Recent research (Pandey, 2017) has revealed that LSTM layers can predict sequences and outperform standard LSTM networks. The reconstruction error is then used to discover anomalies and determine forecast accuracy.

The study (Niimi, 2015) investigated the performance of different fraud detection systems, including Autoencoders, Boltzmann machines, and Bayesian belief networks, and presented their technique for identifying fraudulent transactions. It concluded with suggestions for further research, such as ensemble models and expert knowledge. Deep learning algorithms are used to notice fraudulent activities in credit card transactions. CNNs and RNNs are the most used, while autoencoders have been shown to record common credit card transaction patterns efficiently and detect irregularities. Attention processes have been added to deep learning models to improve performance by focusing on the most important data features. According to the results of the literature study, deep learning algorithms have the potential to expand the performance of credit card anomaly detection systems greatly. More research is needed, however, to fully exploit these algorithms' capabilities and address potential challenges, such as interpretability, in their application to real-world scenarios.

3. Proposed methodology

The proposed work is divided into dataset preparation, preprocessing, and anomaly detection using machine learning and deep learning. Each module is further partitioned into several other steps to carry out various important tasks. The first module will create or extract datasets to pinpoint the specific area of interest, while the second module will explore and implement several anomaly detection techniques. To train the model for anomaly detection, we will utilise a few selected datasets from the NAB actual corpus and some additional real and fake datasets.

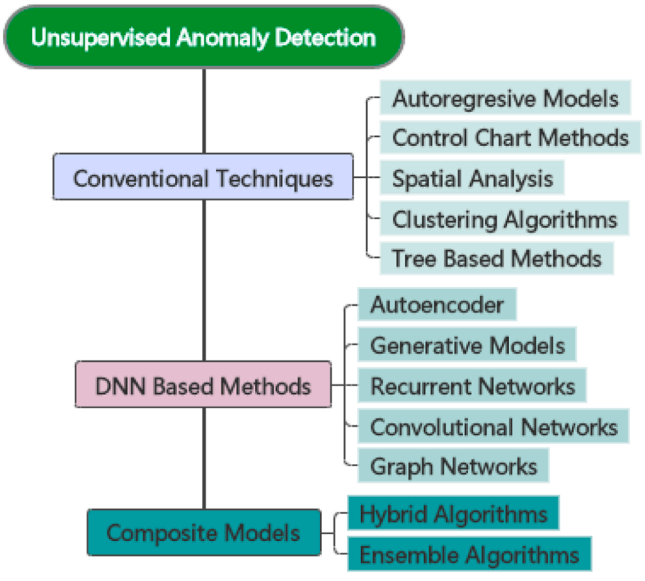

Fig. 1. The generic workflow of the proposed work.

Fig. 1 depicts the planned process in general. Fig. 1 has three branches, which further have subbranches. Firstly, conventional techniques, such as time series anomaly/outlier detection algorithms, have been employed in statistics and machine learning. This section summarises traditional time series anomaly detection approaches that paved the path for more modern data-intensive methods. Secondly, deep learning approaches have recently enhanced anomaly identification in high-dimensional datasets. These techniques can describe complicated, highly nonlinear inter-relationships between several sensors and efficiently capture temporal correlation. Thirdly, deep-learning technologies are combined with traditional methodologies from statistics and signal processing in hybrid algorithms. GRU-AEs and LSTM-AEs, EncDecAD, LSTM-VAEs, OmniAnomaly, Multi-Stage Convolutional Recurrent and Evolving Neural Networks (MSCRED), Convolutional Long-Short Term Memory (ConvLSTM) networks, Deep Autoencoding Gaussian Mixture Model (DAGMM), USAD, and Deep Support Vector Data Description (DeepSVDD) are recent hybrid methods. OmniAnomaly proposes a stochastic RNN model for detecting anomalies in multivariate time series, Multi-Stage Convolutional Recurrent and Evolving Neural Networks (MSCRED) that combines convolution with an LSTM in encoder-decoder architecture, Convolutional Long-Short Term Memory (ConvLSTM) networks effectively capture temporal patterns, and Deep Autoencoding Gaussian Mixture Model (DAGMM)

The research will be carried out using the methodology shown in Fig. 2. This process is divided into several essential steps. Tasks related to each step are carried out independently. Fig. 2 shows tasks that might be carried out in each step.

3.1. Data acquisition

The initial step is to collect time series data from benchmark data sources. The gathered data is now sent into the system, which will be utilised for mining data and analytics. This step is the main role in the time series anomaly detection technique.

3.2. Exploratory data analysis

Scientists use exploratory data analysis (EDA), typically paired with data visualisation tools to explore and analyse data sets. This enables users to manipulate data sources more efficiently to get desired outcomes, making it easier to identify patterns, spot anomalies, test hypotheses, and validate assumptions. Before modelling, EDA examines what the data can tell us. It is challenging to differentiate crucial data features while seeing a single column of numbers or a full spreadsheet. Basic statistics can be time-consuming, tedious, and intimidating, which is where exploratory data analysis tools come in handy.

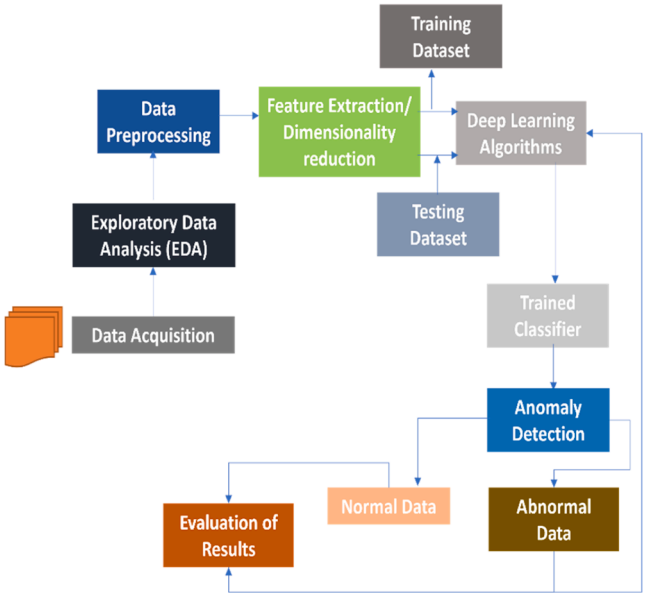

Fig. 2. An overview of proposed methodology for anomaly detection.

3.3. Data preprocessing

Data preprocessing is a central phase in time series anomaly detection data analysis. Data becomes increasingly difficult because of repeats and redundancies in time series. As a filtering approach, data normalisation employs data preprocessing. Examples of data preprocessing include normalisation, cleansing, timestamping, and seasonal trend decomposition. This activity included several completed actions to get the data in the required format. Fig. 3 describes the data preprocessing steps. As we see in Fig. 3, first, we acquire the dataset from the benchmark, then we do some EDA, and then we run preprocessing and clean data steps and check whether the data is in the form of a time series.

3.4. Feature extraction/ dimensionality reduction

Dimensionality reduction is an important procedure in which feature extraction is important in breaking down and reducing vast amounts of raw data into manageable categories. This leads in more efficient processing. The multiple variables inside these enormous datasets are the key components, and analysing them demands tremendous computer resources. Feature extraction is extracting the best features from huge datasets by choosing and combining variables, reducing data size. These features are simple to use and correctly portray the data. Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), the Statistical Method, and TSNE are examples of data feature extraction/ dimensionality reduction methodologies.

In this research, we propose a statistical metric for selecting differentiated or more informative features from a dataset to overcome the limitations of high-dimensional sequence data. The suggested method chooses the most important attributes, which resulted in better categorisation outcomes across diverse types of superior features derived from publicly available established benchmark datasets. The employed strategy is comparable to approaches for classifying sequences without alignment It is superior to the current classification in many ways. Additionally, the suggested method is straightforward, fast, dependable, and robust and requires a little training period. The following processes are used to identify characteristics capable of differentiating various features. The mathematical details of this strategy are shown below.

Denoted by X^i the ith class of data. X_k^i represents the k-th data of the class X^i. $k = 1, 2$. where N_i is the number of classes in the ith data. $X_k^i(j)$, $j = 1, 2, \ldots,$ is the feature vector representing the k-th class of the ith data. The proposed feature selection algorithm is implemented as follows:

After applying this algorithm, a data of n classes is represented by a matrix $n \times m$, where n is the number of classes and m is the number of features.

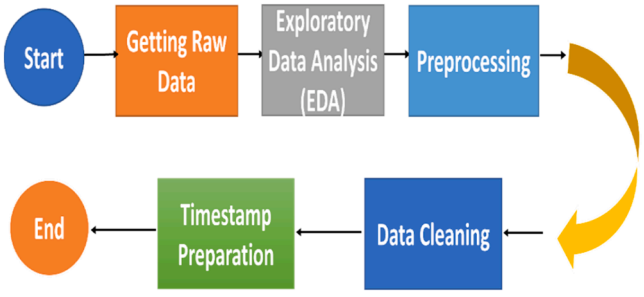

Fig. 3. Different phases/steps under data preprocessing.

Calculate the mean and variance of each data class column by column. As a result, the mean vector for each class is determined as follows:

$$\overline{X}^i(j) = \frac{\sum_{k=1}^{N_i} X_k^i(j)}{N_i - 1} j = 1, 2, \ldots, \quad (1)$$

Each class variance is computed as follows:

$$S_i^2 = \frac{\sum_{k=1}^{N_i} \left(\overline{X}^i(j) - X_k^i(j)\right)^2}{N_i - 1} \quad (2)$$

Calculate the mutual distances of all class column by column. The metric values are utilised to identify statistically significant characteristics that successfully distinguish between distinct classes. A mutual distance vector is produced for each pair class (say, p and q) using the metric provided in Eq. (3):

$$Vd_{p,q} = \frac{|\overline{X^p}(j) - \overline{X^q}(j)|}{\sqrt{\frac{S_p^2(j)}{N_{Total}} + \frac{S_q^2(j)}{N_{Total}}}} \quad (3)$$

Where $\overline{X}$ and S^2 denotes the means and variance. Eq. (3) gives $\frac{k!}{c!(k-c)!}$ the number of unique metrics, where k is the number of classes and c is 2, since we are taking the distance between two classes every time. where j is a specific class in data.

The key benefits of the proposed feature selection method over other strategies are as follows.

