Library Progress International
Vol.44, No.6, Jul-Dec 2024: P.1376-1391

Print version ISSN 0970 1052
Online version ISSN 2320 317X

Original Article

Available online at www.bpasjournals.com

Enhancing Digital Payment Security with AI-Driven Biometric Authentication and Big Data Analytics: A Scalable Approach for Secure Remote Commerce

Somnath Mondal [1], Sujan Das [2], Shib Shankar Golder [3]

Solution Data Architect, EY [1] , Solution Architect – Data, AI & Analytics, Deloitte [2], Senior Solution Architect - Data, AI & Analytics, EY [3]
Motilal Nehru National Institute of Technology, India [1], University of Illinois Urbana Champaign, IL, USA [2], University of Texas at Austin, USA [3]
somnath.mondal@live.com [1], sujandas1985@gmail.com [2] , ece.351@gmail.com [3]

How to cite this article: Somnath Mondal , Sujan Das , Shib Shankar Golder (2024). Enhancing Digital Payment Security with AI-Driven Biometric Authentication and Big Data Analytics: A Scalable Approach for Secure Remote Commerce. *Library Progress International,* 44(6), 1376-1391

Abstract- This conducted research investigates the application of machine learning models to enhance fraud detection in financial transactions. The research work at hand uses a Gradient Boosting Classifier, Random Forest Classifier, and Support Vector Machine models with the main aim of improving fraud detection. The set contains records of fraudulent and non-fraudulent transactions relating to the transaction amount, means of payment, and biometric data. Pre-processing steps may include changing categorical variables into numeric, mapping the target variable into binary values, and scaling features. Each model must be trained and scored on metrics about accuracy, precision, recall, and F1 score. A Gradient Boosting Classifier is an algorithm where performance has been polished iteratively by updating the weights over the misclassified instances. The Random Forest Classifier improves the decision of weak classifiers by aggregation. SVM finds the best hyperplane to separate fraudulent and non-fraudulent transactions. The performances through these models are compared to searching for the best method. The results reveal strengths and weaknesses of each model and draw out insights on how effective these are in real fraud detection cases.

Keywords: Secure Remote Commerce, Digital Wallet, Fraud Detection, Machine Learning, Big Data , Gradient Boosting, Random Forest, Support Vector Machine, Classification Accuracy, Financial Transactions

I: Introduction

A. Research background

Digital payment systems can be predicted to change the global financial transaction landscape. The recent growth rate has been very steep, and it has been propelled by increased smartphone penetration, access to the Internet, and the ease of use of digital platforms. While digital wallets, mobile banking apps, and online payment gateways take over the way through which consumers do daily transactions, it becomes large in volume concerning digital payments across the globe [1]. Therefore, securing these transactions has become quite imperative with their increasing adoption rate for digital payment systems.

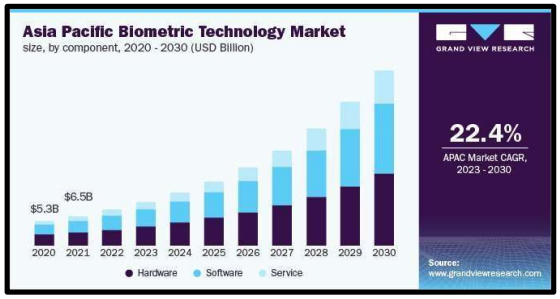


Fig. 1: Biometric Technology Market Size and Report

Integrity, confidentiality, and authenticity of transactions are important in consumers' trust and financial losses. Strong security measures protect sensitive data from unauthorized access. Advanced technologies such as Artificial Intelligence, big data analytics, and biometric authentication are now under the limelight to solve these issues [2]. The AI-driven

algorithms analyze the patterns of transactions in real time to detect and prevent fraudulent activities. Big Data goes a long way in continuously improving these algorithms by processing huge amounts of data and finding out emerging threats.

B. Research aim and objectives

Aim

AI-driven biometric authentication combined with Big Data analytics can be helpful in digital payments for increasing security, efficiency, and user experience.

Objectives

- Current challenges in digital payment security are investigated in this project.
- To assess how AI can identify and block fraud in real time.
- To study henceforth tries to establish whether biometric authentication can ensure secure transactions.
- To design a secure, highly scalable digital payment system.

C. Research Rationale

What is the issue

The financial world is currently focusing more on security because digital transactions have been growing at exponential levels. The cases of cyber-attacks, fraud, and data breaches have increased astronomically as consumers and businesses surge towards online payment methods. Sophisticated methods applied by cybercriminals overshadow the traditional mechanisms of security—for example, passwords, PINs, and basic encryption.

Why is it an issue

The group said that one security breach mean huge financial losses to individual users, not only to the businesses and financial institutions in place. Far from that, these breaches must expose highly sensitive personal data at stake where these destroy consumer trust with long-term reputational injuries. The more integrated digital payment systems are in the everyday lives of people, the higher the risks these can pose. It is, therefore, important to look into these vulnerabilities with focus and at speed.

Why is it an issue now

The problem has come to the fore now because digital payment is being adopted at a fast pace across the world, accelerated in no small part by the COVID-19 pandemic and increasing preference for cashless transactions. This must increase the attack surface for cybercriminals to exploit latent weaknesses manifold. As such, state-of-the-art security solutions, such as AI-powered biometric authentication and Big Data analytics, are required to provide adaptability and robustness against the threats of the modern age.

II: Literature Review

A. Evolution of Digital Payment Systems

Digital payment systems have emerged as a revolutionary tool that has undergone a revolution and disclosed a new era of payment systems across the world due to innovation in technology and the customers' shift in their preferences [3]. The process started in the mid-twentieth century when credit and debit cards entered the market and provided people an opportunity to perform cashless transactions. These plastic cards have been instrumental in providing the necessary impetus to change the way consumers must go about their business – without cash. The last decades of the twentieth century have been marked by the use of electronic banking, and, to a certain extent, the first hints at the use of the Internet for payments. With the development of the Internet in the 1990s, these payment systems began appearing – and online banking platforms have been created. These platforms enabled people to pay for goods and services, pay their bills, and perform numerous other banking transactions online, thus playing a major in the evolution of online banking [4]. This are around the early 2000s when the introduction of the use of online payment gateway like PayPal linked consumers and business entities that needed to have online transactions.

