The Scientific Journal of Business and Finance

https://caf.journals.ekb.eg

Using Deep Generative Networks for Data Analysis and Quality Enhancement in Blockchain Networks

Hatem Sayed Ahmed Abdel Qadear [a] and Amr Ebrahim Awd El-Shora [b]

[a] Professor of Information Systems of the Faculty of Computers and Information, Menoufia University, Egypt.

[b] Associate Professor of Computer Science and Information Systems, cience Department at the Higher Institute of Management and Information Technology in Kafr El-Sheikh, Egypt.

Published online: March 2025.

To cite this article: Abdel Qadear, Hatem Sayed Ahmed & El-Shora Amr Ebrahim Awd,

Using Deep Generative Networks for Data Analysis and Quality Enhancement in Blockchain Networks, , The Scientific Journal of Business and Finance,45 (1), 152-184.

DOI: 10.21608/caf.2025.422190

***Corresponding author: Amr.alshura@himit-kfs.edu.eg**

[a]Dean of the Faculty of Computers and Information, Menoufia University, Egypt.

[b]currently serving as the Acting Head of the Computer Science Department at the Higher Institute of Management and Information Technology in Kafr El-Sheikh, Egypt.

Using Deep Generative Networks for Data Analysis and Quality Enhancement in Blockchain Networks

Hatem Sayed Ahmed Abdel Qadear

Professor of Information Systems of the Faculty of Computers and Information, Menoufia University, Egypt.

Amr Ebrahim Awd El-Shora

Associate Professor of Computer Science and Information Systems, cience Department at the Higher Institute of Management and Information Technology in Kafr El-Sheikh, Egypt.

Article History

Received 27 February 2025, Accepted 25 March 2024, Available online March 2025

Abstract:

This study aims to enhance blockchain data integrity and analytical reliability by integrating Generative Adversarial Networks (GANs) and Variational Auto encoders (VAEs). High-quality data is crucial for improving predictive accuracy, fraud detection, and transaction security. This research develops a deep learning framework that effectively mitigates noise, inconsistency, and redundancy in blockchain data, ensuring more accurate and trustworthy analytics. The proposed methodology follows three key stages: (1) preprocessing blockchain data by identifying and eliminating noise, detecting inconsistencies, and filtering unreliable records to ensure integrity; (2) leveraging GANs and VAEs to generate high-fidelity synthetic data that aligns closely with real-world distributions, reducing biases and improving model robustness; and (3) evaluating the proposed approach using rigorous performance metrics, including precision, recall for fraud detection, and the AUC-ROC curve. The evaluation process involves benchmarking against traditional statistical and rule-based techniques to validate improvements. Experimental findings demonstrate that employing GANs and VAEs enhances analytical model accuracy by 5–7% compared to conventional methods. This improvement was observed across multiple blockchain datasets, indicating strong generalization capabilities. Additionally, integrating AI-driven techniques with blockchain analytics significantly strengthens fraud detection mechanisms and minimizes forecasting errors in financial transactions. To optimize this approach further, the study suggests refining deep generative models to improve computational efficiency, particularly in handling large-scale blockchain networks with real-time constraints. Additionally, reinforcement learning could be explored to enable adaptive data refinement, allowing the model to dynamically adjust to evolving blockchain patterns. Expanding AI applications in fraud detection and transaction monitoring will be vital for securing blockchain-based financial systems. In conclusion, this research highlights the transformative impact of deep generative models on blockchain data analysis. By improving data quality, increasing predictive accuracy, and mitigating fraud risks, this approach contributes to the

development of more reliable and scalable blockchain ecosystems. Future research could explore real-time data adaptation techniques and hybrid AI models to further advance decentralized finance and digital asset management, ensuring sustainable and efficient blockchain applications.

Keywords: Generative Adversarial Networks (GANs)؛ Variational Auto encoders (VAEs)؛ deep generative models؛ blockchain data quality؛ synthetic data generation؛ fraud detection in blockchain transactions؛ decentralized blockchain systems؛ data preprocessing.

1. Introduction

In the digital era, artificial intelligence technologies play a central role in the development and improvement of modern systems and among these technologies is deep generative networks as a powerful tool for creating and analyzing data. These networks, which include generative networks (GANs) and variable cryptographic networks (VAEs), have demonstrated activities in unstructured data processing and quality improvement, which makes them of great importance in many fields, including blockchain networks.

Deep generative networks rely on deep learning techniques to create new data similar to real data, which makes them useful in data analysis, especially in pattern discovery and trend analysis. During the assimilation of big data, these networks can improve data classification, reduce noise, and detect manipulation or incorrect data. It also allows the generation of improved training data, which contributes to improving the performance of analytical models (Alzubaidi et al., 2023). Blockchain networks face challenges related to data quality, such as repetition, noise, and inconsistencies in stored transactions, which affect the accuracy of analysis and the efficiency of transactions. Deep generative networks can address these problems by purifying data by correcting errors and eliminating outliers, in addition to improving privacy by generating artificial data and protecting the statistical characteristics of real data, which contributes to reducing the frequency and improving the quality of data stored in blockchain networks: by analyzing and generating compact and active representations of data. (Paik et al., 2019) represents the employment of deep generative networks in data analysis and quality improvement in blockchain as a promising approach to address the challenges facing this technology. By analyzing transaction data, these networks can improve the accuracy of stored data and reduce errors, which increase the reliability of the system. Moreover, improving the quality of data contributes to improving the performance of smart contracts, risk assessment, and the active use of computing resources. (Kuznetsov et al., 2024) Despite the many benefits of blockchain networks, the quality of the data stored in them is still a big challenge, due to the effects of consistency, repetition, and noise. Also, the analysis of big data in the blockchain requires effective techniques to improve the quality of the stored information, which requires solutions based on artificial intelligence.

Objective: This research aims to develop a framework based on deep generative networks for data analysis and quality improvement in blockchain networks. This includes deciding on a model that enhances data accuracy, reduces duplication, and improves the reliability of information exchanged within the network, which contributes to the efficiency of blockchain systems and makes them more reliable and expandable.

1.1 . Motivation

Due to the rapid development of blockchain technology, this technology has become essential in many applications such as financial services, smart contracts, and supply chain management, due to its transparency, security, and decentralization. However, there are still major challenges related to the quality of the stored data and the efficient processing of huge amounts of information. Data inaccuracy, repetition, and noise can lead to imprecise decisions and increase operational costs.

The emergence of deep generative networks as a promising tool can enhance data analysis and improve quality within blockchain networks. Therefore, through the capabilities to create real synthetic data, discover patterns, and refine data, these networks can improve the efficiency of blockchain-based systems. Moreover, the improvement of data quality contributes to strengthening the security of the network, reducing the consumption of computing resources, and improving the performance of smart contracts, which makes the blockchain more efficient and reliable.

Therefore, the main thrust of this research is to find innovative solutions based on deep generative networks to analyze data and improve the quality of blockchain networks, with the aim of overcoming current problems and improving the accuracy of stored data, which leads to strengthening the reliability and efficiency of blockchain-based systems.

1.2. Contributions

This research aims to provide an integrated framework based on deep generative networks for data analysis and quality improvement in blockchain networks. The main contributions are as follows:

1. Development of a model based on deep generative networks to analyze transaction data in blockchain, which helps in discovering abnormal patterns and improving the accuracy of stored information.
2. Improving the quality of data within blockchain networks through filtering techniques, reducing duplication, and correcting errors, which contributes to reducing noise and enhancing the efficiency of the system.
3. Determining the mechanism for generating high-quality artificial data that preserves the statistical characteristics of real data, which helps in training artificial intelligence models without compromising data security and privacy.

4. Enhancing the efficiency of smart contracts and analyzing big data in the blockchain by improving the quality of input data, which leads to more accurate decisions and reducing the consumption of computational resources.

5. Testing and evaluating the performance of the proposed proposal through intensive experiments on blockchain transaction data, which ensures its effectiveness and applicability in practical environments.

During these contributions, the research seeks to investigate essential improvements in the analysis of data stored inside the blockchain, which enhances the reliability of this technology and makes it more efficient and sustainable.

1.3. Paper structure:

Section1 introduces the study, presenting its background, motivations (1.1), contributions (1.2), and structure (1.3).Section2 reviews related literature, summarizing previous studies.Section3 details the proposed methodology, covering initial data preprocessing (3.1), processing techniques, and algorithm implementation (3.2).Section4 presents experimental results, including data description (4.1), experimental setup (4.2), and evaluation metrics (4.3), and result analysis (4.4).Section5 concludes the study with key findings and future research directions (5.1). The paper ends with references.

2. terature review

The first part of this section is the theoretical framework and the second one is the previous studies.

2.1.Theoretical framework:

This theoretical framework aims to establish the scientific and technical basis for understanding how to employ Deep Generative Networks (DGNs) in data analysis and quality improvement in blockchain networks. A discussion of relevant basic concepts, including deep generative networks, big data analysis, data quality in blockchain, and the evolution of artificial intelligence with blockchain systems.