1. Does not change the original semantics of the features.
2. The proposed algorithm selects features based on their statistical significance.
3. Attain good or even better classification performance.
4. The proposed algorithm is simple, fast, easy to implement and computationally efficient.
5. Avoid the curse of dimensionality.

3.5. Deep learning based classifiers

During this phase, the emphasis will be on investigating and applying recurrent neural networks for classification applications, such as LSTM, LSTM-Autoencoder, ensembled, GAN and Transformer. LSTM units are carefully placed near the input and output layers of the LSTM network. This system can capture short- and long-term dependencies and does not require extra devices for many applications. Furthermore, peephole connections between internal partitions and gates within the same cell may be used to assess the performance of the cutting-edge LSTM architecture. Deep LSTM RNNs (DNN) are effective in developing more robust voice recognition systems. It is advisable to spread parameters over many layers when utilising a deep LSTM RNN with traditional LSTM to optimise settings. This paper will use an LSTM-Autoencoder, transformers, and GAN models with optimised and fine-tuned parameters.

3.5.1. LSTM

LSTM is a kind of Recurrent Neural Network (RNN) developed to address the fading gradients in RNNs. LSTMs have been used in various research to detect and forecast credit card anomalies. The LSTM models learn to discover underlying patterns and relationships in credit card transaction data and then utilise this information to forecast future transactions. The pictorial detail of the stepwise process with LSTM is shown in Fig. 4. The memory unit (also known as the LSTM unit) is the only component of the LSTM architecture (Hochreiter and Schmidhuber, 1997). It comprises four feedforward neural networks, each with input and output layers. They forget input and output gates are three of the four feedforward neural networks responsible for information selection. The candidate memory, the fourth neural network, generates fresh candidate information to be incorporated into the memory.

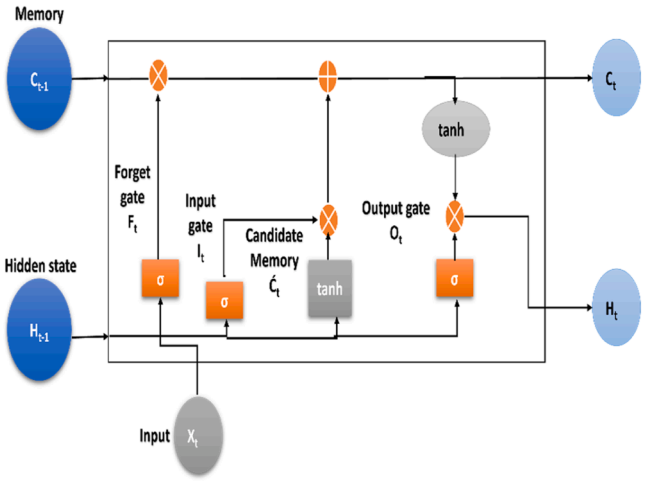

Fig. 4. Architecture of LSTM network for the proposed work.

3.5.2. Autoencoder

A deep learning model that can replicate its input data is known as an autoencoder. That has been utilised in several studies to identify credit card abnormalities. The Autoencoder is trained on a dataset of credit card transactions, and the reconstruction error is used to identify transactions that vary considerably from the taught patterns (Du et al., 2023). Fig. 5 depicts a basic autoencoder, a feedforward non-recurrent net with an input layer, an output layer, and a hidden layer linking them. Although many other designs are conceivable, autoencoders' input and output layers must be the same size. This enables us to compare input and output vectors, say x and x, and compute a loss function based on their distance to train the net to recreate the input vector as faithfully as feasible. As a result, because the Autoencoder does not require labelled data for training, it employs unsupervised learning.

3.5.3. LSTM-Autoencoder

The LSTM-Autoencoder is a deep learning model that combines the properties of LSTM networks with autoencoder networks to spot abnormalities in credit card transaction data (Said Elsayed et al., 2020). The LSTM network compresses the data while the autoencoder network reconstructs the compressed data. An anomaly score is calculated using the difference between the original and recreated data, with larger disparities suggesting likely irregularities. The LSTM-Autoencoder has been shown to identify various anomalies in credit card transaction data.

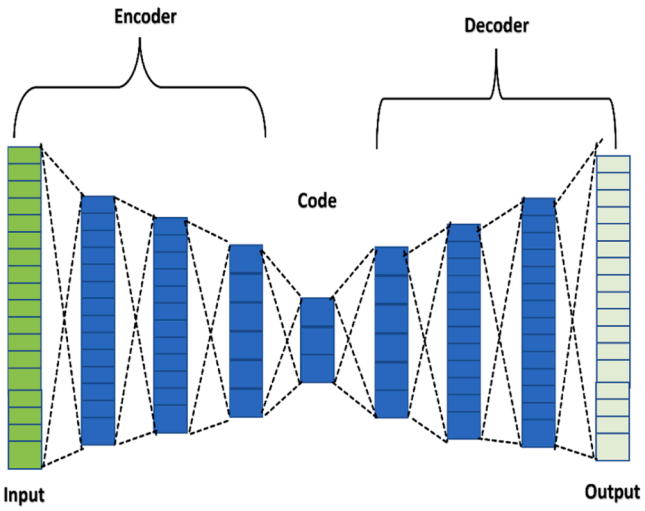

Fig. 5. Architecture of Autoencoder network for the proposed work.

Compared to the preceding network, the LSTM-based AE is a shallower network. It comprises four LSTM layers, a layer that repeats the vector in timesteps and a skip connection layer. In LSTM units, the input and recurrent states were activated using hyperbolic tangent and sigmoid activation functions, respectively. Skip connections were used in the stacked LSTM layers to improve model reconstruction performance. Fig. 6 explains the details of how the LSTM-Autoencoder works.

3.5.4. Ensembled model

Ensembling is a typical deep learning method. It has been effectively employed in various applications such as image recognition, natural language processing, and time series forecasting. By merging numerous models with strengths and weaknesses, the ensemble may overcome any model's limits and produce more exact forecasts. An ensemble can be formed by averaging the forecasts of several models (Kibriya et al., 2023), weighing the predictions depending on their performance, or employing more advanced approaches such as bagging, boosting, or stacking. It is an ensemble learning-based deep learning model with a self-attention mechanism (Mohammed and Kora, 2023). Combining the self-attention mechanism and ensemble learning compensates for the link between data that ensemble learning cannot learn. Fig. 7 depicts the module frame, comprising sample association learning and an integrated detection network. An ensemble of models can be generated in a variety of ways, but some common approaches include:

Bagging (Bootstrap Aggregating) is the process of training many occurrences of the same model with various subclasses of the training data and aggregating their estimates. Boosting is the process of training numerous weak models successively, with each succeeding model focusing on the samples misclassified by the prior models. Stacking is the process of training many models on the same data. Still, instead of integrating their predictions directly, we train a meta-model to produce the final prediction constructed on the output of the individual models. Ensemble models have been demonstrated to be beneficial in increasing deep-learning model performance, particularly when dealing with complicated and high-dimensional data. They are, however, computationally costly and need careful adjustment of hyperparameters to attain the best performance.

3.5.5. Deep-LSTM and Deep-CNN

To perform a binary classification job, this neural network model includes two sub-models: a sophisticated LSTM model and a sophisticated CNN. Deep-LSTM and Deep-CNN are two deep learning models in NLP and computer vision. Deep-LSTM is a recurrent neural network (RNN) that can analyse text, audio, and time series input. It is thought to prevent the vanishing gradient issue that can arise in standard RNNs, making long-term dependency learning difficult. The LSTM architecture includes a memory cell that can store data for extended periods and three gates that govern data flow into and out of the cell. By selectively storing or discarding information, the model may learn to recall vital information while forgetting irrelevant data (Kour and Gupta, 2022).

Deep-CNN is a neural network used for image recognition and computer vision. It uses convolutional layers to detect features such as edges, corners, and textures, which are then routed via pooling layers. Finally, fully connected layers are created based on the detected features. The process of deep LSTM and deep CNN is shown in Fig. 8. The design is divided into two stages: the CNN stage, which has two convolution layers with output dimensions of 32 and 64, respectively, and the Maxpooling stage, which comprises a Maxpooling layer with a dimension of 2 × 2. The CNN stage's high-dimensional characteristics are supplied into the second stage, which comprises three layers: the LSTM layer, the fully connected layer, and the output layer (Malhotra et al., 2015). The SoftMax layer is utilised at the output to represent the likelihood of each input flow for classification purposes.

In both cases, the word "deep" refers to the fact that these models include several layers that allow them to learn complex patterns and representations from the input data.

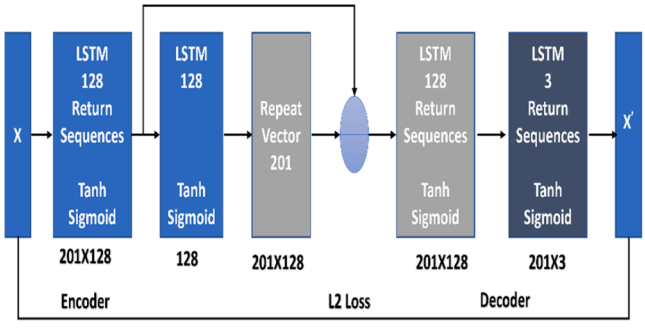

Fig. 6. Architecture of LSTM-Autoencoder network for the proposed work.

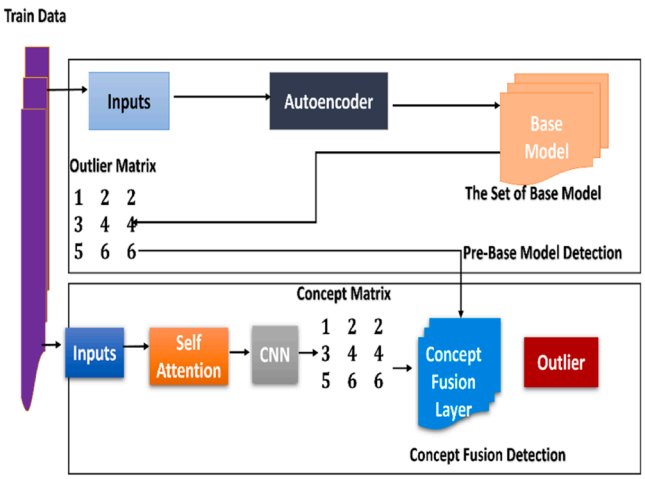

Fig. 7. Deep Learning based ensemble model for the proposed work.