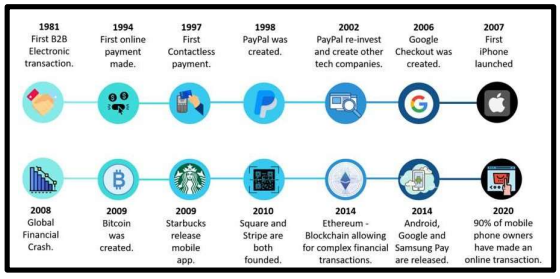


Fig. 2: Evolution of Digital Payment Systems

These also marked the emergence of e-commerce which also created the demand for efficient payment mechanisms in the digital space. With the evolution of the facets of mobile technology, new payment platforms such as Apple Pay, Google

Wallet, and Samsung Pay have been created and enabled users to make payments directly using their mobile devices. AI, machine learning, and biometric authentication have become integrated in the last couple of years to enhance the digital payment system [5]. Fighting fraud and increasing security measures that have been due to AI technologies such as fraud detection and biometric verification methods like fingerprints, facial recognition, and voice recognition have increased the security and efficiency of online transactions. Also, big data has been useful in enhancing the personalization of payments and risk control aspects. The development of payment systems is dynamic, as new technological developments such as the blockchain and cryptocurrencies, advance the way transactions are done around the world. These advancements still are relevant, kicking into the future for secure, efficient, and scalable values of payments as an extension of the digital world [6].

B. Security Challenges in Digital Payments

Currently, there is a wide use of digital payment systems and like all other advanced things, these have also been subjected to insecurity. The first of these is fraud risk, which encompasses identity theft, phishing, and account takeover. Hackers target such sites to have access to credit details and other personal identification numbers (PIN) which these use to make purchases fraudulently [7]. The other major security issues are data breaches where millions of consumers' data are compromised because of poor security protocols put in place. Said breaches can be gained through hacking, using insiders, or lack of strong encryption methods, and the consequences are grave fines for businesses and loss of consumer trust. Developing mobile payments has also brought new threats; especially those that are associated with the malware attack on mobiles and other devices. These devices tend to contain consumer payment data and, as such, are often invaded by hackers.

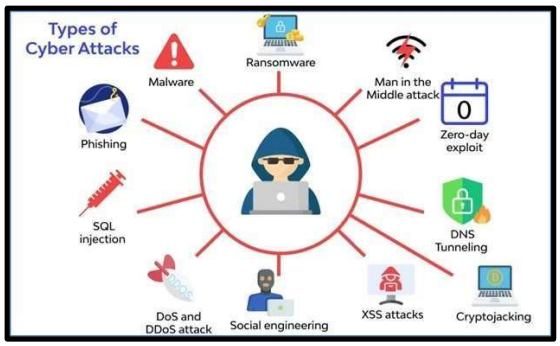


Fig. 3: Security Challenges in Digital Payments

Another security challenge is the mightiness of payment systems that involve banks, payment processors, and third-party vendors. Each link in this chain can be regarded as a weak link that adversaries can take advantage of [8]. Also, these globalized transactions including e-commerce and cross-border transactions expose them to payment fraud menace since criminals can launch attacks from regions that do not have rigid legislation on cybersecurity. Another emergent threat is the complexity of cyber threats, which more and more incorporate artificial intelligence and machine learning to evade known security mechanisms. AI can be used to make phishing attacks more precise, to fake identities through deep fakes, or to seize on new and unguarded payment systems' vulnerabilities. Thus, with such risks in mind, this has led to the need to embrace sophisticated security mechanisms like Artificial Intelligence fraud detection, Biometric authentication, and sophisticated encryption protocols that must tackle the emerging and sundry types of security threats that surround digital payment systems [9].

C. Role of Artificial Intelligence in Enhancing Payment Security

Payment security is a crucial factor in any payment system, and AI is a crucial tool and weapon against modern threats to payments by providing state-of-the-art techniques against new and rapidly developing threats. Using AI in financial portfolios, institutions, and payment processors may enhance fraud detection and minimization of fraud activities, thus enhancing the security of consumer and business transactions [10]. The use of AI in payment security can also be seen in the identification of fraud as one of its major uses. To help explain, it is crucial to understand that the anti-malware traditional rule-based systems directly fail to respond to the dynamic and highly complicated process of cyber threats. While it is not very effective in making fast and crisp decisions, it can swiftly analyze huge volumes of transactional data, recognize patterns, and possibly detect unusual behaviour, that must be linked to fraudulent actions. These therefore can learn from previous experience and their performance gets better with time, and thus can be used to predict and deter fraud more effectively. AI also provides for secure payment by providing biometric forms of payment systems [11].

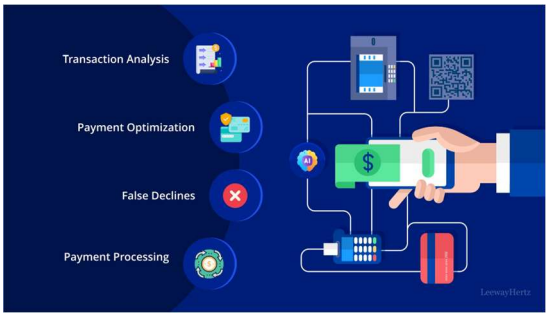


Fig. 4: Role of Artificial Intelligence in Enhancing Payment Security

Tools like face or voice identification, and fingerprint scanning are being included as tools for user identification while making payments. Such AI-based biometric systems provide enhanced security compared to passwords and PINs because these are unique to every person and therefore cannot easily be forged [12]. Apart from fraud detection and authentication, the other areas in which AI is applied include risk assessment and management. The virtue of having an AI solution is that it can compute risk using a wide range of characteristics of the user, their previous transactions, or even information about the device being used to perform the transaction. This makes it easy for pay processors to make sound decisions such as reporting suspicious activities in a certain account to follow up on or possibly rejecting high-risk transactions altogether. Thirdly, AI can also be utilized for supervising the payment system all through and supplying alerts and detection for security personnel. The concept allows organizations to stay ahead of new threats as it can propose efficient methods of how to address security threats thus reducing the effects of break-ins [13].

D. Biometric Authentication in Digital Payments

Therefore biometric authentication has come out as one of the important technologies that has developed the security of digital payments. Being based on fingerprints, face, iris, voice, and biometrics, as means of authentication, is more effective as compared to passwords or PINs. This technology has been widely recognized in the financial area, especially because it offers a higher degree of security compared to traditional methods for protecting against fraud and unauthorized access [14]. However, it has been also underlined that one of its main benefits is the fact that biometric authentication can make the verification more accurate. Biometric data is not subject to be lost, forgotten, or hacked because the data is always with an individual and close to impossible to fake. This makes it extremely difficult for hackers to compromise the system or perform fraudulent transactions and identity thefts. For example, in biometrics, it is common for a person to unlock their smartphone or authenticate the transaction via fingerprint scan which is both secure and easy to employ. Another widespread type of biometric is the facial recognition method, which is actively used in smartphones and other gadgets.