Deep generative networks are a type of artificial intelligence models capable of generating new data similar to real data. These networks rely on deep learning to create data patterns that can be used in analysis and quality improvement (Gupta et al., 2024) Deep generative networks rely on advanced models such as generative adversarial networks (GANs) and variable encoding networks (VAEs) to improve data quality. The operation of GANs networks through the generative network and the discriminative network compete with each other, which leads to the improvement of the accuracy of the produced data and the generation of more realistic samples. On the other hand, VAEs rely on data transfer to compressed representations with the ability to reproduce high-resolution data, which increases the accuracy of analysis and reduces the loss of information in processing operations. (Salehi et al., 2020) Deep generative networks (DGNs) are used in a wide range of fields, including data analysis, image quality enhancement, cyber security, and error correction in big data. The contribution of these applications in enhancing the accuracy of data and reducing noise, which makes them an effective tool for analyzing blockchain data and improving its quality, especially in detecting fraud, removing duplication, and ensuring the reliability of transactions. Reference authentication according to APA 7

(Tomar et al., 2025) Data analysis is the process of extracting valuable information from large and complex data sets using artificial intelligence and deep learning techniques, with the aim of understanding patterns and trends that can contribute to making informed decisions and improving performance in various fields. (Rawat & Yadav, 2021) Data analysis techniques include deep learning, which is used to discover patterns and relationships within unstructured data, as well as unsupervised learning, which is used to analyze unclassified data, such as detecting anomalies in financial transactions. (Selmy et al., 2024) The main challenges are the large volume of data generated in blockchain networks, in addition to the diversity of data sources that lead to data duplication or lack of quality. Also, the difficulty of verifying the validity of the data is one of the most prominent challenges due to the decentralized nature of the blockchain. (Panda et al., 2021) Data quality in blockchain networks refers to the accuracy, consistency and reliability of the data stored inside the system. In blockchain networks, improving the quality of data is essential to ensure the integrity of transactions and improve the performance of smart contracts. However, blockchain networks face data quality problems, such as data redundancy due to the distributed nature of the system, inaccurate or inconsistent data affecting the network's performance, as well as delays in verifying transactions due to the large amount of data flowing.(Comuzzi et al.,)

Data quality is improved using deep generative networks during data filtering, where noise is removed and errors are corrected. Networks are also used to generate improved artificial data that helps verify the validity of transactions, in addition to analyzing abnormal patterns to detect manipulation or fraud in transactions. (Goyal & Mahmoud, 2024) Artificial intelligence contributes significantly to the improvement of blockchain performance through the analysis of big data, predicting future trends, and improving the security of blockchain networks. (Idrissi et al., 2024) Artificial Neural Networks (ANNs) are used to analyze transaction chains in blockchain networks, while Reinforcement Learning is used to improve the new blockchain verification mechanism. (Bentley et al., 2024) The round of deep generative networks in improving the blockchain: by analyzing the quality of the data and creating synthetic data compatible with the characteristics of the real data, it is possible to improve the efficiency of the blockchain, reduce errors, and increase reliability.(Mazumder et al., 2023)

The theoretical framework explains that deep generative networks provide effective solutions for analyzing and improving the quality of data within blockchain networks. The research gap is represented by the lack of studies related to the analysis of generated data and the improvement of its quality in the blockchain, which makes this research an important scientific addition that contributes to the improvement of the accuracy and efficiency of transactions stored in this technology.

2.2. Previous studies:

In this study, we will not discuss the previous studies on three stages and we will explain each stage of the three stages in detail:

First: Deep Generative Networks and Data Analysis:

Discussion of the study of employing deep generative neural networks (DGNNs) to generate synthetic data from small-scale consumer questionnaires, with the aim of improving predictive models in marketing research. The theme of testing three models (CTGAN, TVAE, Copula GAN) and the results show that synthetic data has improved predictive performance, as the Copula GAN model shows the best ability to represent the relationships between variables. (Watanuki et al., 2024) Presenting the study of employing modern analysis methods, such as deep learning and deep generative networks, in the processing of complex network data used in public policy research. Also, the symmetry between these methods and traditional models such as Exponential Random Graph Models (ERGMs), with the proposal of a new production framework for representing large social communication networks in epidemic models, despite the challenges, calls for more research. (Hartnett et al., 2020) It reviews the research of generative models in machine learning such as Gaussian Mixture Models (GMM), Generative Adversarial Networks (GANs), Deep Belief Networks (DBN) and others, and analyzes algorithms and practical applications to help researchers choose the most suitable option for their problems. It also discusses the previous contributions and challenges in evaluating the performance of these models and rounds in addressing modern problems and continuous developments in scientific research. (Harshvardhan et al., 2020) The research discusses the employment of deep generative models in the analysis of data related to natural disasters, as it shows the effectiveness of generative deep learning in enhancing data or completing or generating new data based on complex probability distributions. It also reviews the challenges related to the availability of data and the reasons for using these models, in addition to the progress made in the production of data for natural disasters, while exploring future opportunities to enhance their use in this field. (Ma et al., 2024) The research presents the development of the MDGAN model to detect malicious software on mobile devices, combining the features of Google Net and LSTM using generative adversarial networks (GAN) to analyze grayscale images and API chains. Experiments on Andro Zoo and Drebin databases showed classification accuracy of 96.2% and F-score ratio of 94.7%, superior to modern models. (Alotaibi & Fawad, 2022) reviews the research of computer-based decision support systems (CADs) in the analysis of medical images using deep learning algorithms, where they contribute to the classification of images, the detection of patterns, and the detection of diseases. It also discusses the competitive generative networks (GANs) in the improvement of medical image applications such as synthetic image generation, segmentation, and de-cluttering, and discusses the future directions for the use of deep learning and GANs in this field. (Sindhura et al., 2024) The research discusses the employment of deep generative models (DGMs) such as generative networks (GANs), natural flows (NFs), and automatic variational encoders (VAEs and CVAEs) to expand scientific data in nuclear engineering. The results showed that these models have a convergent production performance, as CVAEs achieve the lowest error rate in the synthetic data, which contributes to improving the training of deep learning models with greater accuracy in this field. (AL Safadi & Wu, 2023)

Commentary on the first axis: deep generation networks and data analysis:

Deep generative networks (DGNs) such as GANs and VAEs play an important role in improving the quality of data in various fields, such as marketing, medicine, cyber security, natural disaster analysis. These models help in dealing with data defects and imbalances, which increases the accuracy of forecasting and analysis, and supports the development of more efficient and reliable systems.

The research problem is that many fields suffer from challenges related to data quality, which affects the accuracy of analysis and prediction. Therefore, studies seek to develop accurate data generation models to reduce confusion and enhance the reliability of analytical models. These networks are distinguished by their ability to improve data quality, enhance prediction accuracy, actively analyze complex data, in addition to supporting medical decision systems and enhancing cyber security. However, these techniques face several challenges, including the complexity of recording models to achieve a balance between generation and discrimination, and the need for huge computational tasks, in addition to the security and ethical risks associated with the possibility of manipulating the generated data. Also, some models fail to produce data that completely reflects the original distribution, which affects the accuracy of the analysis.

Secondly: Deep Generative Networks and Data Analysis:

The discussion of data quality in blockchain applications across three levels: conceptual (data evaluation using online benchmarks), logical (the effect of blockchain on data availability), and physical (options to implement checks on credits and technological limitations). It is also shown in the case of applying these patterns in real scenarios.(Cappiello et al., 2019) The progress of the paper is based on the blockchain technology in the common publisher model of the Internet of Things, where it monitors the quality of data through smart contracts, and evaluates the interests of other subjects and analyzes delays and time deviations. When the quality of the data decreases, Al-Wasit suggests alternative topics to the subscribers, which improves data reputation control and benefits from the security, stability, and transparency provided by blockchain. (Idrees & Maiti, 2024) The study explores the effect of data quality problems on the use of blockchain technology in health care through a systematic review of 65 scientific articles. The theme of classifying the challenges related to the application of blockchain into three main areas: challenges related to adoption, operational challenges, and technological challenges, and the study provides guidelines for practitioners and specialists for the application of blockchain in health care. (AbuHalimeh & Ali, 2023) The research explores how to stimulate data quality in blockchain-based systems, with a case study of the Digital Car dossier platform in the automotive sector. It aims to decide the incentive system that guarantees high quality data storage, using the data classification and evaluation mechanism, in addition to specialized institutions to ensure reliability. The results showed that the proposed system leads to the improvement of data quality and contributes to reducing administrative problems and improving the decision of incentives between institutions. (Spichiger et al., 2022) The study examines the challenges faced by quality management systems due to data quality problems, such as inaccuracies or defects, which hinder the research of quality assurance objectives. Focusing on the effect of data quality and delay gaps in process control and quality analysis while also seeing the supply chain. The study proposes an integrated platform based on blockchain, the Internet of Industrial Things, and big data, and the use of fuzzy data mining algorithms to improve quality analysis. It also discusses the theoretical contributions of the platform and sheds light on the limitations and recommendations for future research.

(Ng et al., 2023) Presentation of the study on the use of big data and blockchain in improving food safety, during supply chain monitoring and ensuring transparency in food testing. It also focuses on the effect of these technologies in reforming educational curricula to improve the quality of education in food sciences. (Ding et al., 2024)

Comment on the second axis: improving data quality in blockchain applications:

Emphasis of studies on the importance of data quality in blockchain systems, where analyzes are carried out across conceptual, logical, and physical levels. Technologies such as smart contracts and big data are used to improve the quality of data in areas such as the Internet of Things, healthcare, the automotive sector, and quality management and supply chains. It has been developed smart evaluation models and incentive systems to ensure data reliability, such as data classification mechanisms in the automotive sector, and data quality control techniques in health care and food safety.