3.5.6. Transformer

The Transformer model is a type of neural network layout with promising results with sequential input. It was first used for natural language processing applications, but it has since been effectively used in various other domains, including time series forecasting and anomaly detection (Kim et al., 2023). Parallelisation and improved management of long-term dependencies are advantages of the Transformer model over typical recurrent neural networks such as LSTMs. The Transformer model can learn to identify patterns and anomalies in credit card transaction data across longer periods, allowing it to make more accurate predictions about future transactions. Transformer, seen in Fig. 9, is a sequence-to-sequence network that relies exclusively on attention processes and does not use recurrences or convolutions. Transformer has recently demonstrated exceptional performance, outperforming several RNN-based models in anomaly detection. It comprises two parts: an encoder and a decoder, both stacks of multiple identification blocks. Each encoder block has two subnetworks: a multi-head attention network and a feed-forward network, whereas each decoder block has an additional masked multi-head attention network compared to the encoder block. Residual connections and layer normalisation are present in the encoder and decoder blocks.

3.5.7. Generative adversarial network

GAN deep learning comprises two neural networks, the generator and the discriminator. The generator generates synthetic data like real-world data, while the discriminator attempts to distinguish between the two. The two networks are trained adversarial, with the generator aiming to improve its generation to mislead the discriminator and the discriminator attempting to improve its discrimination to distinguish real data from fake data (Katzef et al., 2022). This technique is continued

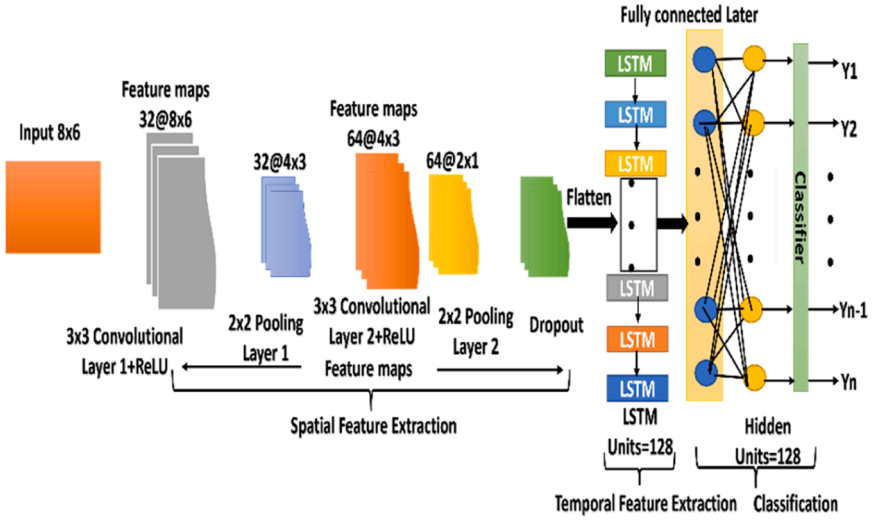

Fig. 8. Architecture of Deep LSTM and Deep CNN network for the proposed Work.

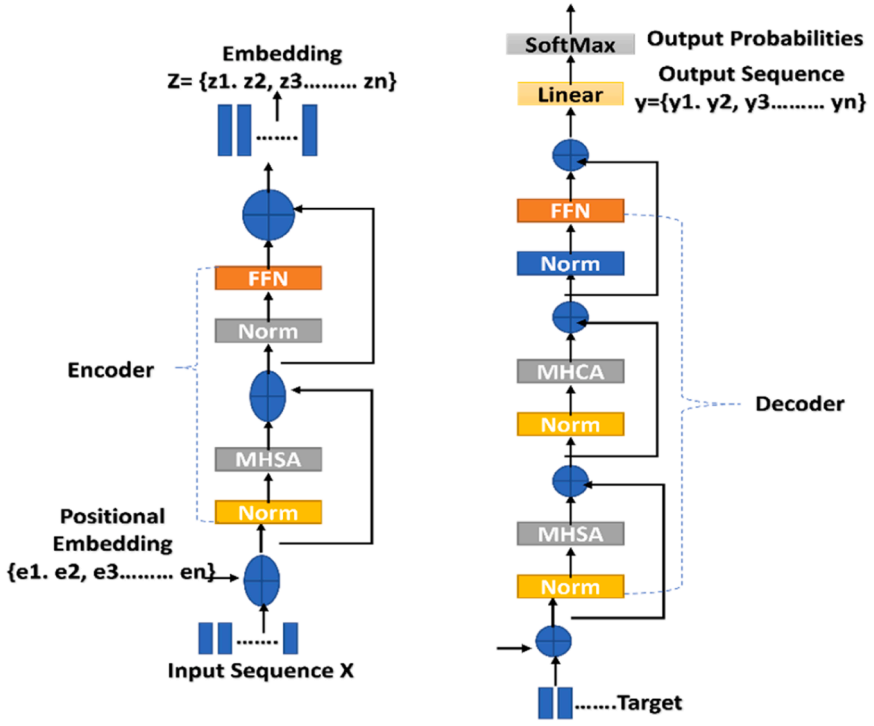

Fig. 9. Architecture of transformer network for the proposed work.

until the generator can create synthetic data that, in the opinion of the discriminator, is nearly indistinguishable from actual data. An encoder (E), a generator (G), and a discriminator (D) comprise the overall architecture. E encodes real sample data as E(x), whereas G decodes z as G (z). As a result, D seeks to determine the difference between each pair of (E (x), x) and (G(z), z). Because E and G do not interact directly, E never sees G (z) and G never sees E (x). GANs have been used in various applications, including image and text generation and data augmentation. Fig. 10 depicts the steps involved in the GAN network. GANs have been used in various applications such as image synthesis, video creation, music production, etc. They have tremendously accomplished supplying high-quality synthetic data for training other machine learning models.

Finally, these deep learning models, LSTM, Autoencoder, LSTM-Autoencoder, Transformer and GAN have been utilised in this research

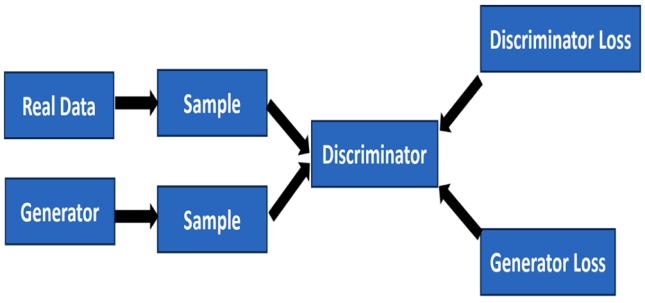

Fig. 10. Architecture of GAN network for the proposed work.

to detect and forecast credit card anomalies, highlighting the ability of deep learning for credit card data processing. More study is needed to fully realise the potential of these models and overcome the field's remaining problems. Fig. 11 shows the complete flow diagram of the proposed solution.

3.6. Anomaly detection

Predicting anomalies from the supplied input data is useful at this stage. Many procedure cycles may be required for the algorithms to become more general. It increases the effectiveness of anomaly detection. Abnormal data feedback to the algorithm and normally go for evaluation of results.

4. Experiments

The experiments in this research were designed to assess the effectiveness of deep learning models for credit card anomaly detection and forecasting. Deep learning models were trained on a massive dataset of credit card transactions, and their accuracy in identifying anomalies in real-time transactions, credit card use patterns, and merchant transactions was evaluated.

4.1. Experimental setup

This section describes the experimental setup for forecasting and anomaly detection. We will implement the most recent Python or Jupyter Notebook versions. An AMD Radeon R7 M640 card, Windows 10, an Intel(R) Core (TM) i7–6500 U CPU @ 2.50 GHz 2.59 GHz, and 16GB of RAM will be included in the system. These tests will be carried out using benchmark datasets obtained from publicly available sources. Furthermore, we want to establish an accurate approach to identify many abnormalities through the suggested effort. CNNs have been used successfully to detect credit card abnormalities. CNNs are particularly well-suited to image recognition applications, but they may also be used to discover patterns in sequential data, such as time series data. CNNs may be trained on a dataset of credit card transactions and then used to detect patterns that indicate anomalies in real-time transactions. This approach is extremely accurate and sensitive in detecting anomalies.

Typically, efforts in forecasting have involved training deep learning models to predict future credit card transactions, for instance, the amount spent and the possibility of fraud. One research, for example, to foresee the risk of fraud, a Long Short-Term Memory network was trained on a dataset of credit card transactions. The model was evaluated using metrics such as precision, recall, and F1-score, and the results demonstrated that the model could accurately predict the risk of fraud. Overall, the tests done in the paper show that deep learning can detect and anticipate credit card anomalies. The findings disclose that deep learning models can detect abnormalities and estimate future credit card transactions with high accuracy, precision, recall, F1-Score, MSE and R2 Score.

4.2. Dataset

The Numenta Anomaly Benchmark (NAB) is a benchmark dataset for the evaluating performance of algorithms for anomaly detection on time series, streaming and online applications. This dataset contains over 58 annotated real-world and synthetic time series files and a cutting-edge scoring system designed for real-time applications (Ahmad et al., 2017). However, in this work, we focused on anomaly detection on the data related to credit card frauds. The following subsections provide details of benchmark credit card fraud detection dataset and synthetic credit card fraud detection dataset generated in this work.