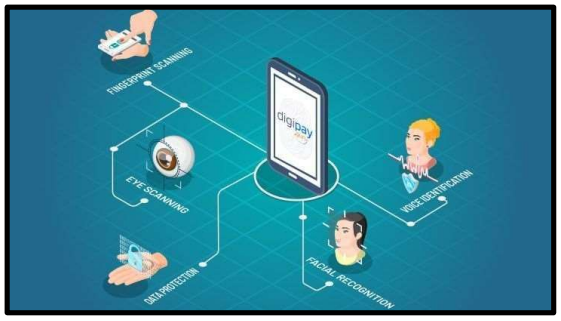


Fig. 5: Biometric Authentication in Digital Payments

These include scans and analyses of the various features on the face to authenticate the user to enable a payment to be made. This method has received favour because it is least intrusive and still able to authenticate users under varying amounts of lighting. Similarly, voice recognition is being embraced for such purposes as authenticating users by their voice, making security a strong point in voice-payment-activated systems [15]. Another advantage of biometric authentication is that it makes payment much easier from the user's perspective. The users do not have to remember ridiculous password combinations and keep pockets full of authentication tokens since the only key to the transaction required is one's biological characteristics. It also enhances the chances of using digital payment systems among the users of the system as it becomes easier for them to make payments [16]. However, biometric authentication has its limitations and these include; privacy, storage space, and misuse of the biometric data. To enhance these problems, strong methods of encryption and security formats that may protect biometric data are exceedingly important.

E. Literature Gap

As evident from the literature there is ample research done on how the integration of AI and biometric authentication can advance the security of digital payment but there is a lack of research done on how AI-enabled biometric authentication for scalable, real-time fraud detection using big data analysis in remote commerce. Moreover, little is known about how well these technologies can be synchronized in various kinds of payments to have a symmetric level of protection, which must also be investigated further.

III: Methodology

A. Research Design

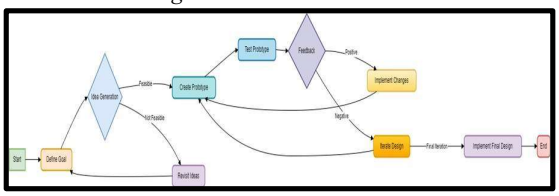


Fig. 6: Flowchart

This targets fraud detection in financial transactions using machine learning models. The research design has been built around the analysis of a transaction dataset in search of fraudulent activities. The methodology entails a few key stages: data exploration, data preprocessing, development, and model evaluation. First, in data exploration, there has been visualization and an understanding of the dataset with transaction records, including such features as the amount of the transaction, means of payment, and device type [17]. This may identify trends in the data and anomalies. It is then followed by the preprocessing of data, where categorical variables are converted to numeric format using one-hot encoding in readiness for fitting machine learning models.

Component	**Description**
Objective	Enhance fraud detection using machine learning.
Dataset	Fraud_Detection.csv with transaction records.
Data Exploration	Visualize distributions and patterns.
Preprocessing	Encode variables, split data, and normalize features.
Models	Gradient Boosting, Random Forest, SVM.
Evaluation	Accuracy, Precision, Recall, F1-Score.
Comparison	Compare model performance using accuracy.
Tools	Python, Pandas, Scikit-learn, Matplotlib, Seaborn.

Table 1: Research Design

The target variable is binary-coded, as it deals with fraud detection. The steps that follow include splitting the dataset into a training and testing set and normalization of features using StandardScaler, also known as feature scaling, to enhance model performance [18]. In the section where the models are developed, training for three machine learning classifiers is done: Gradient Boosting, Random Forest, and Support Vector Machine. The reasons for this choice are that the approaches these three models take toward classification are completely different, and all of them can handle complex patterns in the data.

$$
\begin{aligned}
&\text{“}F(x)=\sum m=1M\alpha mhm(x)\\
&Fm+1(x)=Fm(x)+\eta \cdot hm(x)\\
&F(x)=T1\sum t=1\ Tht(x)\\
&H(Y)=-\sum i=1kpilog2(pi)\\
&f(x)=sign(w\cdot x+b)\\
&f(x)=sign(\sum i=1N\alpha iyiK(xi,x)+b)\\
&Yi(w\cdot xi+b)\geq 1 \text{ for all } i\\
&\sum i=1N\alpha i-21\sum i=1N\sum j=1N\alpha i\alpha jyiyjK(xi,xj)\\
&\sum i=1N\alpha iyi=0\text{”}
\end{aligned}
$$

B. Data Collection

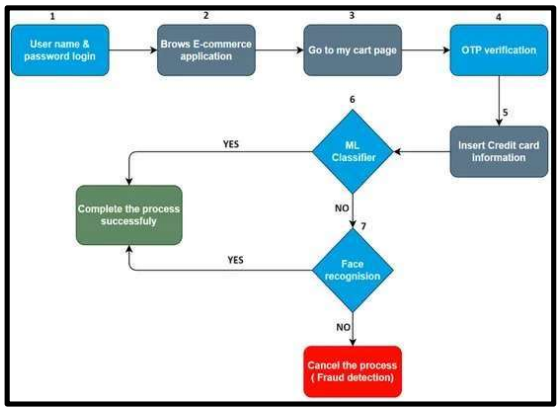


Fig. 7: Framework Working Flowchart

Used in this research is a dataset called Fraud_Detection.csv, which includes full transaction records with a few key features: transaction amount, means of payment, type of device, and biometric data. All these variables are very important in analyzing transaction patterns for establishing fraudulent activities. The dataset has a target variable indicating whether a transaction is classified as fraudulent or not [19]. Downloading a dataset is a part of data collection, which has been already pre-collected and, hence, assumed to have been captured already and is ready and appropriate for analysis. This dataset is designed to provide a representative sample of financial transactions to train and test machine-learning models in fraud detection. Also included is the transaction amount, which conveys information on the value of transactions. The payment method and device type give context about the tools used for transactions. Biometric data supplies an added layer of security information that must improve this capability in fraud detection.

$$
\begin{aligned}
&\text{“}Accuracy=Total\ InstancesTrue\ Positives+True\ Negatives\\
&Precision=True\ Positives+False\ PositivesTrue\ Positives\\
&Recall=True\ Positives+False\ NegativesTrue\ Positives\\
&F1\ Score=2\times Precision+RecallPrecision\times Recall\\
&MSE=n1\sum i=1n(yi-y\hat{}i)2\\
&P(Y=1|X)=1+e-(\beta 0+\beta 1X)1\\
&Loss=21\sum i=1n(yi-y\hat{}i)2\text{”}
\end{aligned}
$$

C. Data Preprocessing

Fig. 8: Data Preprocessing

Data preprocessing is one of the most important steps in preparing a dataset for modelling in machine learning. Several major operations have been done in the preprocessing phase to make the given data suitable for analysis. First of all, variables such as PaymentMethod, BiometricUsed, Location, DeviceType, and BigDataPatternDetected have been converted into numeric format by one-hot encoding. This transformation is imperative since machine learning algorithms require numerical input to process the categorical data efficiently. It has been then made sure that the target variable has been converted into binary variables, FraudDetected, so that classification must become easier.