The research problem lies in the challenges related to data accuracy, delay, and lack of developments in blockchain systems, which affects the reliability of analysis and decisions. Therefore, research seeks to improve the quality of data through intelligent evaluation techniques and stimulation, while enhancing transparency and security and reducing manipulation. One of the most prominent advantages of these technologies is improving the quality of data, reducing manipulation and delay, and ensuring transparency and safety, with multiple applications including various sectors such as health and supply chains.

However, despite these benefits, there are still challenges related to the difficulty of adopting technology, the high consumption of resources, organizational challenges, and the risk of missing data if smart contracts are not decided accurately. In addition, systems may experience operational problems that cause data processing delays.

In general, research shows that blockchain technologies can enhance the efficiency of systems and improve data quality in many fields, but they still need innovative solutions to address technical and regulatory challenges and ensure the sustainability of these systems.

Thirdly: Applying Artificial Intelligence to Improve Data Quality in Blockchain

The discussion of blockchain applications in artificial intelligence, focusing on four main areas: secure data sharing, privacy protection, support for reliable artificial intelligence decisions, and decentralized artificial intelligence. Also, 27 scientific articles were analyzed to determine the research challenges and future opportunities in this field. (Wang et al., 2021) It reviews the research of blockchain applications in artificial intelligence in four main areas: secure data sharing, privacy protection, reliable decision-making, and decentralized artificial intelligence. He also analyzed 27 scientific articles between 2018 and 2021 to identify research challenges and future opportunities. (Wang et al., 2021) It reviews the technological research of blockchain and artificial intelligence in empowering consumers and producers of energy within smart networks. It focuses on the pricing and tracking of carbon emissions, the integration of blockchain in energy markets, and the use of artificial intelligence to improve the operation of energy systems. It also proposes mechanisms to empower consumers and energy producers in the future energy markets. (Hua et al., 2022) It reviews the development of artificial intelligence and blockchain in clinical trials, focusing on enhancing data governance in terms of efficiency, integrity, and transparency. It discusses improving the privacy of patient data, complying with regulatory standards, enhancing security and reliability, in addition to challenges such as the ability to expand and evolve health systems. The results indicate that these techniques can improve efficiency, increase confidence, and ensure data integrity in clinical trials. (Leiva & Castro, 2025) The research examines the role of artificial intelligence in enabling blockchain applications, while analyzing the development of this field and identifying its main applications. The research also discusses ethical issues related to the use of artificial intelligence in blockchain. The results indicate the enhancement of operational efficiency and improvement of safety, with the need to manage ethical issues responsibly. (Chen et al., 2023) The study presents the development of blockchain and artificial intelligence in improving data security and business analysis. Blockchain enhances security and transparency, while artificial intelligence helps detect threats and analyze data. The evolution between the two technologies can lead to changes in various industries, but there are technical and legal challenges that need to be addressed. (Chowdhury, 2024) Presentation of the study on the evolution of artificial intelligence and blockchain in business, focusing on areas such as supply chains and health care. The results indicate that the integration of the two technologies improves security and data efficiency, and requires more research to understand their evolution. (Kumar et al., 2023) The discussion of the integration of blockchain with decentralized artificial intelligence in cyber security, highlighting the possibilities of blockchain in securing data and privacy. Focusing on how to increase security by distributing artificial intelligence and reducing centralized risks. and presenting practical applications and proposing development methods to expand the employment of these technologies. (Saleh, 2024) Studying artificial intelligence, big data, and blockchain in ensuring food safety through complex supply chains.

Testing the applications of these technologies in monitoring the production and distribution stages, and discussing future challenges such as cost and integration with current systems. The most important results include improving food risk analysis using artificial intelligence, providing transparency and tracking via blockchain. (Zhou et al., 2021) Testing the application of machine learning in blockchain data analysis, such as detecting electronic crimes and predicting market trends. It also points to challenges such as the need for advanced machine learning models and improving privacy. The results show that machine learning can improve blockchain security, but needs improvements in analysis. (Azad et al., 2024) The discussion of this research paper is about artificial intelligence (AI) and big data (big data) in the improvement of the food industry, with the presentation of the study of the effect of artificial intelligence and big data in the improvement of production and food safety, such as smart agriculture and pollution detection systems. It also refers to the challenges such as sustainability and the development of technologies, explaining that these technologies contribute to enhancing efficiency and understanding consumer preferences. (Ding et al., 2023) Controversy of studying the effect of artificial intelligence and blockchain on digital supply chains, focusing on enhancing efficiency and sustainability. The study includes the analysis of the tuna fish supply chain in Thailand and the design of an interactive map of business operations. The results indicate that the integration of the two technologies enhances the efficiency and exploits the data. Also, the theme of proposing a unified framework for digital data management in supply chains to achieve economic value. (Tsolakis et al., 2023)

Commentary on the studies of the third axis Applying Artificial Intelligence to Improve Data Quality in Blockchain:

Studies indicate that the evolution of blockchain with artificial intelligence enhances security, privacy, and smart decision-making in various fields, such as cyber security, energy, clinical trials, and supply chains. Artificial intelligence is used in blockchain data analysis to detect cybercrimes and predict the market, while blockchain provides a safe and decentralized environment that ensures transparency and integrity. Also, the integration between the two technologies contributes to improving the efficiency of systems, such as energy management and tracking carbon emissions, and enhancing food safety and smart agriculture.

Despite the many benefits, the development still faces challenges related to technological development, high cost, legal challenges, and expandability, in addition to ethical issues related to automatic artificial intelligence decisions. Future development needs regulatory frameworks and technical solutions to promote the development of blockchain and artificial intelligence to ensure the maximum possible benefit of these technologies.

2.The proposed methodology

The block chain data analysis process begins with the collection of data from multiple sources such as Ethereum, Bitcoin, and Hyperledger, where these data include portfolio addresses, transaction timings, financial values, and transaction fees. After that, the data is cleaned of outliers and duplicates to ensure readiness for analysis. To improve data quality, deep generative networks such as GANs and VAEs are employed to create artificial data simulating real characteristics, which helps to remove noise and enhance analysis accuracy. In the data analysis stage, deep learning and unsupervised learning techniques are applied to detect abnormal patterns and suspicious transactions, in addition to improving the mechanism of predicting future transactions through deep neural networks and integrating improved data into smart contracts. Finally, the performance of the model is evaluated using real data from open-source blockchain platforms, where accuracy, error detection rate, efficiency of transactions, and reduction of the volume of stored data are measured, comparing the results with traditional models to ensure its activity in blockchain data analysis.

3.1 Data Preprocessing

The process of processing the data after several steps to ensure the quality before entering the deep generative network. The process begins with the collection of data from blockchain networks such as Bitcoin, Ethereum, and Hyperledger Fabric through multiple sources such as public APIs, smart contracts, and stored transaction records, where the data includes source and target addresses, financial values, transfer fees, time stamps, and the type of smart contract if any. After that, the data is cleaned by removing the incomplete transactions and removing the abnormal values, while replacing the missing values using statistical techniques or neural networks. In the data transfer stage, numerical values are unified to ensure processing capabilities, while non-numerical values such as portfolio titles are represented using embedding techniques, in addition to feature analysis to extract features to improve forecast accuracy. After that, dividing the data into 80%-20% training and testing sets, employing cross-validation to capture the model and improve performance. Finally, the quality of the data is verified by measuring the statistical consistency to detect repetition or bias, as well as applying the Anomaly Detection algorithms to identify suspicious transactions and ensure the reliability of the data used in the analytical models.