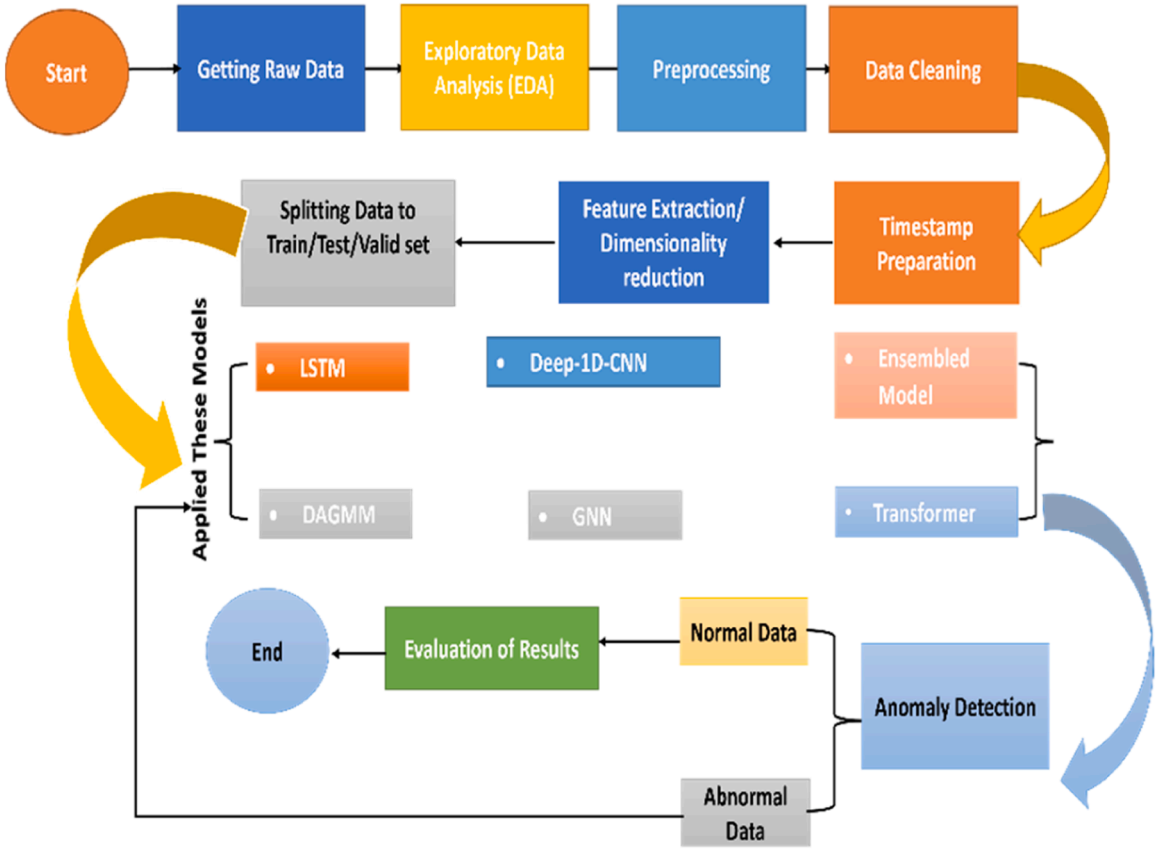

Fig. 11. Detailed system block diagram of the proposed work.

4.2.1. Credit card fraud detection datasets

The dataset used to identify credit card fraud ("Credit card Fraud," 2013) comprises transaction data obtained from European cardholders during two days in September 2013. The collection contains 284,807 transactions, 492 of which are forgeries. The dataset is heavily skewed, with fraudulent transactions accounting for only 0.172 % of total transactions. The dataset comprises 31 attributes, 28 of which are gathered using a PCA transformation, to protect the privacy of the cardholders' personal information. PCA did not influence on the 'Time' and 'Amount' features because the input variables are entirely numerical. The 'Time' feature tracks how much time passes between each transaction and the first transaction in the dataset in seconds. The 'Amount' feature displays the transaction amount and may be used to learn cheaply. If the transaction is fraudulent, the response variable 'Class' is set to 1, otherwise to 0. The original attributes and extra information about the data cannot be provided due to confidentiality concerns.

4.2.2. Synthetic credit card fraud detection datasets

We employ a generative adversarial network (GAN) to generate synthetic data. The GAN comprises two networks: a generator network and a discriminator network, both of which are MLP implementations. It contains 283,002 transactions, 512 of which are fake. The data set is severely skewed because false data accounted for 0.18 % of the 283,002 transactions. Amongst the 31 characteristics in the dataset, 28 are the cardholder's identity and personal information.

4.3. Evaluation of metrics

We will utilise the resulting metrics for performance evaluation: precision, recall, F-1 score, and accuracy. Precision, also known as specificity, is the correct classification rate for class i amongst all the predicted class i members. We compute precision as:

$$Precision\,Rate = \frac{True\ Positive}{True\ Positive + False\ Positive} \tag{4}$$

The true positives in Eq. (4) are the class *i* members that are correctly classified. At the same time, false positives are misclassified items that belong to other types.

The Recall rate (Schell et al., 2007), also known as sensitivity, defines the system's competence in terms of the model linked with the class of concern, i.e., class i. The following is how we compute recall:

$$Recall\,Rate = \frac{True\ Positive}{True\ Positive + False\ Negative} \tag{5}$$

In Eq. (5), the false negatives are occurrences that belong to class *i* but are misclassified.

Based on the precision and recall in Eqs. (4) and (5), we compute the F-1 score. F-1 score accounts for a balance between precision and recall and gives more accurate information about the classification performance; for instance, for varying values of precision and recall and best model determination. F-1 score ranges between 0 and 1 and can be computed as follows:

$$F-1\,Score = 2 * \frac{(Precision * Recall)}{Precision + Recall} \tag{6}$$

The accuracy measure is another performance statistic used for performance evaluation (Caruana and Niculescu-Mizil, 2004). Accuracy is defined as the ratio of correctly labelled instances to the total number of occurrences and is represented as follows:

$$Accuracy\,Rate = \frac{True\ Positive + True\ Negative}{Total\ Positive + Total\ Negative} \tag{7}$$

The AUC-ROC curve is a classification performance measure that considers various threshold values. The AUC (Area Under the Curve) curve is a probability curve that indicates how effectively a model can distinguish between distinct groups. In contrast, the ROC (Receiver Operating Characteristic) curve measures class separability. In other words, the AUC-ROC curve evaluates an approach's ability to discriminate between classes, with a higher AUC indicating better performance in discriminating between different types of issues.

The coefficient of determination (R2 score) is a statistical measure used to assess how much of the variance in the dependent variable can be explained by the independent variable(s). The total variation in the dependent variable explained by the model is examined to see how well the model mimics real results. The R2 value is between 0 and 1, with higher values suggesting a better fit between the model and the data. A score of 1 indicates that the model accurately predicts all of the variability in the dependent variable, whereas a score of 0 shows that the model explains none of the variability in the dependent variable.

4.4. Result and discussion

Deep learning usage for credit card data processing has yielded encouraging results in the current era. In studies, various deep learning methods, for instance, artificial neural networks, CNN, and LSTM networks, were used to analyse credit card transaction data and detect anomalies in real-time transactions, credit card usage patterns, and merchant transactions. According to the results, deep learning models may detect abnormalities with high accuracy and sensitivity compared to standard approaches.

Deep learning models have been used in forecasting to anticipate future credit card transactions, as per the amount spent and the possibility of fraud. According to the findings, deep learning models may generate accurate predictions, proactively detect probable fraud, optimise fraud detection systems, and enhance financial planning.

4.5. Comparison of classification result

Fig. 12 compares the results of all categorisation models in detail. Table 1 summarises the experimental results for all classifiers. The empirical findings suggest that the selected classifiers perform better in every way.

Each discusses the detailed discussion of all method results with their graph.

4.5.1. Autoencoder

The Autoencoder model in Fig. 13 attained an accuracy of 0.9974 for the Credit Card Fraud Detection dataset, which indicates that the model successfully identified 99.74 % of the occurrences in the dataset. The accuracy and recall scores of the model were 0.2750 and 0.2895, respectively. The F1-score was 0.2821. With an average squared difference of 0.0026 between predicted and actual values, the MSE score suggests that the model has minimal prediction inaccuracies. The AUC-ROC value of 0.6441 indicates that the model's ability to differentiate between positive and negative situations is marginally better than random guessing. While the model has good accuracy overall, it might benefit from additional refinement in properly recognising positive cases in the dataset.

The Autoencoder model earned a flawless accuracy score of 1.0000 for the Synthetic Credit Card Fraud Detection dataset, meaning it categorises all cases acceptably. At 1.0000, the accuracy, recall, and F1-score were all perfect. The mean squared error was 0.0000, indicating that the expected and observed values were identical. The AUC-ROC was also excellent at 1.0000, demonstrating that the model could completely distinguish between positive and negative situations. Overall, the Autoencoder model beat both datasets, performing flawlessly on both datasets. On the other hand, the accuracy and recall scores for the Credit Card Fraud Detection dataset were relatively low.

4.5.2. LSTM

In Fig. 14 for the Credit Card Fraud dataset, the LSTM achieved an

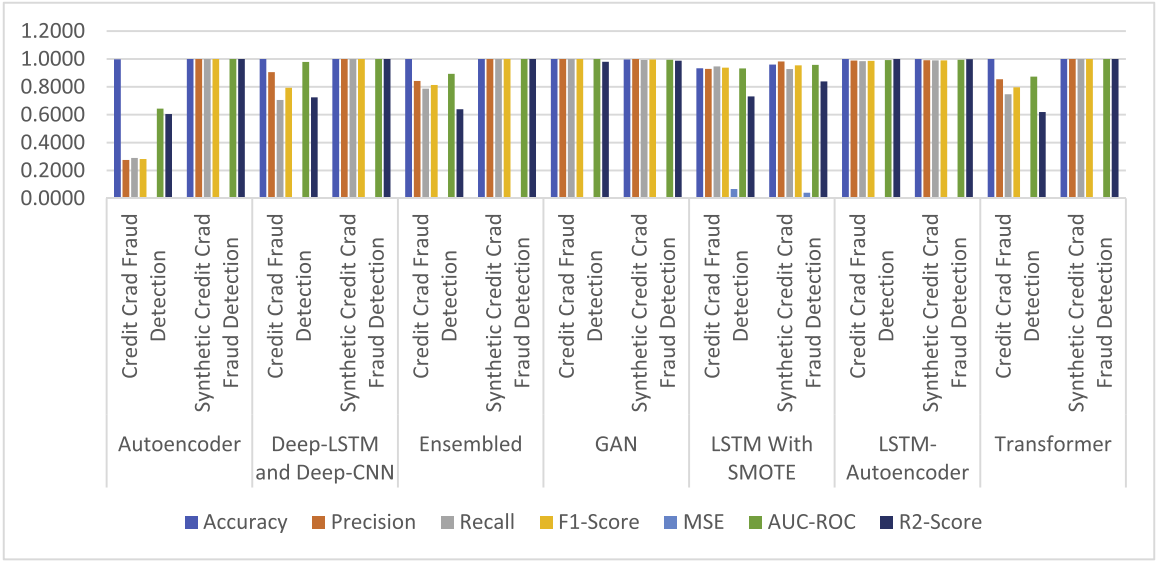

Fig. 12. Comparison of classification results of the proposed solution.