Variable	**Original**	**Transformed**
payment method	Credit, Debit	PaymentMethod_Credit, PaymentMethod_Debit
BiometricUsed	Fingerprint, Iris	BiometricUsed_Fingerprint, BiometricUsed_Iris
FraudDetected	Fraud, No Fraud	1 (Fraud), 0 (No Fraud)

Table 2: Example Data Transformation

In this respect, the target variables 'Fraud' and 'No Fraud' have been encoded as 1 and 0 respectively. In these steps, prepare it to work with classification models for the dataset. Then, the dataset is into features and target variables. In this case, these pursued the independent variables to the target variable. The 'TransactionID,' 'UserID,' and the 'TimeOfTransaction' have been the variables without relevance to the model. Hence, it only remains with the column names consisting of the predictive attributes [20]. The data has been haphazardly divided into 70-30 by training and testing.

$$\text{“One-HotCi=[0}\cdots\text{1}\cdots\text{0]}$$
$$y=\{10 \text{ if Fraud If No Fraud}$$
$$X_{scaled}=\sigma X-\mu$$
$$\text{Training set: (Xtrain, train)}$$
$$\text{Testing set: (Xtest, test)”}$$

D. Model Development

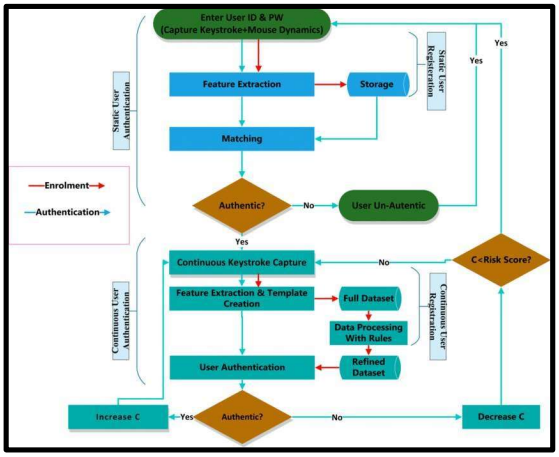


Fig. 9: Authentication Work Flowchart

The three developed machine learning models have been Gradient Boosting Classifiers, improving prediction accuracy by combining many weak learners into one strong predictive model. This is a system where the misclassified instances are given a certain weight and the final tune on the model's performance is realized by iterating the adjustment of weights to misclassified instances. Since the Gradient Boosting Classifier is designed for the correction of misclassifications or errors from the previous iteration, it is quite effective at capturing fraudulent transactions.

Model 1: Gradient Boosting Classifier
Initialize Gradient Boosting Classifier
Train model on X_train and y_train
Predict on X_test
Evaluate performance (accuracy, precision, recall, F1-score)

The second model to be used is a Random Forest Classifier, which works by making an ensemble of a large number of decision trees. Each tree makes a prediction, then the output is taken by aggregating the results across all trees. This mayincrease the accuracy and robustness of the classification, reducing overfitting and increasing model stability. This method has come in very handy while dealing with complex datasets having a high dimensionality of features.

Model 2: Random Forest Classifier
Initialize Random Forest Classifier
Train model on X_train and y_train
Predict on X_test
Evaluate performance (accuracy, precision, recall, F1-score)

Third is SVC, which attempts to construct the best hyperplane to separate these two classes: fraudulent from fraudless transactions. An SVC mayestablish a clear decision boundary by finding a maximum margin intended between the classes.

Model 3: Support Vector Machine (SVM)
Initialize Support Vector Machine Classifier
Train model on X_train and y_train
Predict on X_test
Evaluate performance (accuracy, precision, recall, F1-score)

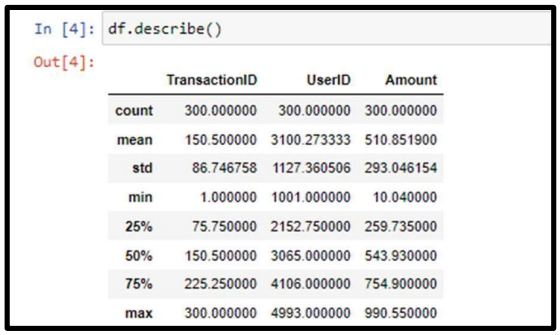
IV: Results And Discussion

A. Result

	TransactionID	UserID	Amount	PaymentMethod	BiometricUsed	Location	DeviceType	TimeOfTransaction	BigDataPatternDetected	FraudDetected
0	1	4174	492.01	Debit Card	Fingerprint	Other	Other	15-08-23	Yes	No Fraud
1	2	4507	985.72	Digital Wallet	Facial Recognition	EU	Mobile	08-02-23	Yes	No Fraud
2	3	1860	245.84	Digital Wallet	Fingerprint	Other	Other	29-10-23	Yes	Fraud
3	4	2294	673.77	Debit Card	Fingerprint	Other	Desktop	29-01-23	No	Fraud
4	5	2130	762.81	Digital Wallet	Fingerprint	Asia	Desktop	14-11-23	No	No Fraud
5	6	2095	241.45	Credit Card	NaN	Asia	Mobile	02-09-23	Yes	No Fraud
6	7	4772	729.58	Credit Card	Facial Recognition	US	Tablet	24-08-23	No	Fraud
7	8	4092	370.94	Debit Card	Facial Recognition	US	Mobile	04-07-23	Yes	No Fraud
8	9	2638	634.14	Bank Transfer	Facial Recognition	Other	Other	29-09-23	No	No Fraud
9	10	3169	635.36	Debit Card	Fingerprint	EU	Tablet	24-08-23	Yes	No Fraud
10	11	1466	538.10	Debit Card	Facial Recognition	Asia	Desktop	27-08-23	No	No Fraud
11	12	2238	94.84	Bank Transfer	Facial Recognition	Asia	Tablet	23-09-23	Yes	No Fraud
12	13	1330	836.13	Bank Transfer	Fingerprint	Other	Desktop	25-12-23	Yes	No Fraud
13	14	2482	324.18	Bank Transfer	Voice Recognition	US	Tablet	06-12-23	No	No Fraud
14	15	3135	190.59	Credit Card	Fingerprint	EU	Tablet	06-03-23	Yes	Fraud

Fig 10: Display the first few rows

This figure shows the initial rows of the dataset, including fields like TransactionID, UserID, Amount, PaymentMethod, and FraudDetected. The dataset, consisting of 300 transactions, captures various details of each payment, such as the method used, biometric authentication, and device type. For example, a user with ID 4174 made a $492.01 purchase via a Debit Card, which is not fraudulent.