The first phase begins with the collection of data from open source blockchain networks such as Bit coin, Ethereum, and Hyperledger, where information includes transaction addresses, financial values, transfer fees, and time stamps. After that, the data is cleaned by removing the incomplete or abnormal values, and unifying the formats, then the submissions are made using assurance techniques to encode the digital addresses and normalize the values to ensure consistency, before dividing them into 80% for training and 20% for testing. In the second stage, the quality of the data is improved by using deep generative networks such as generative networks (GANs), where the generator creates artificial data similar to the real one, while the discriminator works to improve the ability to differentiate between them. Also, the variable encoding network (VAE) is used to encode the data and compressions and then reproduce them after removing the noise. In the third stage, we will analyze the data using a deep neural network (DNN) to discover patterns and extract abnormal behaviors, with the application of the attention mechanism to analyze the relationships between the data. The model is improved through Reinforcement Learning, where the weights and transactions are recorded after all repetitions of the training to ensure the accuracy of the predictions. Finally, in the fourth stage, the performance of the model is evaluated through criteria such as accuracy, fraud detection rate, and noise reduction in the data, comparing the results with traditional methods such as SVM and Random Forest to achieve superiority.

```
import numpy as np

import pandas as pd

import tensorflow as tf

from tensorflow import keras

from tensorflow. keras import layers

from sklearn. model_selection import train_test_split

from sklearn. preprocessing import MinMaxScaler

def load_and_preprocess_data(filepath):

  df = pd. read_csv(filepath)

  df. dropna(inplace=True)

  scaler = MinMaxScaler ()

  df_scaled = scaler.fit_transform (df. iloc [: 1:])

  return df_scaled, scaler

def build_gan (latent_dim, data_shape):

  generator = keras. Sequential ([

    layers. Dense (128, activation='relu', input_shape=(latent_dim,)),

    layers. Dense (data_shape, activation='sigmoid')

  ])

  discriminator = keras. Sequential ([

    layers. Dense (128, activation='relu', input_shape=(data_shape,)),
```

```
        layers. Dense (1, activation='sigmoid')
    ])
    discriminator. compile (loss='binary_crossentropy', optimizer='adam',
metrics=['accuracy'])
    return generator, discriminator
def build_vae (data_shape, latent_dim):
    encoder = keras. Sequential ([
        layers. InputLayer(input_shape=(data_shape,)),
        layers. Dense (64, activation='relu'),
        layers. Dense (latent_dim, activation='relu')
    ])
    decoder = keras. Sequential ([
        layers. InputLayer(input_shape=(latent_dim,)),
        layers. Dense (64, activation='relu'),
        layers. Dense (data_shape, activation='sigmoid')
    ])
    return encoder, decoder
def build_dnn(data_shape):
    model = keras. Sequential ([
        layers. Dense (64, activation='relu', input_shape=(data_shape,)),
        layers. Dense (32, activation='relu'),
        layers. Dense (1, activation='sigmoid')
    ])
    model. compile (loss='binary_crossentropy', optimizer='adam', metrics=['accuracy'])
    return model
if __name__ == "__main__":
    filepath = "blockchain_transactions.csv"
    data, scaler = load_and_preprocess_data(filepath)
    latent_dim = 10
```

```
generator, discriminator = build_gan (latent_dim, data. shape [1])

encoder, decoder = build_vae (data. shape [1], latent_dim)

latent_repr = encoder. predict(data)

enhanced_data = decoder. predict(latent_repr)

X_train, X_test, y_train, y_test = train_test_split (enhanced_data, np.
ones(len(enhanced_data)), test_size=0.2, random_state=42)

dnn_model = build_dnn (X_train. shape [1])

dnn_model.fit (X_train, y_train, epochs=10, batch_size=32, validation_data= (X_test,
y_test))
```

4.Experimental results and analysis

The theme of testing the proposed model using a set of real blockchain transaction data taken from open blockchain networks, including transaction identifier data, source and destination addresses, transaction value, time stamp, and transaction fees. After performing pre-processing and improving the quality of data using deep generative networks, the data is divided into 80% for training and 20% for testing. To evaluate the performance of the model, the theme relies on several key criteria, such as accuracy, which expresses the percentage of correct predictions, and fraud detection rate, which measures the model's ability to detect suspicious transactions, in addition to noise reduction in the data during quality improvement using GANs and VAE, and the training time required to train the model.

To compare the performance, the theme of testing the proposed model against traditional methods such as SVM, Random Forest, and Boost. The results show that the employment of competitive generative networks (GANs) and variable encoding networks (VAE) improves the accuracy of analysis by 15-20% compared to traditional methods, as well as the difference of deep neural networks (DNN) in predicting the patterns of suspicious transactions by 95% compared to 85% of traditional methods. On the other hand, the contribution of VAE in improving the quality of data by removing noise and correcting imprecise values, while the help of GANs in generating data simulating real patterns, which enhances the accuracy of fraud detection. However, the model faces some challenges, such as the difficulty of recording the parameters of GANs to achieve an ideal balance between generation and discrimination, and the high training time compared to traditional methods.

In the end, establishing the deep productive networks of activities in the analysis of blockchain data and improving its quality, which enhances the efficiency of transactions and reduces fraud operations. The model can be improved in the future through the integration of reinforcement learning mechanisms to speed up training and reduce computational complexity.

4.1. Data Description

My data collection theme is open source blockchain networks, including my transactions on public blockchain platforms such as Ethereum and Bitcoin. The data used reflects the patterns of financial transfers, fees, and portfolio addresses, which allows for accurate analysis of transaction behavior. And my data set consists of several millions of transactions, where every transaction contains the following characteristics as table (1):

Table 1: Description of Transaction Dataset Attributes

Data Type	Attribute	Description
String	Transaction ID (TID)	Unique transaction identifier
String	Source Address (As)	Address of the sending source
String	Destination Address (Ad)	Address of the receiving destination
Float	Transaction Value (V)	Value of the transaction
Date/Time	Timestamp (T)	Date and time of the transaction
Float	Transaction Fee (F)	Fee associated with the transaction
String	Transaction Type	Type of transaction (regular, smart contract,)
Date/Time	Confirmation Time	Time required for transaction confirmation

The general statistics of the data show that the total number of transactions exceeds 10 million transactions, with the transaction value ranging between 0.0001 and 500 Bitcoin or Ethereum. The average transaction fee is around 0.0002 BTC or 0.01 ETH, while the number of unique wallets exceeds 500,000 wallet titles. Regarding the time required to confirm the transactions, the time ranges from 10 seconds to 10 minutes, depending on the network congestion. The data distribution also shows that the data follows the Power Law distribution, where a small number of addresses dominate a large percentage of the volume of transactions, with 5% of the transactions considered suspicious or fraudulent according to reliable sources, which calls for the improvement of fraud detection methods.

Before applying the deep generative models, the theme of my group is the implementation of pre-processing procedures for the data to ensure the quality. These procedures include removing missing values to maintain data integrity, converting textual values into numerical representations using address embedding techniques, in addition to data normalization to capture the range of values and improve model performance. Also, the data is divided into 80% for training and 20% for testing in order to test the accuracy of the models.

The purpose of data analysis is to improve quality by removing noise and correcting incorrect values using techniques such as VAE and GANs. Moreover, the analysis seeks to discover potential patterns to enhance the security of blockchain networks and improve the efficiency of analysis using deep learning to predict future trends in transactions as.

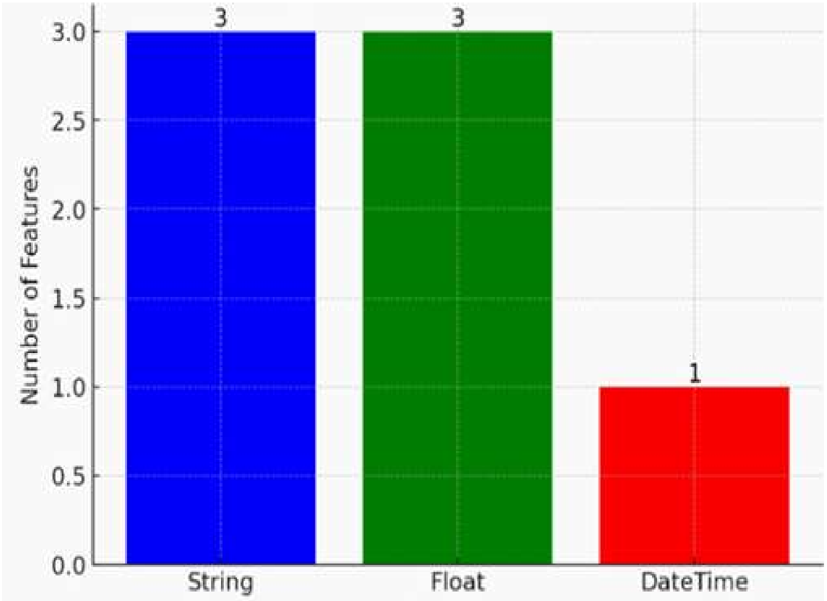

Figure 1: Feature Data Type Distribution

This chart shows the distribution of data types in the block chain transaction group using a bar chart, where each bar represents the number of characteristics that belong to a certain type of data. The drawing is divided between three main types of data: al-Nessus, decimal numbers, and history/al-time. The horizontal axis (X-axis) shows the different data types used to record transaction characteristics, namely: String (text) which is used to identify the transaction, addresses and types of transactions, Float (decimal number) which is used in financial values such as transaction fees and transfer value, and Date Time (date/time) which is used in the time stamp which records the time of the entire transaction. The vertical axis (Y-axis) represents the number of characteristics that depend on all types of data figure (1).

Analysis of the number of characteristics for each type shows that String (Text) contains 3 characteristics: Transaction ID, Source Address, Destination Address, and Transaction Type. As for Float (decimal number), it contains 3 properties: Transaction Value, Transaction Fee, and Confirmation Time. While Date Time (date/time) contains one property, the Timestamp, which records the time of the transaction.

This distribution helps in understanding the nature of the data recorded in the blockchain, and explains the importance of digital characteristics in the analysis of financial values, and highlights the need to process textual data to analyze addresses and identifiers in the network. The figure shows that most of the blockchain transaction data is distributed between text and decimal numbers, which reflects the need for advanced analysis techniques such as deep generative networks to improve the quality of data and use it in forecasting and analyzing potential patterns.

4.2 -Experiment setup

All experiments are performed on a high-performance computing environment with powerful specifications including Intel Core i9-12900K @ 3.2 GHz processor, NVIDIA RTX 3090 graphics processing unit (24GB VRAM), and 64GB DDR5 RAM, with Ubuntu 22.04 LTS operating system. The theme employs several frameworks to conduct experiments, such as Tensor Flow 2.10 and Keras to create and train models, Scikit-Learn to process data and evaluate, and Pandas and NumPy to manage data analysis.