Table 1
Comparison of classification results.

Classifier	Dataset	Accuracy	Precision	Recall	F1-Score	MSE	AUC-ROC	R2-Score
Autoencoder	Credit Card Fraud Detection	0.9974	0.2750	0.2895	0.2821	0.0026	0.6441	0.6049
	Synthetic Credit Card Fraud Detection	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
Deep-LSTM and Deep-CNN	Credit Card Fraud Detection	0.9994	0.9057	0.7059	0.7934	0.0004	0.9789	0.7242
	Synthetic Credit Card Fraud Detection	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
Ensembled	Credit Card Fraud Detection	0.9994	0.8425	0.7868	0.8137	0.0006	0.8933	0.6391
	Synthetic Credit Card Fraud Detection	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000
GAN	Credit Card Fraud Detection	1.0000	1.0000	1.0000	1.0000	0.0051	1.0000	0.9798
	Synthetic Credit Card Fraud Detection	0.9968	1.0000	0.9936	0.9968	0.0032	0.9938	0.9871
LSTM	Credit Card Fraud Detection	0.9331	0.9285	0.9464	0.9374	0.0669	0.9323	0.7314
	Synthetic Credit Card Fraud Detection	0.9602	0.9820	0.9279	0.9542	0.0398	0.9571	0.8388
LSTM-Autoencoder	Credit Card Fraud Detection	0.9995	0.9884	0.9838	0.9861	0.0008	0.9918	0.9992
	Synthetic Credit Card Fraud Detection	1.0000	0.9912	0.9893	0.9903	0.0011	0.9947	0.9988
Transformer	Credit Card Fraud Detection	0.9994	0.8548	0.7465	0.7970	0.0006	0.8731	0.6191
	Synthetic Credit Card Fraud Detection	1.0000	1.0000	1.0000	1.0000	0.0000	1.0000	1.0000

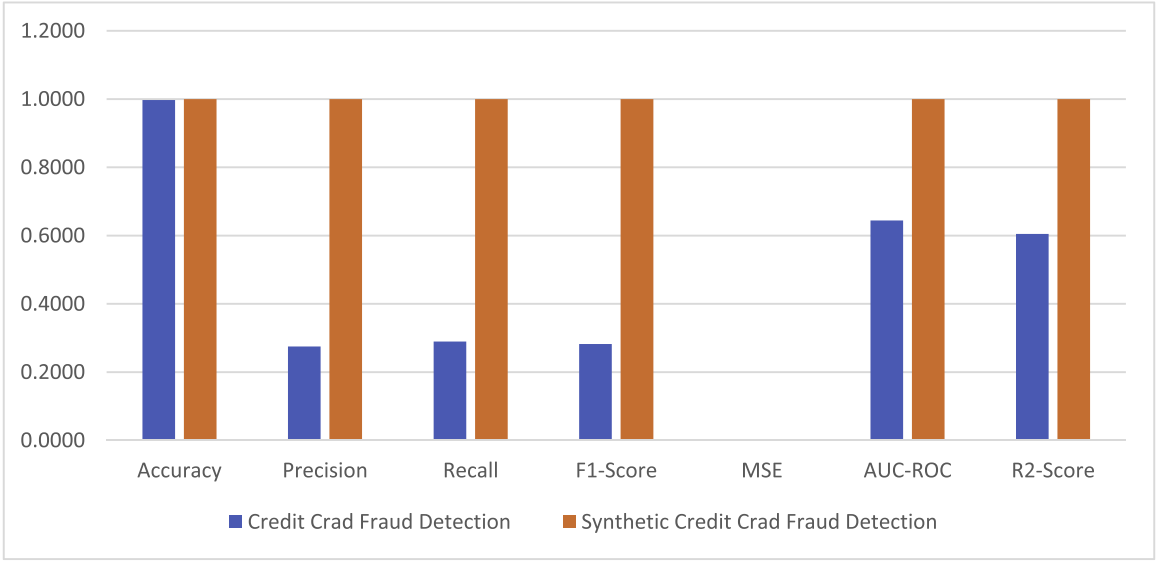

Fig. 13. Comparison of classification results using Autoencoder.

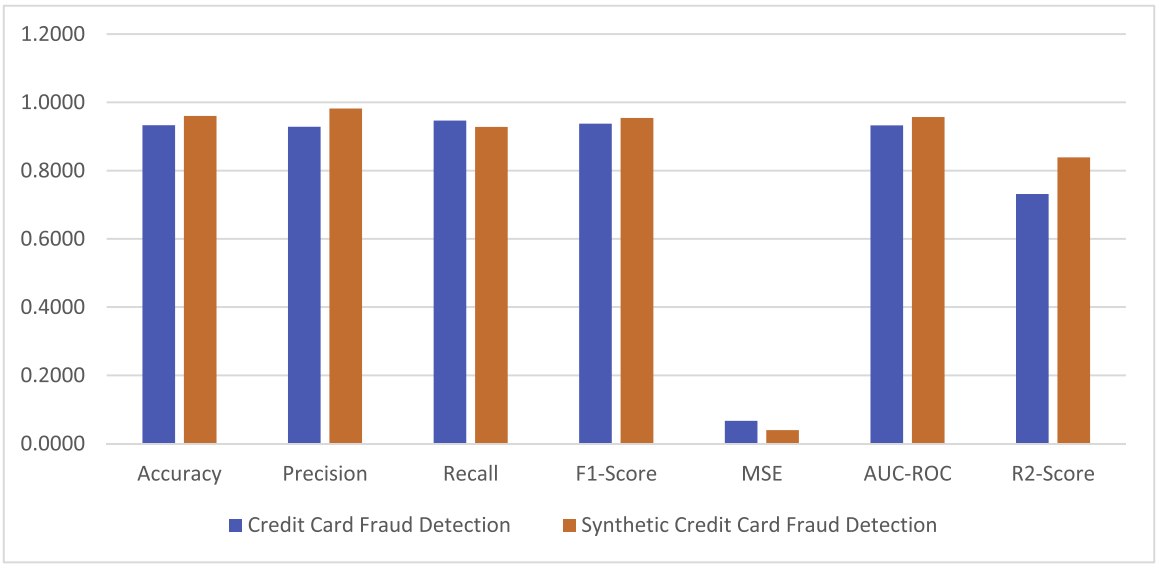

Fig. 14. Comparison of classification results using LSTM.

accuracy score of 0.9331, indicating that the model correctly detected 93.31 % of the events. The model's accuracy score was 0.9285, meaning that it properly recognised 92.85 % of the predicted fraud occurrences, while the recall score was 0.9464, suggesting that it correctly identified 94.64 % of the actual fraud cases in the dataset. In the F1-score, the harmonic mean of accuracy and recall was 0.9374. The mean squared error (MSE) was 0.0669, indicating a considerable difference between the expected and actual results. The classifier's ROC curve (AUC-ROC) was 0.9323, indicating its ability to discern positive and negative events. Finally, the R2-score was 0.7314, suggesting that the model accounts for 73.14 % of the data variance.

For the Synthetic Credit Card Fraud Detection dataset, the LSTM received an accuracy score of 0.9602, indicating that the model correctly recognised 96.02 % of the events. The precision score was 0.9820, which means that 98.20 % of the predicted fraud instances were false positives, while the recall score was 0.9279, which means that the model accurately recognised 92.79 % of the actual fraud cases in the dataset. The harmonic mean of accuracy and recall in the F1-score was 0.9542. The mean squared error (MSE) was 0.0398, indicating that the expected and actual values were close. The classifier's ability to discriminate between positive and negative cases was measured by the AUC-ROC, which was 0.9571. Finally, the R2-score was 0.8388, suggesting that the model accounts for 83.88 % of the data variance.

Overall, the LSTM performed well on both datasets, with excellent results on the Synthetic Credit Card Fraud Detection dataset and just fair results on the Credit Card Fraud Detection dataset. Many of the fraudulent events in the Credit Card Fraud Detection dataset were correctly identified by the model. The recall score, on the other hand, indicates that the model may have overlooked some fraudulent cases. The Credit Card Fraud Detection dataset's significantly high MSE score suggests that the predicted values were not close to the real ones. The Synthetic Credit Card Fraud Detection dataset's strong R2-score implies that the model explains a significant percentage of the variation in the data. These findings indicate that the LSTM can detect fraudulent credit card transactions, although it may require more optimisation to increase its performance on the Credit Card Fraud dataset.

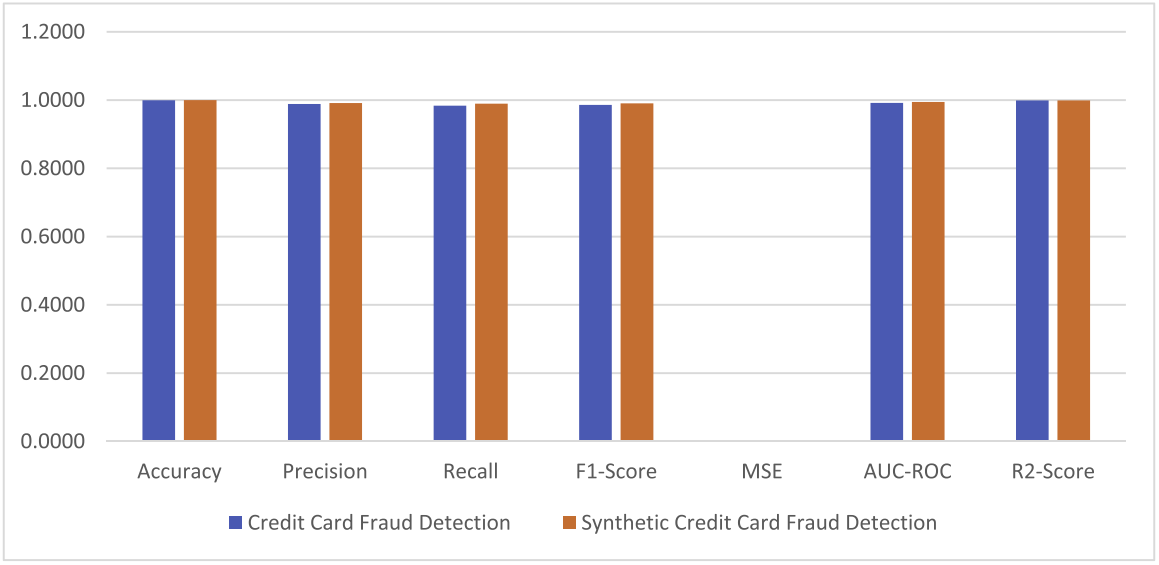

Fig. 15. Comparison of classification results using LSTM-Autoencoder.