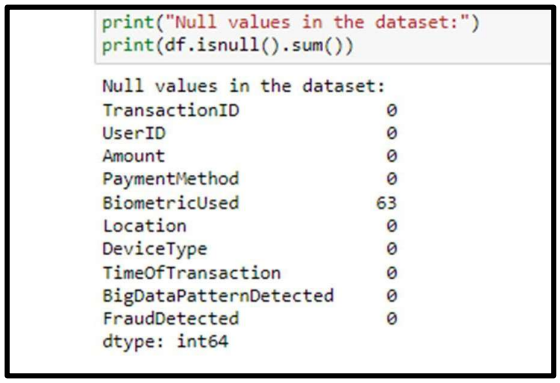
```
In [4]: df.describe()
```

Out[4]:

	TransactionID	UserID	Amount
count	300.000000	300.000000	300.000000
mean	150.500000	3100.273333	510.851900
std	86.746758	1127.360506	293.046154
min	1.000000	1001.000000	10.040000
25%	75.750000	2152.750000	259.735000
50%	150.500000	3065.000000	543.930000
75%	225.250000	4106.000000	754.900000
max	300.000000	4993.000000	990.550000

Fig 11: Description of the dataset

This figure summarises the dataset's statistical properties. The Amount field, with a mean of $510.85 and a maximum of $990.55, shows significant variability. Key statistics like mean, standard deviation and percentiles help understand the distribution of transaction amounts. The dataset's detailed description aids in identifying trends and anomalies, which are crucial for developing effective fraud detection models.



```
print("Null values in the dataset:")
print(df.isnull().sum())
```

```
Null values in the dataset:
TransactionID              0
UserID                     0
Amount                     0
PaymentMethod              0
BiometricUsed             63
Location                   0
DeviceType                 0
TimeOfTransaction          0
BigDataPatternDetected     0
FraudDetected              0
dtype: int64
```

Fig 12: Null values Check

The above figure shows the null values or missing data checked into the dataset. It is noticed that in the column BiometricUsed, there are 63 missing values. All the other fields, like TransactionID, Amount, and FraudDetected, are complete. Finding out the missing values is important in data quality and especially in columns like BiometricUsed, which might impact the analysis of fraud detection and model accuracy.

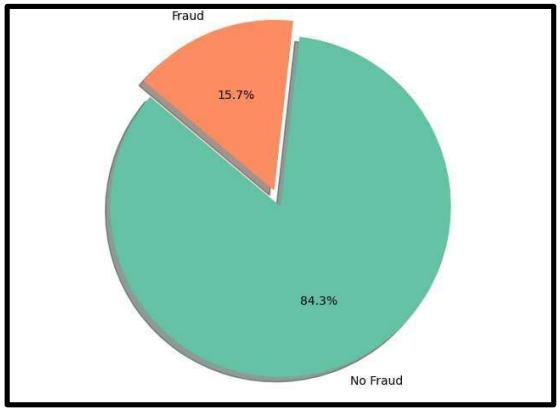


Fig 13: Distribution of Fraud vs No Fraud

This pie chart indicates the distribution of instances between these two classes: cases of fraud and non-fraudulent transactions. As observed, 84.3% are non-fraudulent, while 15.7% are fraudulent cases. This figure indicates the imbalance of the dataset, which is directly related to the challenge of fraud detection building an accurate model that must be able to efficiently detect these less frequent fraudulent transactions [21].

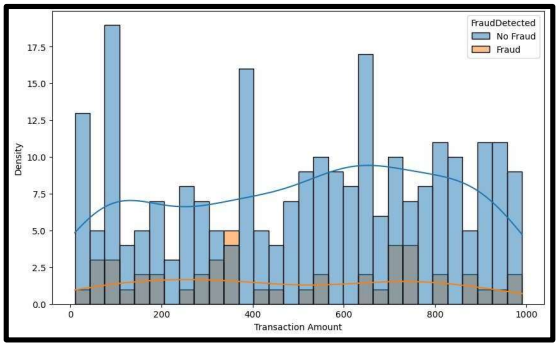


Fig 14: Transaction Amount Distribution by Fraud Detection

This histogram is the result of visualizing the individual transaction amounts while differentiating between the fraudulent and the non-fraudulent transactions. Fake payments (orange) perform various values, yet there are some groups of fraudulent transactions with lower values. The blue colour demonstrates the prevalence of non-fraudulent behaviour across the entire range and can be attributed to a large number of transactions in the dataset and the challenge presented by various fake transactions in certain ranges.

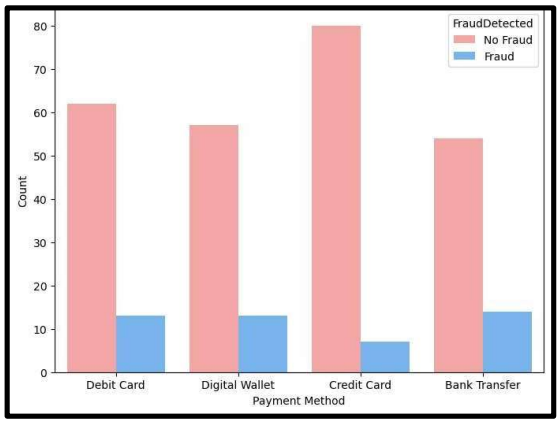


Fig 15: Payment Method Distribution by Fraud Detection

This bar chart represents the payment methods for both fraudulent and non-fraudulent transactions. Credit Cards is the most frequently used method with Digital Wallets coming as the second most common. Although transactions involving fraud are relatively low, they are slightly higher with Credit Cards and Digital Wallets.

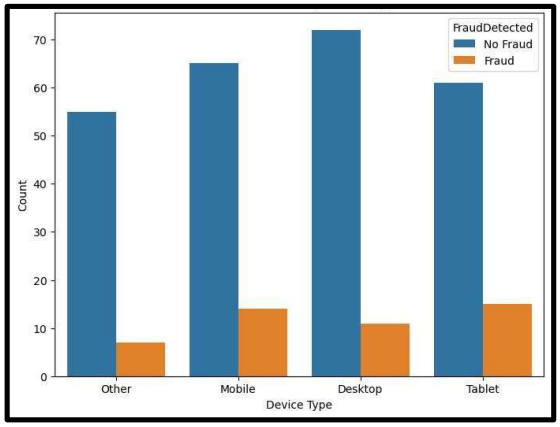


Fig 16: Fraud Detection by Device Type

This figure shows how fraud is divided among devices. Fraud is more evenly spread across different devices though Desktop and Mobile devices are used more often than others. It must be seen that the simple presence of a device is not sufficient to support the fraud hypothesis and analysis of usage across multiple types of devices can be helpful for more targeted efforts in this sphere.