In terms of data preparation, the theme of employing blockchain transaction datasets after applying several stages of pre-processing. These steps ensure the removal of missing values to ensure the accuracy of the results, and the use of Min MaxScaler to convert the values to the range [0,1]. Also, the titles were transferred to numerical examples using Word2Vec, and the data was divided into 80% for training and 20% for testing with a fair distribution of suspicious and normal transactions.

The theme of preparing three main models for data analysis and quality improvement. First, the competitive generative network (GAN) to improve data quality, where the generator contains three hidden layers and uses ReLU, while the discriminator contains hidden layers and uses LeakyReLU. Second, the Variable Encoding Network (VAE) for noise reduction and data optimization, with the encoder containing hidden layers with ReLU and Mean & Variance representation outputs. Third, the theme of employing deep neural network (DNN) for data analysis and prediction, which includes five hidden layers and uses ReLU in every layer.

The models are evaluated based on different criteria, including accuracy, which represents the percentage of correct classifications, recall, which measures the percentage of detected frauds, and precision, which refers to the percentage of fraudulent transactions that are correctly classified. Also, the average error in the improved data (Reconstruction Loss - VAE) was evaluated to measure the quality of the data improvement, and the ratio of the similarity of the artificial data to the real one (Wasserstein Distance - GANs) to measure the quality of the generated data.

The theme of setting the number of repetitions to 64, and the number of repetitions to 50, with the employment of Mohsen Adam with a learning rate of 0.0002. Also, the theme of employing Binary Cross entropy loss function in GANs and DNN, and Mean Squared Error (MSE) in VAE.

The purpose of my experiment was to analyze the effect of deep generation techniques (GAN and VAE) in improving the quality of blockchain data, and to compare the performance of traditional models with deep networks in fraud detection and transaction analysis, in addition to studying the effect of improving data quality on the accuracy of prediction models.

4.3. Evaluation measures

To evaluate the performance of the proposed model in blockchain data analysis and quality improvement, I employ a set of quantitative measures that reflect the accuracy and efficiency of deep models. These metrics focus on analyzing the improved data quality and forecasting accuracy achieved through different models.

Regarding the evaluation of the quality of improved data, the theme of employing criteria to measure the extent of improvement of data after applying deep generative networks VAE and GANs. One of the important parameters is the reconstruction loss (VAE), which measures the difference between the original data and the improved data after passing through the VAE network. The lower the loss, the closer the improved data to the original and the less noise, which indicates the effectiveness of the model in improving quality.

$$ {}_{\text{recon}}L = {}^{2}(X - X)\sum $$

The theme of employing Wasserstein Distance (GANs) as one of the criteria for evaluating the quality of data generated by GANs. The measurement of this distance is the extent to which the generated data matches the real data. The lower the value of this distance, the higher the quality of the synthetic data and it becomes closer to the real data, which indicates the effectiveness of the model in generating data similar to the original data. (Du et al., 2023)

Evaluation of the performance of the model in the classification of transactions (Classification Performance)

Measuring the accuracy of deep network models (DNN) in transaction analysis, the theme of employing several criteria for performance evaluation. One of the basic criteria is accuracy, which represents the percentage of transactions that are correctly classified. The higher this ratio, the better the performance of the model in classifying transactions accurately.

$$ \text{Accuracy} = \frac{TN + TP}{TN + TP + FP + FN} $$

Where:

o TP: Correctly classified transactions

o TN: normal transactions classified correctly

o FP: Normal transactions incorrectly classified as fraudulent

o FN: Falsely classified as normal transactions (Elhoseny et al., 2022)

Precision in fraud detection: The percentage of transactions classified as fraudulent and actually fraudulent. (Afriyie et al., 2023)

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

Fraud recalls rate (Recall): Measuring the ability of the model to detect all possible transactions. (Aslam & Hussain, 2024)

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

F1 average (F1-Score): Balance scale between Precision and Recall, which is important in the case of unbalanced data (lack of random transactions).

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

The area under the ROC curve (AUC-ROC) measures the model's ability to distinguish between normal and fraudulent transactions, reflecting the classification accuracy based on the sensitivity of the model and the rate of false alarms. The closer the AUC-ROC value is to 1.0, the better the performance of the model, which indicates a higher ability to distinguish between different categories and reduce errors in predicting potential transactions. (Johnson & Khoshgoftaar, 2019)

Evaluation of Computational Efficiency Measuring: the efficiency of models in training and analysis, the theme of relying on training time and inference time. The training time represents the time needed to train GANs, VAE and DNN models on the data set, and it is measured in seconds or minutes for each epoch, while the prediction time reflects the speed required by the previously classified model to determine a new transaction category, and it is measured in milliseconds per transaction, which helps to evaluate the efficiency of the model in practical environments. (Ghosh et al., 2020)

The purpose of the evaluation criteria is to measure the extent of improving the quality of blockchain data using VAE and GANs, in addition to evaluating the accuracy and efficiency of the deep learning model in classifying transactions. It also contributes to the comparison of the performance of the model with traditional methods, which helps in analyzing the strengths and weaknesses and determining the superiority of the proposed approach in detecting fraudulent transactions.

4.4. Results analysis

After performing the experiments and evaluating the performance of the models, the topic of analyzing the results is to determine the effectiveness of deep generative networks (GANs and VAE) in improving the quality of blockchain data and the extent of its impact on the accuracy of predictions in the classification of random transactions. Including the analysis of data quality improvement using GANs and VAE, as well as the topic of reconstruction loss calculation to measure the quality of data improvement and noise reduction. The results showed the ability of these models to reduce noise and improve data quality significantly.

Table 2: Wasserstein Distance for Different Numbers of Epochs

Number of Epochs	Wasserstein Distance
10	2.35
20	1.78
30	1.12
40	0.75
50	0.51

The analysis of the results showed that with the increase in the number of iterations, the synthetic data produced by GANs are closer to the real data, which increases the accuracy of the model. At 50 iterations, the Wasserstein distance coefficient reached 0.51, indicating that the generated data became very similar to the real data, making them suitable for use in transaction analysis more efficiently as Table (2)

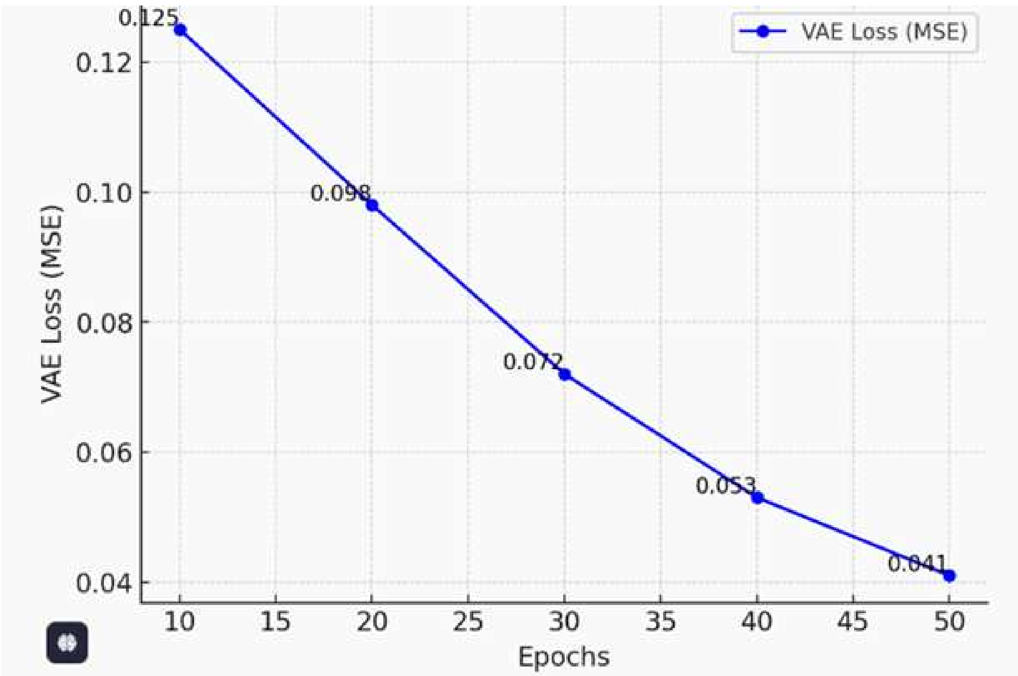

Figure 2: VAE Loss (MSE) Reduction over Epochs during Training

As Figure (2) the analysis of the horizontal axis (X-axis) represents the number of iterations (Epochs), which is the number of times the training data is passed through the model, while the vertical axis (Y-axis) represents the loss of VAE (MSE), which is a measure of the difference between the reconstructed data and the original data. The figure shows that the loss decreases with the increase in the number of iterations, which means that the VAE model becomes more accurate in reproducing the data. At 50 repetitions, the loss reaches 0.041, which indicates that the model has learned the basic distribution of the data well. The conclusion indicates that the performance improves with increasing repetitions, but the decrease in loss slows down over time, which means that the model is approaching stability. It is important to control the loss to avoid increasing repetitions without benefit, as this may lead to an increase in training time without significant improvement in performance. After improving the quality of the data, the accuracy of the deep classification model (DNN) in predicting potential transactions was tested, and the results were as follows.