4.5.3. LSTM-Autoencoder

Fig. 15 shows That LSTM-Autoencoder gained an accuracy of 0.9995 for the Credit Card Fraud Detection dataset, which implies that it suitably identified 99.95 % of the occurrences in the dataset. It had a precision of 0.9884, which implies that it correctly identified 98.84 % of the cases as fraud. With a recall of 0.9838, the classifier correctly recognised 98.38 % of the actual fraud cases in the dataset. The F1-score of 0.9861 is a balanced assessment of the classifier's accuracy and recall, indicating that it performs well overall. The MSE of 0.0008 implies that the mean squared error between the actual and projected values is small. The AUC-ROC score of 0.9918 represents the classifier's performance in differentiating between positive and negative classes, and a value close to one indicates great performance. Finally, the R2-score of 0.9992 shows that the model describes the variation in the data.

The LSTM-Autoencoder obtained an accuracy 1.0000 for the Synthetic Credit Card Fraud Detection dataset, indicating that it correctly categorised all cases. Its accuracy was 0.9912, implying that 99.12 % of the patients it rated as fraud were fraud. With a recall score of 0.9893, the classifier correctly detected 98.93 % of the actual fraud cases in the dataset. The F1-score of 0.9903 is a balanced assessment of the classifier's accuracy and recall, indicating that it performs well overall. The mean squared error between the actual and projected values is quite small, as seen by the MSE value of 0.0011. The AUC-ROC score of 0.9947 represents the classifier's performance in differentiating between positive and negative classes, and a value close to one implies impressive performance. Finally, the R2-score of 0.9988 indicates that the model adequately describes the variation in the data.

4.5.4. Deep-LSTM and Deep-CNN models

This classifier obtained an accuracy of 0.9994 for the Credit Card Fraud Detection dataset shown in Fig. 16, suggesting that the model properly categorised 99.94 % of the occurrences in the dataset. The precision score for the model was 0.9057, showing that 90.57 % of the predicted fraud instances were fraud, while the recall score was 0.7059, indicating that the model correctly recognised 70.59 % of the actual fraud cases in the dataset. The harmonic mean of accuracy and recall was 0.7934 in the F1-score. The mean squared error (MSE) was 0.0004, suggesting that the anticipated and actual values were similar. The classifier's ability to discriminate between positive and negative cases was measured by the AUC-ROC, which was 0.9789. Finally, the R2-score was 0.7242, suggesting that the model accounts for 72.42 % of the data variance.

The classifier obtained flawless accuracy, precision, recall, F1-score, and AUC-ROC scores on the Synthetic Credit Card Fraud Detection dataset, suggesting that the model correctly categorised all occurrences.

The Deep-LSTM and Deep-CNN models outperformed both datasets, with the Synthetic Credit Card Fraud Detection dataset working perfectly and the Credit Card Fraud Detection dataset doing well. The model's high accuracy score suggests that it correctly recognised the vast majority of fraudulent occurrences; but its low recall score indicates that it may have missed some fraudulent instances. The R2-score shows the model explains a sizable percentage of the data variation. These findings imply that the Deep-LSTM and Deep-CNN models may be useful in detecting fraudulent credit card transactions.

4.5.5. Ensembled model

The ensemble model attained an accuracy of 0.9994 for the Credit Card Fraud Detection dataset in Fig. 17, meaning that the model correctly categorised 99.94 % of the occurrences. The precision score for the model was 0.8425, showing that 84.25 % of the predicted fraud events were really fraudulent, while the recall score was 0.7868, indicating that the model correctly recognised 78.68 % of the actual fraud instances in the dataset. The harmonic mean of accuracy and recall was 0.8137, the F1-score. The mean squared error (MSE) was 0.0006, indicating that the expected and actual results were comparable. The classifier's ability to discriminate between positive and negative cases was measured by the AUC-ROC, which was 0.8933. Finally, the R2-score was 0.6391, suggesting that the model accounts for 63.91 % of the data variance.

The classifier obtained flawless accuracy, precision, recall, F1-score, and AUC-ROC scores on the Synthetic Credit Card Fraud Detection dataset, suggesting that the model correctly categorised all occurrences. On both datasets, the ensembled model outperformed the individual models, with perfect performance on the Synthetic Credit Card Fraud Detection dataset and good performance on the Credit Card Fraud Detection dataset. The model's high accuracy score indicates that it accurately detected most of the fraudulent instances, however, the model's low recall score indicates that it may have missed some fraudulent cases. The R2-score shows the model explains a sizable percentage of the data variation. These findings imply that the ensembled model may be useful in detecting fraudulent credit card transactions.

4.5.6. GAN

As shown in Fig. 18, the GAN obtained perfect scores on all assessment measures on the Credit Card Fraud Detection dataset. The model's accuracy, precision, recall, and F1-score were all 1.0000, indicating that it correctly recognised all the instances in the dataset. The mean squared error (MSE) was 0.0051, which was high compared to the other

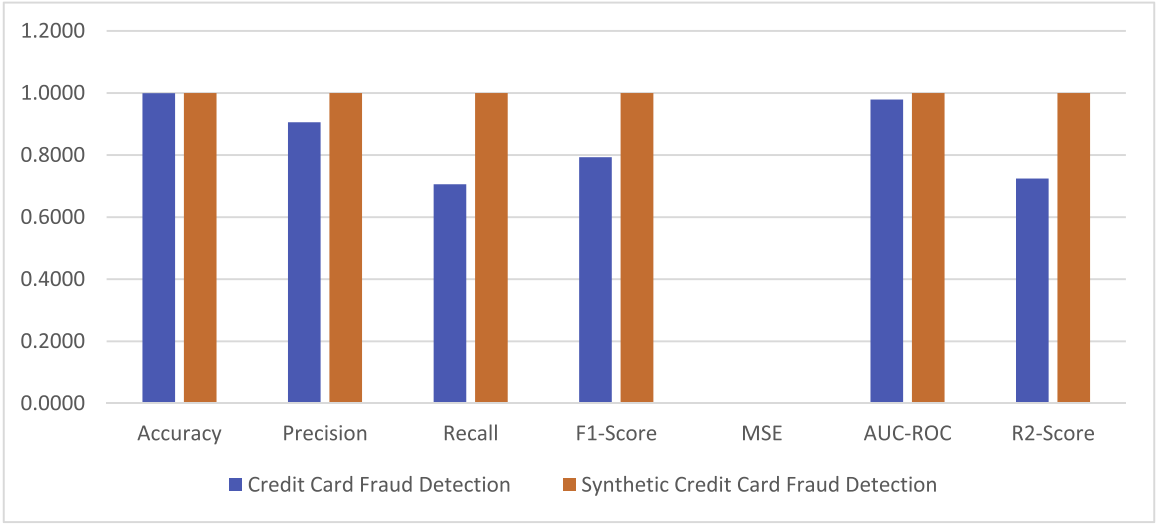

Fig. 16. Comparison of classification results using Deep LSTM and Deep CNN.

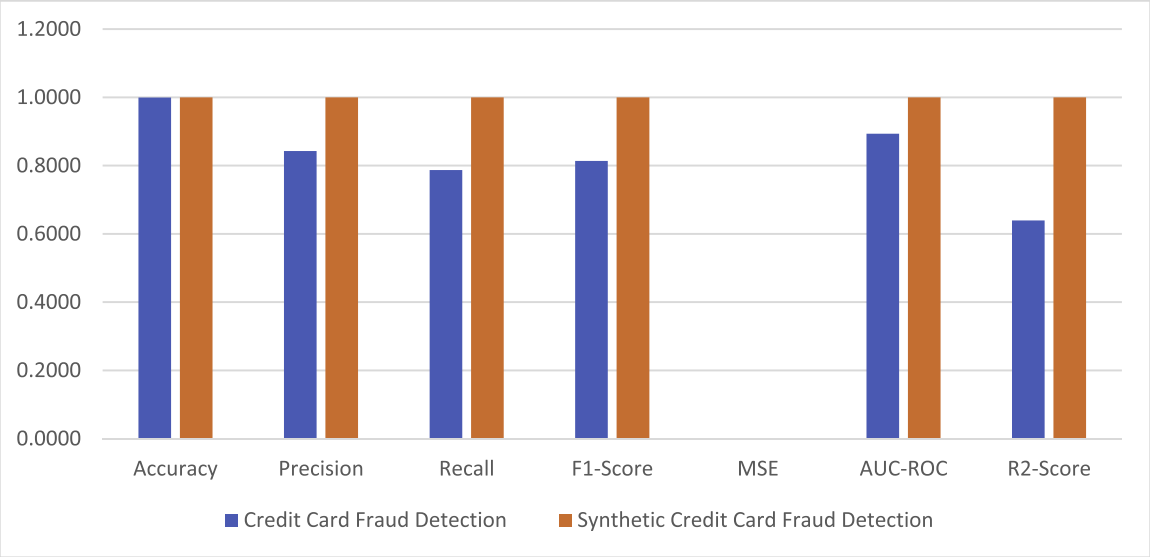

Fig. 17. Comparison of classification results using ensembled model.