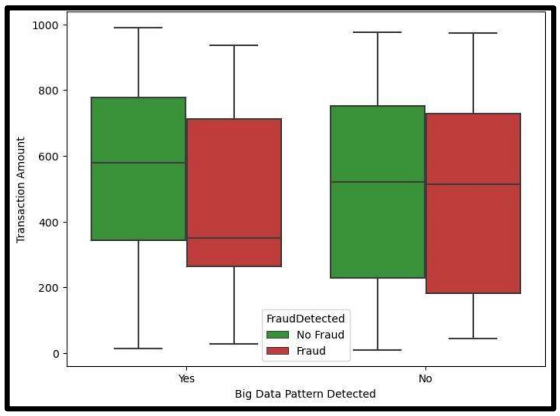


Fig 17: Transaction Amount by Big Data Pattern Detected vs Fraud Detection

The above box plot shows the transaction amounts considering Big Data pattern detection and fraud status. Analyzing the characteristics of the transactions, non-fraudulent ones have a wider range when exceptions in terms of Big Data behaviour patterns are detected while fraudulent ones have a uniform distribution. This means that the two factors such as transaction amount and Big Data pattern detection might not fully capture fraud hence the necessity for other factors that have a complexity.

Gradient Boosting Accuracy: 0.778

	precision	recall	f1-score	support
0	0.83	0.92	0.87	75
1	0.14	0.07	0.09	15
accuracy			0.78	90
macro avg	0.49	0.49	0.48	90
weighted avg	0.72	0.78	0.74	90

Fig 18: GB model accuracy and classification report

The above classification report has the Gradient Boosting model performance at an accuracy of 77.8%. It seems great for deciding nonfraudulent transactions but is poor in fraud detection, reflected in the low recall of 0.07 for fraud. More specifically, a low F1-score for fraud may indicate the model's struggle to find a balance between precision and recall for fraud [22].

Random Forest Accuracy: 0.822

	precision	recall	f1-score	support
0	0.83	0.99	0.90	75
1	0.00	0.00	0.00	15
accuracy			0.82	90
macro avg	0.42	0.49	0.45	90
weighted avg	0.69	0.82	0.75	90

Fig 19: Accuracy and Classification report of the RF model

The Random Forest model shows an improvement in accuracy to 82.2% with a Random Forest model. The GB model, it still has a hard time detecting Fraud cases since Precision and Recall in fraudulent transactions are 0.00. This implies a challenge in handling imbalanced datasets whereby the model perfectly predicts non-fraudulent transactions but fails to identify fraudulent ones.

Support Vector Machine Accuracy: 0.833

	precision	recall	f1-score	support
0	0.83	1.00	0.91	75
1	0.00	0.00	0.00	15
accuracy			0.83	90
macro avg	0.42	0.50	0.45	90
weighted avg	0.69	0.83	0.76	90

Figure 20: SVM model accuracy and classification report

This figure shows the performance of the ***"Support Vector Machine (SVM)"*** model, which yields an accuracy of 83.3%. In general, the SVM model is excellent at classifying non-fraudulent cases since it has a recall of 1.00 and an F1 score of 0.91. However, for fraudulent cases, the model did poorly, failing to detect such cases properly; both precision and recall equalled 0.00. This may indicate challenges with imbalanced datasets [23].

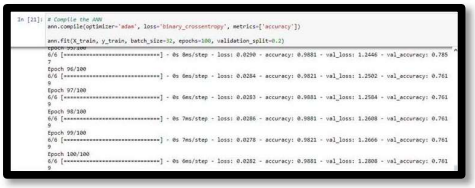


Figure 22: ANN model fit

This figure illustrates the process of fitting the Artificial Neural Network model during training. The model is trained on 100 epochs with a batch size of 32, the Adam optimizer, and binary cross-entropy as the loss function. The model was able to achieve a training accuracy of 98.81% and a validation accuracy of 76.1% by the end of training. The loss on training data is approximately 0.0290, which increases very slightly as the epochs progress, hence probably meaning that the model must be overfitting in some manner. This is due to the validation accuracy plateaus at around 76.1%. It can thus be interpreted that although this model works well on training data, it can still have limited generalization capacity on unseen data.

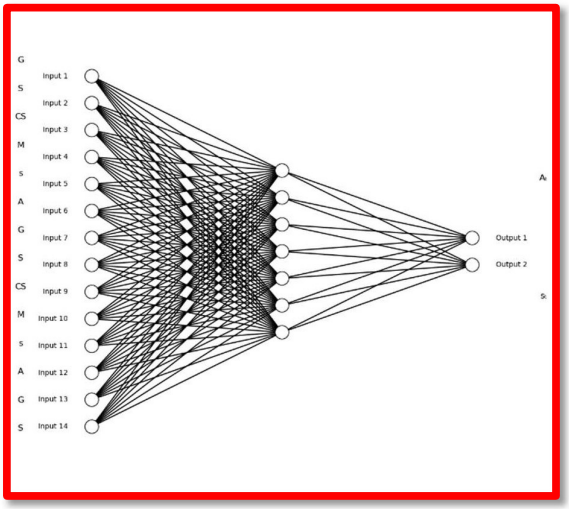


Fig 21: ANN model structure

```
3/3 [==============================] - 0s 1ms/step

ANN Model Accuracy: 0.778
              precision    recall  f1-score   support

           0       0.84      0.91      0.87        75
           1       0.22      0.13      0.17        15

    accuracy                           0.78        90
   macro avg       0.53      0.52      0.52        90
weighted avg       0.74      0.78      0.75        90
```

Figure 23: ANN Model Accuracy

The figure shows the overall accuracy of the ANN model on the test dataset, which is 77.8%. The classification performance of the model is shown through precision, recall, and F1-score metrics for both the classes: Fraud and No Fraud. The model is very precise; however, for the Fraud class, the model's performance drastically goes down to a precision of 0.22 with a recall of 0.13. In this case, a clear performance imbalance exists; the model is very poor at correctly classifying fraud cases. This must be caused by a class imbalance in the dataset or model underfitting for the minority class.