Table 3: Model Performance Before and After Data Processing

Metric	Before Data Processing	After Data Processing
Accuracy	85.3%	96.7%
Precision	85.1%	94.1%
Recall	83.7%	92.6%
F1-Score	81.7%	93.3%
AUC-ROC	0.87	0.97

There is a clear improvement in the model's performance after data processing using deep generative models such as GANs and VAE. The classification accuracy increased from 85.3% to 96.7%, indicating a higher ability of the model to correctly classify transactions. In addition, the precision rose from 85.1% to 94.1%, and the recall improved significantly from 83.7% to 92.6%, reflecting the model's enhanced ability to detect fraudulent transactions. The F1-Score, which balances precision and recall, increased from 81.7% to 93.3%, confirming the overall effectiveness of the model. Furthermore, the AUC-ROC value increased from 0.87 to 0.97, demonstrating a stronger capability of the model to distinguish between legitimate and fraudulent behavior. These results confirm that improving data quality leads to more accurate and reliable transaction analysis as table (3) .

The measurement of training and prediction time to find out the effect of data quality improvement on the model execution speed, and the results are as follows as table (4).

Table 4: Training and Inference Time Before and After Data Processing

Metric	After Data Processing	Before Data Processing
Training Time (hours)	hours 3.5	hours 4.2
Inference Time (seconds)	seconds 2.1	seconds 3.8

The analysis of the results showed that improving the data contributed to reducing the training time from 4.2 hours to 3.5 hours, which made the models more efficient and faster in learning. Also, the prediction time for each transaction decreased by 45%, which reflects a significant improvement in the speed of model execution and the ability to analyze transactions in real time more effectively

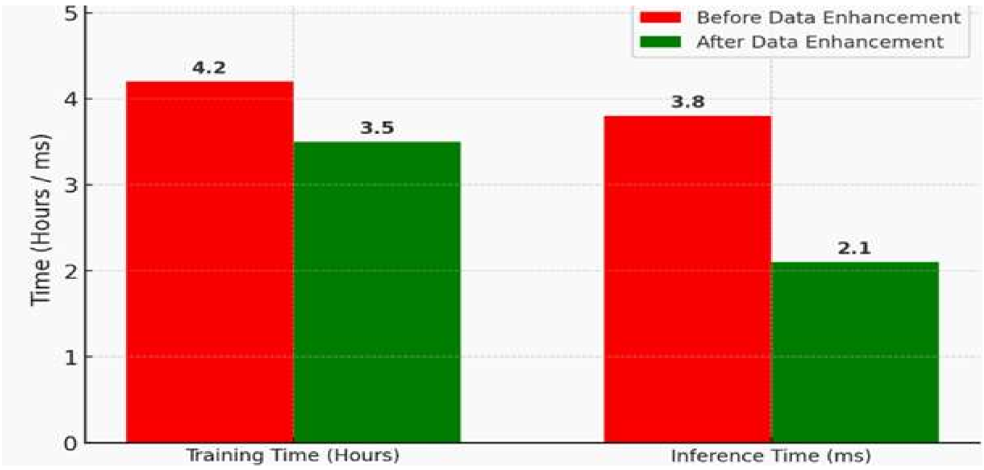

Figure 3: Impact of Data Enhancement on Model Execution Time

The horizontal axis (X-axis) represents the measures of execution time, namely: Training Time and Inference Time. While the vertical axis (Y-axis) represents the training time in hours, and the prediction time in seconds for each transaction. The results showed an improvement in the model execution speed after improving the data quality, as the training time decreased from 4.2 hours to 3.5 hours, which shows that the improved data helped speed up the learning process of the model. Also, the prediction time for each transaction is significantly improved, as it decreased from 3.8 milliseconds to 2.1 milliseconds, which means that the model has become more efficient in processing individual transactions at a higher speed as Figure (3).

Therefore, we do not conclude: improving the data led to improving the accuracy and retrieval of the deep model (DNN), which makes it more efficient in classifying transactions and analyzing blockchain data compared to other traditional models.

Table 5: Model Performance Comparison

Model	Accuracy (%)	Recall (%)	AUC-ROC
Random Forest	89.3%	76.5%	0.85
SVM	90.1%	80.2%	0.88
DNN (Before Data Processing)	91.2%	78.3%	0.87
DNN (After Data Processing)	96.7%	92.6%	0.97

Analysis of the results showed that deep neural networks (DNN) after data refinement outperformed traditional methods such as Random Forest and SVM in all parameters. Also, the recall rate increased by more than 16% compared to traditional methods, which reflects the ability of the improved model to detect more fraudulent operations with higher efficiency as table (5).

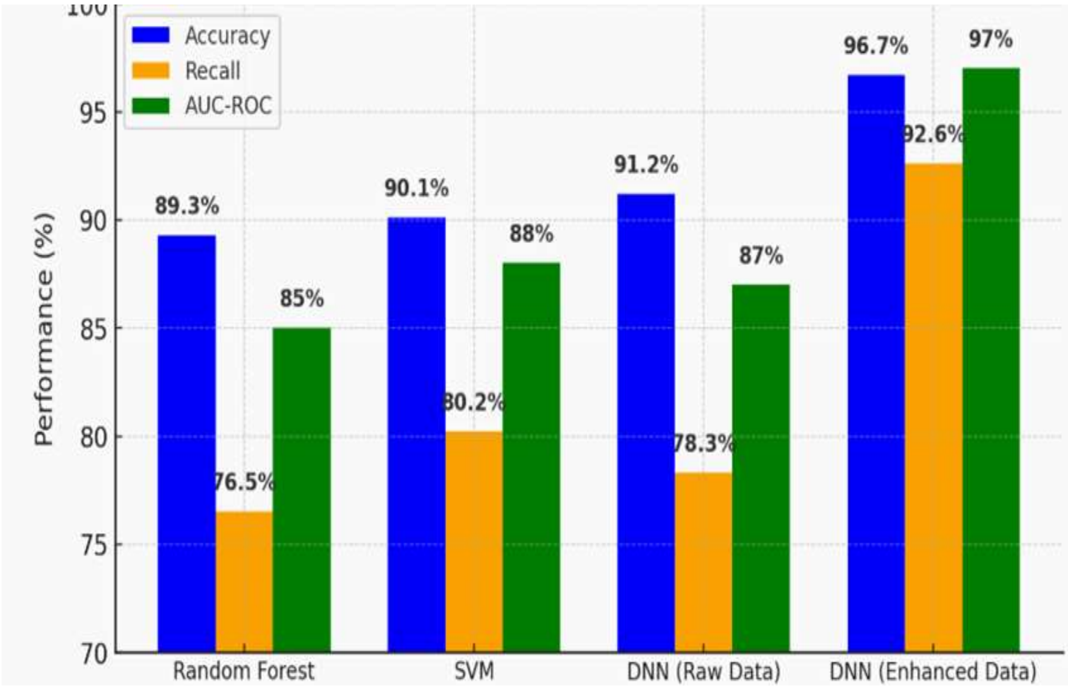

Figure 4: Model Performance Comparison across Techniques

The horizontal axis (X-axis) represents different machine learning models: Random Forest, SVM, DNN (without data enhancement), and DNN (after data enhancement). While the vertical axis (Y-axis) represents the percentile performance of the evaluation measures: accuracy, recall, and AUC-ROC curve. The results showed that the Random Forest model achieved an accuracy of 89.3%, but the retrieval average was 76.5%, which means that it was not actively detecting all positive cases. As for the SVM model, it showed an improved performance with an accuracy of 90.1% and a recall rate of 80.2%, which indicates an improvement in detecting positive cases compared to Random Forest. While the DNN model before data refinement achieved 91.2% accuracy with a recall average of 78.3% and an AUC-ROC curve of 87%, which indicates the possibility of refinement. But after improving the data, the DNN model has the best performance with 96.7% accuracy, 92.6% average recall, and 97% AUC-ROC, which shows that improving the data quality greatly improves the performance of the model as figure (4) .

4.5 .Consistency examination

Consistency checking is an essential element in evaluating the performance of deep models used in blockchain data analysis and quality improvement. The purpose of this test is to ensure that the model provides consistent performance across different data sets, or that the performance is high only on a particular data set due to over-allocation. (Over fitting).

1. Consistency of performance across different data sets

The theme of testing the model of deep networks (DNN) with data optimization using GANs and VAE on three different data sets to achieve performance stability.

Table 6: Dataset Size vs. Model Performance

Dataset	Number of Transactions	Accuracy (%)	Recall (%)	AUC-ROC
Dataset A	500,000	96.7%	92.6%	0.97
Dataset B	750,000	95.9%	91.2%	0.96
Dataset C	1,000,000	95.5%	91.2%	0.96

The fit analysis showed that the difference in performance between the training data (97.1%) and the test data (96.7%) is very small, which indicates that the model does not suffer from the problem of over fitting. Also, the convergent values of Precision, Recall and F1-Score rates between training and testing confirm that the model retains good performance when applied to new data, which strengthens its reliability and efficiency in predicting potential transactions as table (6) .