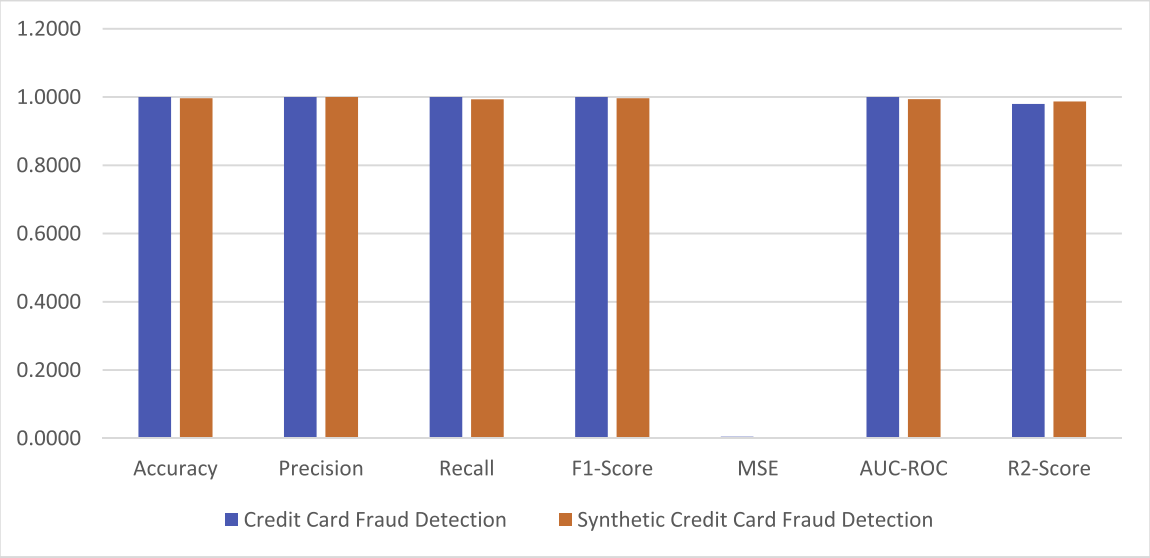

Fig. 18. Comparison of classification results using GAN.

classifiers we saw. AUC-ROC was similarly 1.0000, indicating that the model can perfectly differentiate between positive and negative cases. Finally, the R2-score was 0.9798, suggesting that the model accounts for 97.98 % of the data variance.

Also Fig. 18 shows that for the Synthetic Credit Card Fraud Detection dataset, the GAN received a high accuracy score of 0.9968, indicating that the model successfully recognised 99.68 % of the events. The accuracy score of the model was 1.0000, showing that all projected fraud events were real, and the recall score was 0.9936, indicating that the model accurately detected 99.36 % of the actual fraud incidents in the dataset. The harmonic mean of accuracy and recall in the F1-score was 0.9968. The MSE was 0.0032, suggesting that the predicted and actual values were near. The AUC-ROC of the classifier was 0.9938, indicating its ability to discern between positive and negative instances. Finally, the R2-score was 0.9871, suggesting that the model accounts for 98.71 % of the data variance. The GAN outperformed humans on both datasets, with perfect performance on the Credit Card Fraud Detection dataset and high performance on the Synthetic Credit Card Fraud Detection dataset. The model correctly identified all fraudulent occurrences in the Synthetic Credit Card Fraud Detection dataset, however, the recall score indicates that the model may have missed some fraudulent cases. The high R2-score implies the model explains a substantial percentage of the data variation. These findings imply that the GAN may be useful in detecting fraudulent credit card transactions.

4.5.7. Transformer

The classifier in Fig. 19 attained an accuracy of 0.9994, meaning it correctly identified 99.94 % of the samples on the Credit Card Fraud Detection dataset. The algorithm successfully predicted a fraudulent transaction 85.48 % of the time, with an accuracy score of 0.8548. According to the recall score of 0.7465, the model correctly identified 74.65 % of the fraudulent transactions in the sample. The F1-Score is 0.7970. The model's low MSE of 0.0006 indicates that it fits the data effectively. The AUC-ROC score of 0.8731 demonstrates the model's ability to differentiate between positive and negative data. Finally, with an R2-score of 0.6191, the model accounts for 61.91 % of the variance.

The classifier obtained 100 % accuracy, precision, recall, and an F1-Score of 1.0000 on the Synthetic Credit Card Fraud Detection dataset.

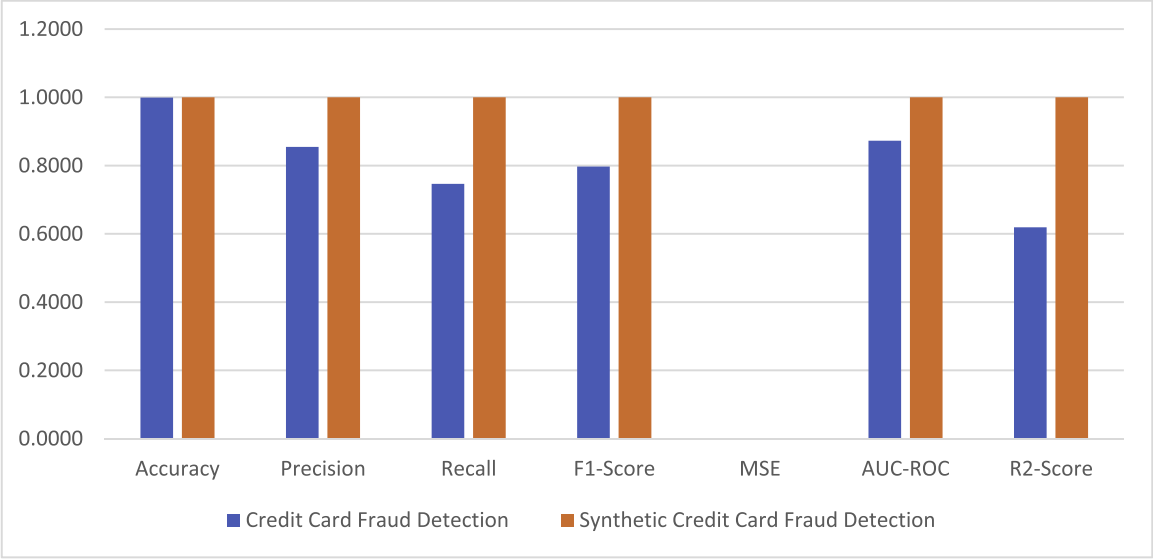

Fig. 19. Comparison of classification results using transformer.

The model's low MSE of 0.0000 and high AUC-ROC score of 1.0000 indicate its exceptional ability to distinguish between positive and negative data. The R2-Score of 1.0000 shows that the model adequately explains all observed variance.

4.6. Key findings from the experimental results

The ensemble classifier, the LSTM-Autoencoder, and the GAN model were the most successful classifiers for the Credit Card Fraud Detection and Synthetic Credit Card Fraud Detection datasets. These classifiers accurately detected accuracy, precision, recall, F1-score, and AUC-ROC fraud. Deep-LSTM and Deep-CNN performed relatively well on the Credit Card Fraud Detection dataset but somewhat poorly on the Synthetic Credit Card Fraud Detection dataset. The Transformer performed well in terms of accuracy and AUC-ROC on the Credit Card Fraud Detection dataset, but it performed low in terms of precision and recall when compared to other classifiers. The LSTM demonstrated low accuracy, precision, and recall on both datasets, showing that data may be imbalanced as other classifiers for fraud detection. Based on these findings, the ensembled classifier, the LSTM-Autoencoder, and the GAN are the most promising options for detecting credit card fraud.

The proposed approach might be suitable for any time series-related data and tasks. This is also useful for stock market (Yıldız et al., 2022) predictions and anomaly identification (Alkanhel et al., 2023). This is also important for weather forecasting and identifying (Kaya et al., 2023). The intrusion detection problem is a time series issue that necessitates. The intrusion detection problem (Al-Ghuwairi et al., 2023) is a time series issue that necessitates time series modelling. This can also be applicable in the prediction/forecasting and anomaly detection of time series data. The purpose of using multiple datasets in our work is to enhance and strengthen the learning capabilities of the proposed models.

The outcomes of the studies mentioned in this paper demonstrate the promise of deep learning for credit card anomaly detection and forecasting and the need for more research in this field. Deep learning algorithm advances and increased availability of annotated data are expected to contribute to further improvements in credit card data processing accuracy and efficiency.

5. Conclusion

This study investigated how deep learning techniques may be utilised for time series forecasting and anomaly detection, both critical jobs in data analysis with extensive applications in various sectors. This project aimed to use autoencoders and various deep-learning methods to detect credit card transaction records abnormalities. Experiments have shown that RNNs, particularly LSTM and GAN variations, are excellent for time series forecasting, particularly when long-term dependencies exist. Additionally, we demonstrated that RNNs and CNNs can successfully capture complicated and nonlinear patterns in time series data, resulting in enhanced predictions and better detection of abnormalities.

In this work, for the Credit Card Fraud Detection and Synthetic Credit Card Fraud Detection datasets, the ensembled classifier, the LSTM-Autoencoder, and the GAN model were determined to be the most successful classifiers. These classifiers performed admirably in identifying accuracy, precision, recall, F1-score, and AUC-ROC fraud. Also, the suggested statistical method for the feature selection approach outperforms current strategies in numerous ways. It does not change the original meanings of the features, selects characteristics based on statistical relevance, achieves high or even superior classification performance, is simple, fast, easy to implement, and computationally efficient, and avoids the dimension plague. It also does not change the original meanings of the features. This paper suggests that deep learning can improve the accuracy of anomaly identification and forecasting compared to standard approaches.

Finally, the results of our experiments indicate that Deep learning techniques for time series forecasting and anomaly detection have tremendous promise in various applications, such as finance, healthcare, manufacturing, and cybersecurity. As data-driven decision-making gets traction, these models will become more common. Deep learning for credit card anomaly detection and forecasting has the potential to improve the accuracy and efficiency of credit card data analysis significantly. Future research will lead to additional breakthroughs in this area.

CRediT authorship contribution statement

Amjad Iqbal: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Visualization, Writing – original draft. **Rashid Amin:** Writing – review & editing, Supervision, Project administration, Resources.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Ahmad, S., Lavin, A., Purdy, S., Agha, Z., 2017. Unsupervised real-time anomaly detection for streaming data. Neurocomputing 262, 134–147.

Ahmed, J., Gharakheili, H.H., Raza, Q., Russell, C., Sivaraman, V., 2019. Real-time detection of DNS exfiltration and tunneling from enterprise networks. Paper presented at the. In: Proceedings of the IFIP/IEEE Symposium on Integrated Network and Service Management (IM).

Ahmed, S.F., Alam, M.S.B., Hassan, M., Rozbu, M.R., Ishtiak, T., Rafa, N., Gandomi, A.H., 2023. Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artif. Intell. Rev. 56, 1–97.

Al-Ghuwairi, A.R., Sharrab, Y., Al-Fraihat, D., AlElaimat, M., Alsarhan, A., Algarni, A., 2023. Intrusion detection in cloud computing based on time series anomalies utilizing machine learning. J. Cloud Comput. 12 (1), 127.