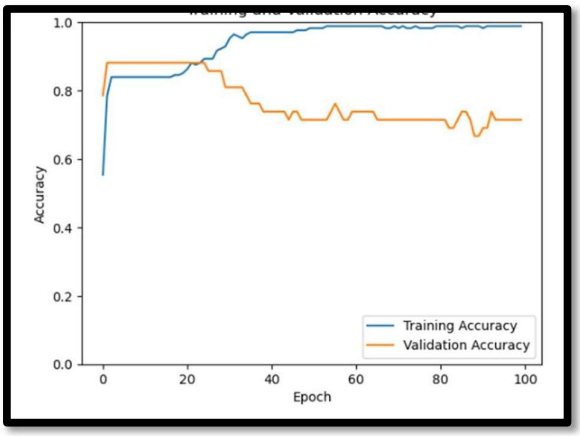


Figure 24: Training and Validation Accuracy

The training and validation set over 100 epochs. The training accuracy rapidly rises to above 90% within the first 10 epochs and continues increasing steadily as training continues to about 99%. In contrast, validation accuracy achieves about 80% early before starting to decline at around 20 epochs, fluctuating further between 70% and 80%. This divergence of training and validation accuracy might be indicative of possible overfitting—very good performance during training but not generalizing onto the validation data—a clear indication that model complexity might be too large for given data.

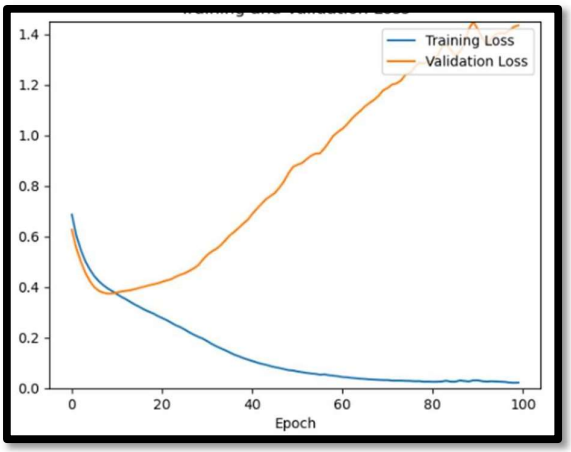


Figure 25: Training and Validation Loss

Loss curves of training and validation sets during model training which are shown in the above figure. The training loss decreases consistently, reaching close to 0 by the end of 100 epochs. In contrast, the validation loss drops at the beginning but from 30 epochs onward increases steadily until it has overtaken the training loss by a significant amount. This again

confirms that the model is overfitting on the training data by the increase in validation loss while the training loss is decreasing. In itself, this overfitting can already be taken as an indication that the model learned to fit noise in the training data rather than generalizing to unseen data, hence suggesting the use of regularization techniques or simpler model architectures.

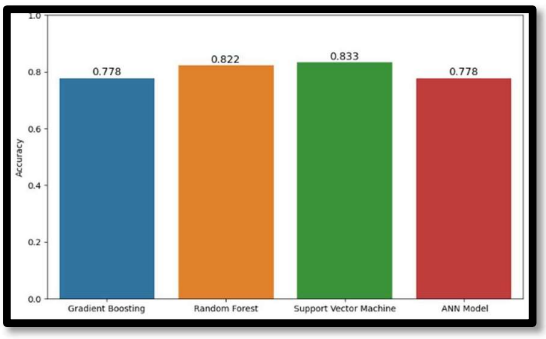


Fig 26: Model Accuracy Comparison

This bar chart compares the accuracy of four model: Gradient Boosting, Random Forest, SVM, and the ANN model. Among these, the SVM model achieves the highest accuracy at 83.3%, followed by RF at 82.2%. Both the Gradient Boosting and ANN models perform similarly, with an accuracy of 77.8%. The relatively lower performance of the ANN model compared to SVM and Random Forest suggests that traditional "machine-learning models" must be better suited for this dataset. The comparison reveals that the proposed SVM model has the highest overall accuracy compared to the other two models that were trained with the data set.

B. Discussion

The evaluation of the three machine learning models is based on their efficiency in identifying fraudulent transactions where Gradient Boosting, Random Forest, and the Support Vector Machine Sonnet (SVM) were used. From the results, the SVM model shows the highest accuracy of 83.3%, while Random Forest is 82.2%, and Gradient Boosting at 77.8%. Also, the ANN models perform similarly to the GB model, with an accuracy of 77.8%Yet, all models seem to have issues with fraud detection as shown through the precision, recall, and F1-score values for fraudulent transactions. This is an issue common with datasets where the minority class is fraud as models learn to predict non-fraudulent transactions incredibly well but struggle with the rarely occurring fraud.

Model	**Accuracy**	**Precision (Fraud)**	**Recall (Fraud)**	**F1-Score (Fraud)**
Gradient Boosting	0.778	0.14	0.07	0.09
Random Forest	0.822	0.00	0.00	0.00
Support Vector Machine	0.833	0.00	0.00	0.00
ANN	0.778	0.22	0.13	0.17

Table 3: Performance Summary

V: Conclusion And Recommendation

A Critical Evaluation

AI-driven biometric authentication and Big Data analytics in digital payment security yield several key implications for the future. The coupling of AI with Biometrics allows an adequate mechanism for fraud detection in real-time, which brings down the probability of conducting transactions without a user's direct permission. In this regard, the power of analyzing large datasets poses AI as capable of adapting to new threats continuously, therefore increasing resilience in payment systems. Biometric authentication just adds a layer of security on top, making it as hard as possible for malicious actors to gain unauthorized access. However, the proposed approach has limitations. Powerful as today's AI must be, the

strong predictive abilities that the software can provide have to do with data, in terms of quality and quantity; incomplete or biased data must introduce inaccuracies and must further compromise security [24]. It is simply that biometric data raises questions of privacy and data protection, since this sort of sensitive information must be misused, or a breach thereof must have very serious consequences indeed. Scalability is also problematic: huge computational resources and infrastructure must be required for the deployment of AI-driven biometric systems.

B Research recommendation

Digital payment security and fraud detection based on AI can be further enhanced to detect emerging patterns of fraud by integrating AI, Big Data analytics, and biometric authentication. These systems can imbibe automatically adjusting AI models to new threats in real-time and, hence, further repel sophisticated attacks. Big Data analytics can be directed toward the fusion and analytics of transaction data over all platforms to better identify fraudsters. Improved fusion in techniques of data may help in accepting various data sources, hence improving security measures by rigidity. Multi-modal biometric systems combining a multitude of biometric traits may fortify the process of authentication and bring down the count of false positives and negatives. It has been very critical to make sure that these biometric systems are resistant to spoofing and hence offer security integrity. There must be a focus on the development of standardized frameworks for industry adoption concerning the interoperability of biometric data use amongst multiple systems and between different providers.

C Future work

It is requisite to see future research in digital payments in some areas. First of all, advanced AI models need further investigation, like deep learning networks, to build better fraud detection capabilities. Second, newer biometric technologies can provide even higher security authentication methods, in particular, behavioural biometrics and multimodal systems. Third, privacy concerns about biometric data are to be taken seriously by improving anonymization and encryption techniques. Blockchain can provide a decentralized and tamper-proof record of transactions, thereby potentially increasing transparency and reducing fraud [25]. For protecting digital payments, quantum-resistant encryption methods are in dire need due to the arrival of quantum computing. It must leverage synergies with AI for fraud detection and blockchain for secure transaction recording. It is in the area of behavioural analytics that research can make a difference in fraud detection through the identification of abnormal behaviour patterns in users.

References

[1] Awad, A.I., Babu, A., Barka, E. and Shuaib, K., 2024. AI-powered biometrics for Internet of Things security: A review and future vision. Journal of Information Security and Applications, 82, p.103748.

[2] Hassan, M., Aziz, L.A.R. and Andriansyah, Y., 2023. The role artificial intelligence in modern banking: an exploration of AI-driven approaches for enhanced fraud prevention, risk management, and regulatory compliance. Reviews of Contemporary Business Analytics, 6(1), pp.110-132.

[3] Sambrow, V.D.P. and Iqbal, K., 2022. Integrating Artificial Intelligence in Banking Fraud Prevention: A Focus on Deep Learning and Data Analytics. Eigenpub Review of Science and Technology, 6(1), pp.17-33.

[4] Olweny, F., 2024. Navigating the nexus of security and privacy in modern financial technologies. GSC Advanced Research and Reviews, 18(2), pp.167-197.

[5] Majeed, A. and Hwang, S.O., 2021. Data-driven analytics leveraging artificial intelligence in the era of COVID-19: an insightful review of recent developments. Symmetry, 14(1), p.16.

[6] Nayak, A., Patnaik, A., Satpathy, I. and Patnaik, B.C.M., 2024. Data Storage and Transmission Security in the Cloud: The Artificial Intelligence (AI) Edge. In Improving Security, Privacy, and Trust in Cloud Computing (pp. 194-212). IGI Global.

[7] Tyagi, A.K., Aswathy, S.U. and Abraham, A., 2020. Integrating blockchain technology and artificial intelligence: Synergies perspectives challenges and research directions. Journal of Information Assurance and Security, 15(5), p.1554.

[8] Suresh, H.N. and GC, N.S., 2022. Enhancing Security in E-Banking through Artificial Intelligence. Educational Administration: Theory and Practice, pp.220-224.

[9] Jain, R., Prajapati, D. and Dangi, A., 2023. Transforming the financial sector: A review of recent advancements in FinTech. Available at SSRN 4380348.

[10] Davenport, T.H. and Mittal, N., 2023. All-in on AI: How smart companies win big with artificial intelligence. Harvard Business Press.

[11] Shoetan, P.O., Oyewole, A.T., Okoye, C.C. and Ofodile, O.C., 2024. Reviewing the role of big data analytics in financial fraud detection. Finance & Accounting Research Journal, 6(3), pp.384-394.

[12] Bhattacherjee, A. and Badhan, A.K., 2024. Convergence of Data Analytics, Big Data, and Machine Learning: Applications, Challenges, and Future Direction. In Data Analytics and Machine Learning: Navigating the Big Data Landscape (pp. 317-334). Singapore: Springer Nature Singapore.

[13] Kaushik, K., Khan, A., Kumari, A., Sharma, I. and Dubey, R., 2024. Ethical Considerations in AI-Based Cybersecurity. In Next-Generation Cybersecurity: AI, ML, and Blockchain (pp. 437-470). Singapore: Springer Nature Singapore.
[14] Zhou, J., Chen, C., Li, L., Zhang, Z. and Zheng, X., 2022. FinBrain 2.0: when finance meets trustworthy AI. Frontiers of Information Technology & Electronic Engineering, 23(12), pp.1747-1764.
[15] Das, S. and Ganguly, D., 2024. Protecting Your Assets: Effective Use of Cybersecurity Measures in Banking Industries. In Next-Generation Cybersecurity: AI, ML, and Blockchain (pp. 265-286). Singapore: Springer Nature Singapore.
[16] Kaloudi, N. and Li, J., 2020. The ai-based cyber threat landscape: A survey. ACM Computing Surveys (CSUR), 53(1), pp.1-34.
[17] Xu, J., 2022. AI Theory and Applications in the Financial Industry. Future And Fintech, The: Abcdi And Beyond, 74.
[18] Khan, A., Jafar, S.H. and El-Chaarani, H., 2024. Evolution of Fintech in the Financial Sector: Recent Trends and Future Perspectives. The Adoption of Fintech, pp.17-33.
[19] Elkhodr, M., Khan, S. and Gide, E., 2024. A novel semantic IoT middleware for secure data management: blockchain and AI-driven context awareness. Future Internet, 16(1), p.22.
[20] Hung, A.H.C., 2023. Examining the AI-Based Biometric Surveillance Data Collected by Employers: A Review Based on Federal and State Laws Protecting Employee Privacy Rights. Hofstra Lab. & Emp. LJ, 41, p.25.
[21] Tan, E., 2022. Chapter 2: The role of big data, AI and blockchain technology in digital public governance. In The new digital era governance: How new digital technologies are shaping public governance (pp. 193-204). Wageningen Academic Publishers.
[22] Kapadiya, K., Patel, U., Gupta, R., Alshehri, M.D., Tanwar, S., Sharma, G. and Bokoro, P.N., 2022. Blockchain and AI-empowered healthcare insurance fraud detection: an analysis, architecture, and future prospects. IEEE Access, 10, pp.79606-79627.
[23] Kumar, S. and Aithal, P.S., 2023. Tech-Business Analytics in Tertiary Industry Sector. International Journal of Applied Engineering and Management Letters (IJAEML), 7(4), pp.349-454.
[24] Rane, N., 2023. Integrating leading-edge artificial intelligence (AI), internet of things (IOT), and big data technologies for smart and sustainable architecture, engineering and construction (AEC) industry: Challenges and future directions. Engineering and Construction (AEC) Industry: Challenges and Future Directions (September 24, 2023).
[25] Ali, S.I., 2024. Consideration of Web Technology and Cloud Computing Inspiration for AI and IoT Role in Sustainable Decision-Making for Enterprise Systems. Journal of Information Technology and Informatics, 3(2).