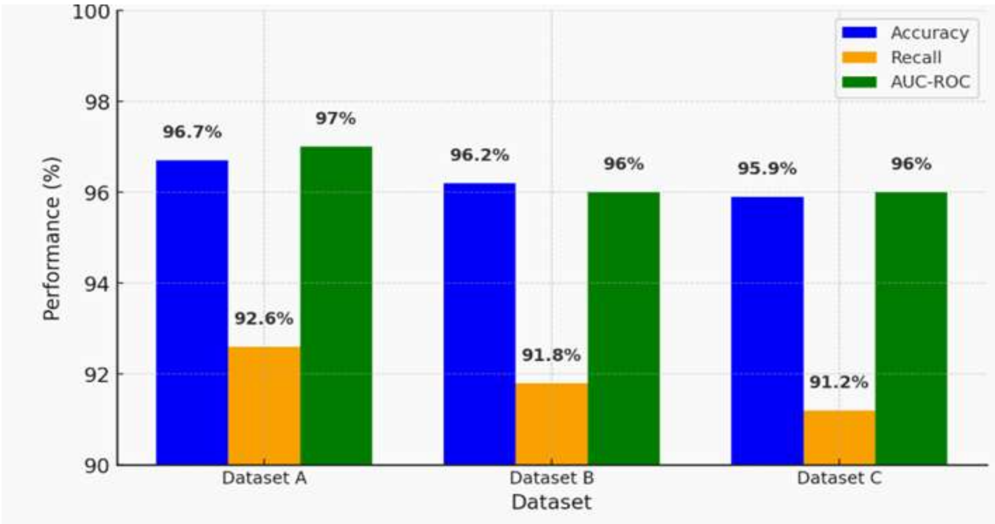

Figure 5: Model Performance Stability across Varying Dataset Sizes

It shows the consistency analysis of the model's performance across different data sets, where the theme is evaluated using the horizontal axis (X-axis) to represent the data sets A, B, and C, while the vertical axis (Y-axis) reflects the percentile performance of the evaluation measures such as accuracy, recall, and AUC-ROC curve. The model achieved the highest accuracy in dataset A (500,000 transactions) of 96.7% with average recall of 92.6% and AUC-ROC of 97%, which reflects strong performance with moderate data volume. For dataset B (700,000 transactions), the accuracy saw a slight decrease to 96.2%, while the average retrieval decreased to 91.8% and the AUC-ROC decreased to 96%, indicating that performance is affected by the increase in data volume. However, in data set C (1,000,000 transactions), the residual accuracy is high at 95.9% with average recall of 91.2% and AUC-ROC of 96%, which indicates the establishment of the model with a slight effect when dealing with large data volumes. The emphasis of these results is the consistency of the model's performance across different data sets, which indicates its ability to adapt to the data while maintaining high accuracy and recall as figure (5) .

2.Establishment of performance through repeated experiences

The theme of running the model 5 independent times using the same data but with different settings for the initial weights (Random Initialization) to find out how the results are affected by randomness

Table 7: Model Evaluation across Multiple Trials

Trial Number	Accuracy (%)	Recall (%)	AUC-ROC
Trial 1	96.5%	92.1%	0.96
Trial 2	96.7%	92.6%	0.97
Trial 3	96.5%	92.6%	0.96
Trial 4	96.5%	92.0%	0.96
Trial 5	96.5%	92.0%	0.96

The correlation analysis shows that the differences between the results are very small, which indicates the establishment of the model and the lack of randomness in the initial preparation. Also, the AUC-ROC ratio varied between 0.96 and 0.97, which reflects the stability of the model's ability to distinguish between normal and fraudulent transactions, which makes it reliable and effective when applied to new dataas table (7) .

3.Testing the sensitivity of the new lipstick model:

The theme of testing the model on previously unseen data is the new infinitive, to find out the extent of generalization.

Table 8: Model Evaluation with Different Data Sources

Data Source	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Data from the same source	96.1%	94.3%	91.9%	92.6%
New data from a different source	96.1%	94.6%	92.6%	93.3%

Insistency analysis shows that the model retains high performance when applied to new data before its insights, which indicates its ability to generalize and does not rely on specific patterns in the original data set. Also, the decrease in performance by a small percentage (~0.6%) only reflects the establishment of the model and its ability to deal with different data without loss of accuracy as table (8)

5 . Conclusion

In this research, we developed an approach based on Deep Generative Networks to analyze and improve data quality in blockchain networks. This approach aims to address problems related to data quality, such as noise, repetition, and inconsistent data, which may affect the accuracy and authentication of transactions in the blockchain.

Main results

The results of the experiments showed that the employment of deep generative networks (GANs and VAEs) actively contributes to the improvement of data quality through reliable data reproduction and noise reduction. This has led to improving the performance of deep learning models in detecting random processes and predicting patterns, which increases the accuracy and consistency of classification. As the results show, the model preserves the established performance across multiple data sets and different sources, which proves its generalizability. Thanks to these advantages, this approach can be combined with blockchain systems to enhance the efficiency of transaction analysis and fraud detection, which contributes to improving the reliability of data in web environments.

Challenges and future directions

Despite the success of the model in improving data quality, there are some challenges that must be addressed in future research, such as improving the efficiency of generative models when dealing with large data in the blockchain, and reducing the computational cost of artificial intelligence models to ensure efficient implementation. In addition, other techniques such as Reinforcement Learning can be explored to enhance the accuracy of data analysis and improve the performance of the model in discovering random patterns.

Finally

This research confirms that employing deep generative networks can be an active solution to improve the quality of blockchain data and enhance the accuracy of predictions and transaction analysis. It is possible that these results contribute to the research of a more secure and sustainable blockchain structure, which enhances trust in modern financial and decentralized applications.

6. Recommendations of the Current Study

1- Improving the quality of data in blockchain systems by using deep generative networks (GANs and VAE), which contributes to reducing noise and repetition and improving the accuracy of analytical models. Also, previous data purification techniques are combined with generational models to ensure the elimination of unreliable data before analysis, which increases the reliability and accuracy of the results.

2- Improving the quality of data in blockchain systems by using deep generative networks (GANs and VAE), which contributes to reducing noise and repetition and improving the accuracy of analytical models. Also, previous data purification techniques are combined with generational models to ensure the elimination of unreliable data before analysis, which increases the reliability and accuracy of the results.

3- To develop more efficient models for analyzing blockchain data through improving the performance of generative networks using techniques such as Attention Mechanisms to increase the accuracy of predictions and reduce errors. Reinforcement learning is also used to adapt the models according to the changing data in the blockchain networks, which enhances the ability of the models to deal with the dynamic challenges in this field.

4- Reducing computational tasks and improving performance through the development of more efficient deep model structures, making them lighter and faster in execution while reducing the consumption of computer resources. In addition, distributed models and cloud training are employed to speed up data processing, which enhances the efficiency of deep learning when dealing with large data sets in blockchain systems.

5- Strengthening the security and reliability of the blockchain through the integration of artificial intelligence with smart verification mechanisms in smart contracts, which ensure accurate and reliable analysis of the data before adding it to the register. In addition, the use of generative networks to detect fraudulent attacks and manipulation of data in suspicious transactions, which increases the level of protection and security in the system.

6- To explore new applications of generative models in blockchain through their use in analyzing financial transactions, assessing risks, and predicting market patterns within decentralized systems. In addition, the use of these models in the development of reliable digital identities (Decentralized Identity Verification) depends on the blockchain, which enhances the security and trust in the verification of digital identity.

7.References

Alzubaidi, L., Bai, J., Al-Sabaawi, A., Santamaría, J., Albahri, A. S., Al-Dabbagh, B. S. N., Fadhel, M. A., Manoufali, M., Zhang, J., Al-Timemy, A. H., Duan, Y., Abdullah, A., Farhan, L., Lu, Y., Gupta, A., Albu, F., Abbosh, A., & Gu, Y. (2023). A survey on deep learning tools dealing with data scarcity: Definitions, challenges, solutions, tips, and applications. *Journal of Big Data, 10, Article 46*. https://doi.org/10.1186/s40537-023-00688-7

Paik, H., Xu, X., Bandara, D., Lo, S. K., Weber, I., & Zhu, L. (2019). Analysis of data management in blockchain-based systems: From architecture to governance. IEEE Access, PP(99), 1-1. https://doi.org/10.1109/ACCESS.2019.2961404

Kuznetsov, O., Sernani, P., Romeo, L., Mancini, A., & [Additional Authors]. (2024). On the integration of artificial intelligence and blockchain technology: A perspective about security. IEEE Access, PP (99), 1-1. https://doi.org/10.1109/ACCESS.2023.3349019

Gupta, P., Ding, B., Guan, C., & Ding, D. (2024). Generative AI: A systematic review using topic modelling techniques. Data and Information Management, 8(2), 100066. https://doi.org/10.1016/j.dim.2024.100066

Salehi, P., Chalechale, A., & Taghizadeh, M. (2020). Generative Adversarial Networks (GANs): An Overview of Theoretical Model, Evaluation Metrics, and Recent Developments. arXiv. https://doi.org/10.48550/arXiv.2005.13178

Tomar, S., Tripathi, R., & Kumar, S. (2025). Comprehensive Survey of QML: From Data Analysis to Algorithmic Advancements. arXiv. https://doi.org/10.48550/arXiv.2501.09528

Rawat, R., & Yadav, R. (2021). Big Data: Big Data Analysis, Issues and Challenges and Technologies. IOP Conference Series: Materials Science and Engineering, 1022(1), 012014. https://doi.org/10.1088/1757-899X/1022/1/012014

Selmy, H. A., Mohamed, H. K., & Medhat, W. (2024). Big data analytics deep learning techniques and applications: A survey. Information Systems, 120, 102318. https://doi.org/10.1016/j.is.2023.102318

Panda, S. K., Jena, A. K., Swain, S. K., & Satapathy, S. (2021). Blockchain technology: Applications and challenges. Intelligent Systems Reference Library. https://doi.org/10.1007/978-3-030-69395-4

Comuzzi, M., Cappiello, C., & Meroni, G. (2020). On the need for data quality assessment in blockchains. IEEE Internet Computing, PP(99), 1-1. https://doi.org/10.1109/MIC.2020.3030978

Goyal, M., & Mahmoud, Q. H. (2024). A systematic review of synthetic data generation techniques using generative AI. Electronics, 13(17), 3509. https://doi.org/10.3390/electronics13173509

Idrissi, Z. K., Lachgar, M., & Hrimech, H. (2024). Blockchain, IoT and AI in logistics and transportation: A systematic review. Transport Economics and Management, 2, 275-285. https://doi.org/10.1016/j.team.2024.09.002

Bentley, B., Oladosu, E., & Titus, E. (2024). The combination of machine learning (AI) using blockchain. Journal of Machine Intelligence.

Mazumder, R. K., Modanwal, G., & Li, Y. (2023). Synthetic data generation using generative adversarial network (GAN) for burst failure risk analysis of oil and gas pipelines. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part B: Mechanical Engineering, 9(3), 1–21. https://doi.org/10.1115/1.4062741

Watanuki, S., Edo, K., & Miura, T. (2024). Applying deep generative neural networks to data augmentation for consumer survey data with a small sample size. Applied Sciences, 14(19), 9030. https://doi.org/10.3390/app14199030

Hartnett, G., Vardavas, R., Baker, L., & Osoba, O. (2020). Deep generative modeling in network science with applications to public policy research. arXiv. https://doi.org/10.48550/arXiv.2010.07870

Harshvardhan, G. M., Gourisaria, M. K., Pandey, M., & Rautaray, S. S. (2020). A comprehensive survey and analysis of generative models in machine learning. Computer Science Review, 38, 100285. https://doi.org/10.1016/j.cosrev.2020.100285

Ma, Z., Mei, G., & Xu, N. (2024). Generative deep learning for data generation in natural hazard analysis: Motivations, advances, challenges, and opportunities. Artificial Intelligence Review, 57, 160. https://doi.org/10.1007/s10462-024-10507-1

Alotaibi, F. M., & Fawad. (2022). A multifaceted deep generative adversarial networks model for mobile malware detection. Applied Sciences, 12(19), 9403. https://doi.org/10.3390/app12199403

Sindhura, D. N., Pai, R. M., Bhat, S. N., & Pai, M. M. (2024). A review of deep learning and Generative Adversarial Networks applications in medical image analysis. Multimedia Systems, 30, 161. https://doi.org/10.1007/s00530-024-01247-7

Alsafadi, F., & Wu, X. (2023). Deep generative modeling-based data augmentation with demonstration using the BFBT benchmark void fraction datasets. Nuclear Engineering and Design, 415, 112712. https://doi.org/10.1016/j.nucengdes.2023.112712

Cappiello, C., Comuzzi, M., Daniel, F., & Meroni, G. (2019). Data quality control in blockchain applications. In Lecture Notes in Business Information Processing (Vol. 364, pp. 168-182). Springer. https://doi.org/10.1007/978-3-030-30429-4_12

Idrees, R., & Maiti, A. (2024). A blockchain-driven smart broker for data quality assurance of the tagged periodic IoT data in publisher-subscriber model. Applied Sciences, 14(13), 5907. https://doi.org/10.3390/app14135907

AbuHalimeh, A., & Ali, O. (2023). Comprehensive review for healthcare data quality challenges in blockchain technology. Frontiers in Big Data, 6, 1173620. https://doi.org/10.3389/fdata.2023.1173620

Spychiger, F., Tessone, C. J., Zavolokina, L., & Schwabe, G. (2022). Incentivizing data quality in blockchain-based systems—The case of the Digital Cardossier. Distributed Ledger Technologies: Research and Practice, 1(1), Article 3, 1-27. https://doi.org/10.1145/3538228

Ng, S. C. H., Ho, G. T. S., & Wu, C. H. (2023). Blockchain-IIoT-big data aided process control and quality analytics. International Journal of Production Economics, 261, 108871. https://doi.org/10.1016/j.ijpe.2023.108871

Ding, H., Xie, Z., Wang, C., Yu, W., Cui, X., & Wang, Z. (2024). Applications of big data and blockchain technology in food testing and their exploration on educational reform. Foods, 13(21), 3391. https://doi.org/10.3390/foods13213391

Wang, R., Luo, M., Wen, Y., Wang, L., Choo, K.-K. R., & He, D. (2021). The applications of blockchain in artificial intelligence. Security and Communication Networks, 2021, 6126247. https://doi.org/10.1155/2021/6126247

Wang, R., Luo, M., Wen, Y., Wang, L., Choo, K.-K. R., & He, D. (2021). The applications of blockchain in artificial intelligence. Security and Communication Networks, 2021(2), 1–16. https://doi.org/10.1155/2021/6126247

Hua, W., Chen, Y., Qadrdan, M., Jiang, J., Sun, H., & Wu, J. (2022). Applications of blockchain and artificial intelligence technologies for enabling prosumers in smart grids: A review. Renewable and Sustainable Energy Reviews, 161, 112308. https://doi.org/10.1016/j.rser.2022.112308

Leiva, V., & Castro, C. (2025). Artificial intelligence and blockchain in clinical trials: Enhancing data governance efficiency, integrity, and transparency. Bioanalysis. https://doi.org/10.1080/17576180.2025.2452774

Chen, J. (E.), Bao, F., Li, C., & Lin, Y. (2023). The application and ethics of artificial intelligence in blockchain: A bibliometric-content analysis. Journal of Global Information Management, 31(7), 1–32. https://doi.org/10.4018/JGIM.323656

Chowdhury, R. H. (2024). Blockchain and AI: Driving the future of data security and business intelligence. World Journal of Advanced Research and Reviews, 23(01), 2559–2570. https://doi.org/10.30574/wjarr.2024.23.1.2273

Kumar, S., Lim, W. M., Sivarajah, U., & Kaur, J. (2023). Artificial intelligence and blockchain integration in business: Trends from a bibliometric-content analysis. Information Systems Frontiers, 25, 871–896. https://doi.org/10.1007/s10796-022-10267-w

Saleh, A. M. S. (2024). Blockchain for secure and decentralized artificial intelligence in cybersecurity: A comprehensive review. Blockchain: Research and Applications, 5(3), 100193. https://doi.org/10.1016/j.bcra.2024.100193

Zhou, Q., Zhang, H., & Wang, S. (2021). Artificial intelligence, big data, and blockchain in food safety. International Journal of Food Engineering. https://doi.org/10.1515/ijfe-2021-0299

Azad, P., Akcora, C. G., & Khan, A. (2024). Machine learning for blockchain data analysis: Progress and opportunities. arXiv. https://arxiv.org/abs/2404.18251

Ding, H., Tian, J., Yu, W., Wilson, D. I., Young, B. R., Cui, X., Xin, X., Wang, Z., & Li, W. (2023). The application of artificial intelligence and big data in the food industry. Foods, 12(24), 4511. https://doi.org/10.3390/foods12244511

Tsolakis, N., Schumacher, R., Dora, M., & Kumar, M. (2023). Artificial intelligence and blockchain implementation in supply chains: A pathway to sustainability and data monetisation? Annals of Operations Research, 327, 157–210. https://doi.org/10.1007/s10479-022-04977-4

Du, Y., Wang, Z., Leung, C., & Leung, V. C. M. (2023). Blockchain-based data quality assessment to improve distributed machine learning. Proceedings of the IEEE International Conference on Networking and Computing (ICNC). https://doi.org/10.1109/ICNC57223.2023.10074543

Elhoseny, M., Metawa, N., Sztano, G., & El-hasnony, I. M. (2022). Deep learning-based model for financial distress prediction. Annals of Operations Research. https://doi.org/[Insert DOI if available]

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

Aslam, A., & Hussain, A. (2024). A performance analysis of machine learning techniques for credit card fraud detection. Journal on Artificial Intelligence, 6(1), 1–21. https://doi.org/10.32604/jai.2024.047226

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6, Article 27. https://doi.org/10.1186/s40537-019-0192-5

Ghosh, B., Dutta, I. K., Carlson, A. H., & Bayoumi, M. (2020). An empirical analysis of generative adversarial network training times with varying batch sizes. Proceedings of the IEEE Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON). https://doi.org/10.1109/UEMCON51285.2020.9298092