Alkanhel, R., Khafaga, D., El-kenawy, E., Abdelhamid, A., Ibrahim, A., Amin, R., El-den, B., 2023. Hybrid grey wolf and dipper throated optimization in network intrusion detection systems. Comput. Mater. Contin. 74, 2695–2709.

Beju, D.G., Făt, C.M., 2023. Frauds in Banking System: Frauds with Cards and Their Associated Services *Economic and Financial Crime, Sustainability and Good Governance*. Springer, pp. 31–52.

Brause, R., Langsdorf, T., Hepp, M., 1999. Neural data mining for credit card fraud detection. Paper presented at the. In: Proceedings of the 11th International Conference on Tools with Artificial Intelligence.

Carcillo, F., Dal Pozzolo, A., Le Borgne, Y.A., Caelen, O., Mazzer, Y., Bontempi, G., 2018. Scarff: a scalable framework for streaming credit card fraud detection with spark. Inf. Fusion 41, 182–194.

Caruana, R., Niculescu-Mizil, A., 2004. Data mining in metric space: an empirical analysis of supervised learning performance criteria. Paper presented at the. In: Proceedings of the tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

Choi, K., Yi, J., Park, C., Yoon, S., 2021. Deep learning for anomaly detection in time-series data: review, analysis, and guidelines. IEEE Access 9, 120043–120065.

Credit card Fraud. (2013). https://data.world/raghu543/credit-card-fraud-data.

Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C., Bontempi, G., 2017. Credit card fraud detection: a realistic modeling and a novel learning strategy. IEEE Trans. Neural. Netw. Learn. Syst. 29 (8), 3784–3797.

Deb, C., Zhang, F., Yang, J., Lee, S.E., Shah, K.W., 2017. A review on time series forecasting techniques for building energy consumption. Renew. Sustain. Energy Rev. 74, 902–924.

Du, H., Lv, L., Guo, A., Wang, H., 2023. AutoEncoder and LightGBM for credit card fraud detection problems. Symmetry (Basel) 15 (4), 870.

Elkourchi, A., El Oualidi, M.A., Ahlaqqach, M., 2023. Demand forecast of pharmaceutical products during COVID-19 using holt-winters exponential smoothing. Paper presented at the. In: Proceedings of the International Conference on Artificial Intelligence & Industrial Applications.

Fiore, U., De Santis, A., Perla, F., Zanetti, P., Palmieri, F., 2019. Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. Inf. Sci. (Ny) 479, 448–455.

Gao, J., Zhou, Z., Ai, J., Xia, B., Coggeshall, S., 2019. Predicting credit card transaction fraud using machine learning algorithms. J. Intell. Learn. Syst. Appl. 11 (3), 33–63.

Ghosh, S., Reilly, D.L., 1994. *Credit card fraud detection with a neural-network*. Paper presented at the system sciences, 1994. In: Proceedings of the Twenty-Seventh Hawaii International Conference on.

Hasib, K.M., Azam, S., Karim, A., Al Marouf, A., Shamrat, F.J.M., Montaha, S., Rokne, J. G., 2023. MCNN-LSTM: Combining CNN and LSTM to Classify Multi-Class Text in Imbalanced News Data. IEEE Access.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Katzef, M., Cullen, A.C., Alpcan, T., Leckie, C., 2022. Generative adversarial networks for anomaly detection on decentralised data. Annu. Rev. Control 53, 329–337.

Kaya, Ş.M., İşler, B., Abu-Mahfouz, A.M., Rasheed, J., AlShammari, A., 2023. An intelligent anomaly detection approach for accurate and reliable weather forecasting at IoT edges: a case study. Sensors 23 (5), 2426.

Ketepalli, G., Tata, S., Vaheed, S., Srikanth, Y.M., 2022. Anomaly detection in credit card transaction using deep learning techniques. Paper presented at the. In: Proceedings of the 7th International Conference on Communication and Electronics Systems (ICCES).

Kibriya, H., Amin, R., Kim, J., Nawaz, M., Gantassi, R., 2023. A novel approach for brain tumor classification using an ensemble of deep and hand-crafted features. Sensors 23 (10), 4693.

Kim, J., Kang, H., Kang, P., 2023. Time-series anomaly detection with stacked Transformer representations and 1D convolutional network. Eng. Appl. Artif. Intell. 120, 105964.

Koundal, D., Guo, Y., Amin, R., 2023. Deep Learning in Big Data, Image, and Signal Processing in the Modern Digital Age, 12. MDPI, p. 3405.

Kour, H., Gupta, M.K., 2022. An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM. Multimed. Tools Appl. 81 (17), 23649–23685.

Kurien, K.L., Chikkamannur, A.A., 2019. Benford's law and deep learning autoencoders: an approach for fraud detection of credit card transactions in social media. Paper presented at the. In: Proceedings of the 4th International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT).

Lesot, M.J., Revault d'Allonnes, A., 2012. Credit-card fraud profiling using a hybrid incremental clustering methodology. Paper presented at the. In: Proceedings of the International Conference on Scalable Uncertainty Management.

Liang, D., Wang, J., Gao, X., Wang, J., Zhao, X., Wang, L., 2022. Self-supervised pretraining isolated forest for outlier detection. Paper presented at the. In: Proceedings of the International Conference on Big Data, Information and Computer Network (BDICN).

Lin, T.H., & Jiang, J.R. (2020). *Anomaly Detection with autoencoder and random forest.* Paper presented at the Proceedings of the International Computer Symposium (ICS).

Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., Alsaadi, F.E., 2017. A survey of deep neural network architectures and their applications. Neurocomputing 234, 11–26.

Malhotra, P., Vig, L., Shroff, G., & Agarwal, P. (2015). *Long short term memory networks for anomaly detection in time series.* Paper presented at the Proceedings of the Esann.

Mohammed, A., Kora, R., 2023. A comprehensive review on ensemble deep learning: opportunities and challenges. J. King Saud Univ. 35, 757–774.

Nguyen, H., Tran, K.P., Thomassey, S., Hamad, M., 2021. Forecasting and anomaly detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. Int. J. Inf. Manage. 57, 102282.

Niimi, A. (2015). *Deep learning for credit card data analysis.* Paper presented at the Proceedings of the World Congress on Internet Security (WorldCIS).

Noor, T.H., Almars, A.M., Alwateer, M., Almaliki, M., Gad, I., Atlam, E.S., 2022. SARIMA: a seasonal autoregressive integrated moving average model for crime analysis in Saudi Arabia. Electronics (Basel) 11 (23), 3986.

Panchal, U.K., .& Verma, S. (2019). *Identification of potential future credit card defaulters from non defaulters using self organizing maps.* Paper presented at the Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT).

Pandey, Y., 2017. Credit card fraud detection using deep learning. Int. J. Adv Res. Comput. Sci. 8 (5).

Pumsirirat, A., Liu, Y., 2018. Credit card fraud detection using deep learning based on auto-encoder and restricted boltzmann machine. Int. J. Adv. Comput. Sci. Appl. 9 (1), 18–25.

Rafique, R., Gantassi, R., Amin, R., Frnda, J., Mustapha, A., Alshehri, A.H., 2023. Deep fake detection and classification using error-level analysis and deep learning. Sci. Rep. 13 (1), 7422.

Raza, M., Qayyum, U., 2019. Classical and deep learning classifiers for anomaly detection. Paper presented at the. In: Proceedings of the 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST).

Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., Beling, P., 2018. Deep learning detecting fraud in credit card transactions. Paper presented at the. In: Proceedings of the Systems and Information Engineering Design Symposium (SIEDS).

Said Elsayed, M., Le-Khac, N.A., Dev, S., Jurcut, A.D., 2020. Network anomaly detection using LSTM based autoencoder. Paper presented at the. In: Proceedings of the 16th ACM Symposium on QoS and Security for Wireless and Mobile Networks.

Salazar, A., Safont, G., Vergara, L., 2018. Semi-supervised learning for imbalanced classification of credit card transaction. Paper presented at the. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN).

Santos, L.J.S., Ocampo, S.R., 2018. Bayesian method with clustering algorithm for credit card transaction fraud detection. Rom. Stat. Rev. 66 (1), 103– 120.

Schell, M.J., Yankaskas, B.C., Ballard-Barbash, R., Qaqish, B.F., Barlow, W.E., Rosenberg, R.D., Smith-Bindman, R., 2007. Evidence-based target recall rates for screening mammography. Radiology 243 (3), 681–689.

Seyedhossein, L., Hashemi, M.R., 2010. Mining information from credit card time series for timelier fraud detection. Paper presented at the. In: Proceedings of the 5th International Symposium on Telecommunications.

Shaukat, M.W., Amin, R., Muslam, M.M.A., Alshehri, A.H., Xie, J., 2023. A hybrid approach for alluring ads phishing attack detection using machine learning. Sensors 23 (19), 8070.

Silva, J.C.S., Macêdo, D., Zanchettin, C., Oliveira, A.L., de Almeida Filho, A.T., 2021. Multi-class mobile money service financial fraud detection by integrating supervised learning with adversarial autoencoders. Paper presented at the. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN).

Tingfei, H., Guangquan, C., Kuihua, H., 2020. Using variational auto encoding in credit card fraud detection. IEEE Access 8, 149841–149853.

Wu, T.Y., Wang, Y.T., 2021. Locally interpretable one-class anomaly detection for credit card fraud detection. Paper presented at the. In: Proceedings of the International Conference on Technologies and Applications of Artificial Intelligence (TAAI).

Yee, O.S., Sagadevan, S., Malim, N., 2018. Credit card fraud detection using machine learning as data mining technique. J. Telecommun. Electr. Comput. Eng. (JTEC) 10 (1–4), 23–27.

Yıldız, K., Dedebek, S., Okay, F.Y., Şimşek, M.U., 2022. Anomaly detection in financial data using deep learning: a comparative analysis. Paper presented at the. In:

Proceedings of the Innovations in Intelligent Systems and Applications Conference (ASYU).
Zamini, M., Montazer, G., 2018. Credit card fraud detection using autoencoder based clustering. Paper presented at the. In: Proceedings of the 9th International Symposium on Telecommunications (IST).
Zhang, A., Zhao, X., Wang, L., 2021. CNN and LSTM based encoder-decoder for anomaly detection in multivariate time series. Paper presented at. In: Proceedings of the IEEE 5th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC).