IJRAR.ORG E-ISSN: 2348-1269, P-ISSN: 2349-5138

INTERNATIONAL JOURNAL OF RESEARCH AND ANALYTICAL REVIEWS (IJRAR) | IJRAR.ORG

An International Open Access, Peer-reviewed, Refereed Journal

Forecasting the Evolution of Cyber Attacks in FinTech Using Transformer-Based Time Series Models

Adarsh Reddy Bilipelli
Independent Researcher

Abstract—In the fast-paced digitalizing financial landscape, cyberattacks against FinTech platforms become more complex and pose an ever-increasing threat to their operations. To meet the necessities of a proper and timely threat prediction, the proposed study presents the Alert BERT, a transformer-based deep learning model specialized in predicting the evolution of cyber threats within FinTech. The model is trained using a robust preprocessing with the help of the IEEE-CIS fraud dataset, in which data cleaning, normalization, categorical encoding, and SMOTE-based class balancing are performed. Alert BERT uses the contextual learning ability of BERT on the structured transaction data and is capable of capturing sequential patterns of cyber threats. The model outperformed more conventional models like Deep Neural Networks, Random Forest, and GBDT, with a high accuracy rate of 97.2%, precision of 91.8%, recall of 94.1%, and an F1-score of 92.9%. Comparative assessment using confusion matrices, ROC curves and learning curves also supports its predictive capability and robustness. This structure makes it possible to anticipate threats, which enhances FinTech cyber resiliency to a great extent. Alert BERT provides a cost-effective, data-driven tool to optimize the cybersecurity posture in complex financial environments through the integration of real-time forecasting and high-performance sequence modeling attempts to deliver a scalable approach to enhancing cybersecurity posture with a high degree of confidence in dynamic financial environments.

Keywords—Cybersecurity, FinTech, Transformer Models, Time Series Prediction, Financial Technology.

I. INTRODUCTION

The Financial sector has always been a leading force in technological innovation whereby the industry has moved away from conventional forms of banking and embraced technological platforms, which provide efficient and more convenient and easier-to-use forms of financial services. Financial technology (FinTech) has transformed the financial service delivery making it efficient, accessible and innovative due to the rapid development around the field [1]. FinTech, which includes robo-advisors, blockchain-based payments, digital banking, and mobile payments, has upended the traditional financial system. Yet, the digitalization of this industry has led to the industry being vulnerable to more advanced and dynamic cyber-attacks. Cybercriminals are taking advantage of flaws on digital infrastructures to coordinate diverse attacks, such as phishing, ransomware, data breaches, and advanced persistent threat (APT) [2]. The stakes could not be higher when it comes to active and intelligent cybersecurity as FinTech platforms are processing vast amounts of sensitive data and high-frequency transactions on a regular basis. It is necessary to forecast the development of cyberattacks to be ready to face potential risks, build the most effective defense systems, and make the FinTech infrastructures resistant to cyberattacks [3].

The conventional cybersecurity systems tend to use rule-based models and past data trends to extract threats. Although these approaches produce results to a certain level, they do not offer the flexibility and the ability to predict the future that can be used to deal with the emergence of new threats that dynamically change over time. Deep learning (DL) and machine learning (ML) models have been used to fill this gap more frequently and provide better accuracy and flexibility [4]. One of these, transformer-based models has grown to become state-of-the-art sequence modelling tasks because of their capability to learn long-range dependency, and contextual relations in data. Transformers were originally designed to be used in natural language processing (NLP), but their performance in many different fields has been limited only by self-attention and the parallel nature of their computation, leading to applications in time series forecasting. [5].

Time series forecasting is a very important method of predicting the temporal pattern of cyberattacks. With regard to FinTech, predicting how frequently, what kind of, and how significant cyber threats will be in the future, may be useful in terms of early warning systems, risk mitigation measures, and resource deployment [6]. As opposed to the classic recurrent neural networks (RNNs) and long short-term memory (LSTM) model, the time series models based on transformers do not have the problem of vanishing gradient and are more adaptive to large-scale data with complicated temporal dynamics. The ability of such models to learn regularities in cyberattack data, e.g. to find spikes in data during financial quarters or trends in anomalies related to new FinTech products, can effectively empower cybersecurity predictive intelligence [7].

The transformation of time series analysis through the use of transformer architectures is a potential source for predicting cyber threats in the FinTech domain. These models may offer an insight into future attack vectors by utilizing historical data regarding cyber incidents as well as past network traffic records, threat intelligence feeds, and telemetry data in real-time. The latter aspect is especially relevant to FinTech organizations that serve in a highly-regulated environment and are within the responsibility of ensuring the safety of all

customer data and the integrity of all service elements. Moreover, accurate forecasting allows stakeholders to implement timely security patches, reinforce system defenses, and inform strategic decisions regarding cybersecurity investments.

A. Motivation and Contributions of the Study

The high growth rate and digitalization of the FinTech industry have also brought about more advanced cyberattacks that put financial systems and consumer confidence at high risk. Conventional reactive cybersecurity is ineffective in identifying the changing attack patterns timely hence leading to huge losses and disruption of operations. Consequently, there is an urgent need for predictive models capable of forecasting the evolution of cyber threats to enable proactive defense strategies. Leveraging advanced NLP and DL techniques, particularly transformer-based models, presents a promising direction for capturing complex temporal and contextual patterns in cyber-attack data. The main key contributions include:

- Utilized the IEEE-CIS fraud dataset for comprehensive cyber-attack data relevant to FinTech transactions.
- Conducted rigorous data pre-processing including data cleaning, categorical encoding, numerical feature normalization, and synthetic minority oversampling (SMOTE) to balance class distribution.
- Implemented Alert BERT, a transformer-based deep learning architecture tailored for capturing sequential and contextual information for cyber-attack forecasting.
- The dataset should be included in training and testing sets to guarantee model generalizability and reliable assessment.
- Evaluated the model using accuracy, precision, recall, and F1-score metrics to validate its effectiveness in predicting cyber-attack patterns and enhancing FinTech security.

B. Justification and Novelty

The novelty of the proposed Alert BERT-based framework lies in its integration of a pre-trained BERT model with a specialized architecture tailored for fraud and cyber-attack detection in FinTech environments a domain where contextual understanding of sequential transactional data is crucial. In contrast to the conventional models, Alert BERT uses deep contextual embeddings to model complicated temporal relationships and intricate behavioral patterns and is far more likely to predict optimal outcomes. Class balancing using SMOTE makes the study objective better in identifying minority cases of fraud, which is one of the limitations of imbalanced data. In addition, the framework has been designed to run in real-time and thus it can be used in high-speed financial environments where high-speed threat detection is a primary concern. The demonstrated strong performance across multiple evaluation metrics justifies its effectiveness and establishes Alert BERT as a robust and innovative solution for proactive FinTech cybersecurity.

C. Organization of the paper

The paper is organized as follows: Section II presents a review of related work. Section III details the proposed methodology and evaluation techniques. Section IV discusses the results, and comparative analysis. Section V concludes and future research to enhance predictive cybersecurity in FinTech environments.

II. Literature Review

This section reviews recent advances in AI, anomaly detection, and predictive analytics for cybersecurity in FinTech, focusing on ML and DL methods that aim to forecast evolving cyber-attacks, strengthen proactive defenses, and enhance risk management in financial technology environments. Some of the reviewed works are:

Qasaimeh et al. (2022) developed a model that uses a DNN to predict upcoming network-based cyberattacks on financial institutions. Some of the most significant cyberattacks on financial institutions throughout the previous three years made up the dataset used to train and test the algorithm. After that, the forecasting model's performance was assessed in a real banking setting, yielding a predicting accuracy of 90.36% [8].

Akintoye et al. (2022) investigated how Nigerian Deposit Money Banks' financial innovation is fuelled by cybersecurity. The basic data that was gathered was examined using both descriptive and inferential statistics. Based on bank monitoring and risk management, the study discovered that cybersecurity positively and statistically significantly impacted Nigerian deposit money institutions' financial innovation. $F(2,55)=23.274$, $p<0.05$, adj.$R2 = 0.447$ [9].

Lacruz and Saniie (2021) proposed how techniques for AI and ML may help detect credit card fraud.They create two distinct approaches for high-accuracy fraud detection after taking a theoretical approach to the topic: Autoencoder (semi-supervised learning) and Logistic Regression (supervised learning). The results from both methods are promising since they could predict fraudulent transactions with 94% certainty. [10].

Ochoa et al. (2021) proposed a quantitative model to help organizations analyze cyber risks in financial terms, particularly in the financial sector. A Peruvian corporation optimized its cybersecurity spend by 32.2% after implementing the methodology. This is particularly important as more companies are becoming victims of ransomware infections, and few organizations can calculate their impact in monetary terms [11].

Al Duhaidahawi et al. (2021) conducted research on how cybersecurity is affected by Fintech factors. They developed hypotheses based on statistical results, finding a positive correlation coefficient at a significance level of 0.01. A positive effect factor between the research variables was also discovered by the authors; when financial technology was connected to cybersecurity, its influence increased dramatically to 0.908. Because the effect coefficient was strong and increased to 0.908, suggesting a positive association between the study variables, this suggests complementarity of the independent variable's domains. [12].

Pattabhi (2020) discussed how deep learning could be used to forecast data breaches and especially in the financial sector. The model combines information on previous security breaches, metadata, LSTM, and CNN models to analyze IT infrastructure data, records of user behavior as well as threat intelligence feeds. The paper describes why the traditional risk management methods are incompetent and elaborates on data management, feature gathering, model architecture design, and assessment. An average F1-score of 12% was attained when the model was evaluated using a reference dataset and compared to more conventional machine learning methods like Support Vector Machines and Random Forests [13].

Table I provides an overview of the previous work that has been conducted related to cybersecurity and financial technology, and its associated methods, data, and results, along with limitations and directions to future research with a particular emphasis on forecasting and risk quantification

TABLE I. SUMMARY OF RELATED STUDIES ON CYBER-ATTACK FORECASTING AND CYBERSECURITY IN FINTECH

Author	Methodology	Data	Key Findings	Limitation/Future Work
Qasaimeh et al. 2022	Deep Neural Network to forecast cyber-attacks	Historical data of major cyber-attacks on banking institutions (3 years)	Developed a prediction model with a 90.36% accuracy rate for upcoming network-based cyberattacks.	Limited to network-level attacks in traditional banking; does not cover wider FinTech or new attack vectors
Akintoye et al. 2022	Survey research; descriptive & inferential statistics	Structured questionnaire from 56 senior staff in Nigerian Deposit Money Banks	discovered that cybersecurity has a statistically significant beneficial influence on financial innovation through risk management and monitoring (Adj.R^2 = 0.447, F (2,55) =23.274, $p < 0.05$)	Focused on organizational practices, not technical forecasting; does not predict future attack trends
Lacruz and Saniie 2021	AI & ML: supervised logistic regression and semi-supervised autoencoder for fraud detection	Credit card transaction datasets	Both models detect fraud with ~94% accuracy; showing AI's effectiveness for fraud detection	Detects current fraud but does not forecast new cyber-attack patterns; limited to card fraud use case
Ochoa et al. 2021	Quantitative model to calculate ransomware impact in financial terms	Cyber-risk data from a Peruvian financial company	The proposed model enabled a 32.2% optimization in cybersecurity investment decisions	Focuses on cost quantification, not predictive modeling of evolving attacks
Al Duhaidahawi et al. 2021	Statistical analysis: correlation & regression	Survey data linking FinTech variables to cybersecurity	Strong positive relationship between FinTech adoption and cybersecurity needs; high influence factor (0.908)	Explores correlations but does not forecast future threats; lacks predictive models
Pattabhi, 2020	Hybrid deep learning: LSTM & CNN vs. SVM & RF	Historical security breaches, user behavior, threat feeds	Hybrid model predicts data breaches better than classic ML; highlights deep learning's advantage for proactive security	F1-score still low (12%); limited real-world FinTech-specific validation; requires richer, real-time datasets

III. METHODOLOGY

The proposed methodology introduces a useful method of fraud identification in FinTech based on the IEEE-CIS dataset. It consists of thorough pre-processing of data, where data cleaning is performed to remove inconsistencies, encoding of categorical variables to handle non-numeric variables, normalization of numerical features, and the balancing of datasets based on SMOTE to counter the class imbalance. After the pre-processing, the data is partitioned into testing and training data. The main aspect of this methodology is that it involves the adoption of the AlertBERT, which is a DL model based on BERT, specifically designed to predict fraud. Alert BERT receives the pre-processed training data and uses it to train and test itself. Among the classification performance metrics used to assess performance and guarantee the reliability and robustness of the model are accuracy, precision, recall, and F1-score. The approach illustrated in Figure 1, warrants that the model receives high-quality and well-balanced data, resulting in a better predictions and effective results of fraud detection.

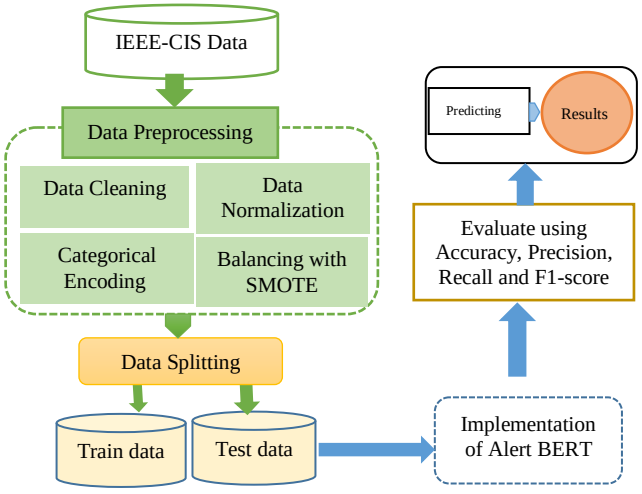

Fig. 1. Proposed Methodology for Fraud Detection using AlertBERT on IEEE-CIS Dataset

A. Data collection

The IEEE-CIS dataset is among the most extensive publicly accessible datasets on transactional fraud detection. It includes associated device and browser information together with anonymized portions of a variety of financial activities. Since the dataset is extremely unbalanced and contains very few fraud incidents, it may be used to evaluate the sensitivity of detection. The visualizations of the data are given below:

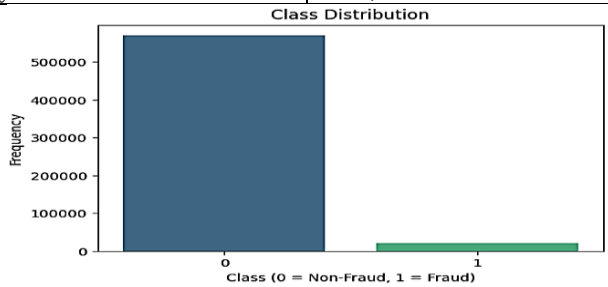

Fig. 2. Class Distribution of Fraud vs. Non-Fraud Classes

This bar chart in Figure 2 displays the class distribution of a dataset, likely related to fraud detection. It shows a significant imbalance, with a very large number of "Non-Fraud" instances (Class 0) dominating the dataset, while "Fraud" instances (Class 1) are a small minority. The frequency of non-fraud is over 500,000, whereas fraud is less than 50,000.

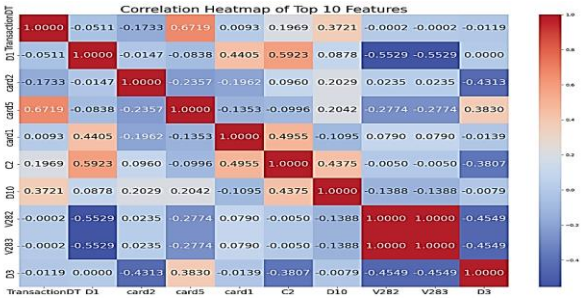

Fig. 3. Correlation Heatmap of the Dataset

The correlation heatmap shown in Figure 3 illustrates the connections among the dataset's top 10 attributes. The color intensity and shade (red for positive, blue for negative) indicate the strength and direction of the correlation, with values ranging from -1 to 1. The diagonal shows perfect positive correlation (1.0000) of each feature with itself. Key features like 'TransactionDT', 'card1', 'card2', 'card5', 'C2', 'D10', 'V282', 'V283', and 'D3' are analyzed, revealing various degrees of linear correlation among them.

B. Data Preprocessing

A critical and necessary stage in creating and implementing ML models is data preparation. By making certain that the input data is clear and reliable, and structured correctly, appropriate preprocessing enables the model to learn pertinent patterns and function well on fresh, unknown data. The following preprocessing steps were undertaken:

C. Data Cleaning

Data cleaning involved removing duplicates, correcting inconsistencies, and eliminating records that aren't crucial to ensure the reliability and correctness of the data. To preserve dataset completeness without skewing the results, mean imputation for numerical features and mode imputation for categorical characteristics were used to manage missing values.

D. Data Normalization

In order to create a new range of data from an old one, normalization is used to reduce the data's variances and make them look more similar in a well-behaved. Min-Max Normalization with the [L,U] range (usually ranging from 0 to 1) [14], Equation (1) formula is used to fit each feature with value x inside a predetermined boundary, normalising it in terms of the values for the minimum and maximum, x_{max} and x_{min}, Respectively.

$$x' = \frac{(x - x_{min})}{x_{max} - x_{min}} \times (U - L) + L \quad (1)$$

E. Categorical Encoding

Categorical feature encoding is a preprocessing step used to convert categorical data into a numerical format that ML algorithms can interpret. Since most models require numerical input, encoding techniques are applied to transform non-numeric categories into a suitable numerical representation. Label encoding and one-hot encoding are two popular techniques that are selected according to the characteristics of the categorical data. When label encoding is used with ordinal features, a unique integer value is assigned to each category according to its order. Nominal features are encoded using one-hot encoding, each category is represented as a binary vector, with only the element at the category position set to 1, and the rest to 0.

F. Synthetic Minority Oversampling Technique (SMOTE)

The SMOTE is used to enrich class balance in the datasets of fraud detection. This method creates artificial examples of the minority category by interpolating between an example of the minority category xi and one of its k-neighbours x_{nn}, as in the equation below, Equation (2)

$$x_{new} = x_i + \lambda(x_{nn} - x_i), \lambda \epsilon [0,1] \quad (2)$$

where x_{nn} is among the k-nearest neighbors of x_i, and λ Is a random number between 0 and 1. This method assists the model to learn more about the better boundaries of decisions because it offers more informative samples, as opposed to just copying the existing ones. The balanced class distribution of the binary classification task is as follows:

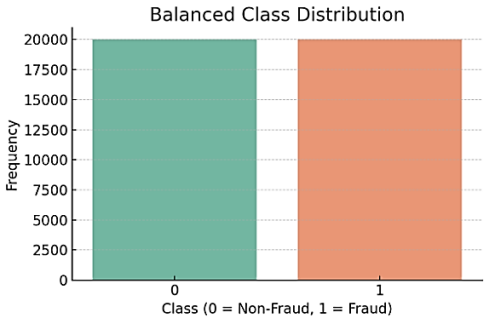

Fig. 4. Class distribution of binary classification

Figure 4 shows a balanced distribution of classes. In reference to the binary classification issue, transactions in Class 0 are non-fractional, whereas transactions in Class 1 are fraudulent. The number of samples is the same for each class (20,000) so that the frequency of the two categories was equal. Such a balanced dataset is important in training the ML models because it reduces bias towards the majority class and improves the model to give accurate results for the instances as either being a fraud or not being a fraud.

G. Data Splitting

The model was trained using 70% of the data set, which was split into two halves, in order to find patterns, and the other 30% set aside, to test the model, in order to determine its performance and capability to generalize to new untested information. This segment offers a solid assessment on the predictive performance of the model.

H. Classification of AlertBERT Model

AlertBERT is a time series-specific transformer, that has anomaly detection and forecasting tasks in mind. AlertBERT uses transformer architecture to describe temporal relations in sequential data, which is part of the BERT (Bidirectional Encoder Representations from Transformers) revolution in natural language processing. In contrast to the conventional recurrent models, transformers utilize self-attention functionality, which enables the model determine each timestamp's relative importance in relation to the others in the input sequence, so that the model may have a more significant understanding of long-range dependencies and intricate temporal patterns.

At its core, AlertBERT processes a time series input X={x1,x2,...,xT}, where T is the sequence length. The input is first embedded into a latent space through positional encoding PE to retain temporal order, as transformers inherently lack sequence-awareness as shown in Equation (3):

$$Z_0 = X + PE \quad (3)$$

The model then employs feed-forward neural networks and numerous layers of multi-head self-attention. Equation (4) is used by the self-attention mechanism to calculate attention scores:

$$Attention(Q, K, V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right) \quad (4)$$

where d_k He dimensions of the key vectors and queries Q, keys K, and values V are linear projections of Z_0.

I. Performance Metrics

Performance metrics are crucial for assessing prediction algorithms' dependability and efficacy, especially in classification tasks. One of the most widely used tools for visualizing the outcomes of a confusion matrix is a classification model that provides thorough insight into the model's performance, as indicated by the number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). This matrix helps identify not only how accurately the model classifies instances but also where it makes errors. Based on the confusion matrix, several performance metrics are derived to quantitatively assess model accuracy, precision, recall, F1-score, and more. These metrics are essential for evaluating several models and choosing the best one for the task. They are listed below:

1) Accuracy

Accuracy, the main performance assessment parameter, calculates the classifier's proportion of accurate predictions to total predictions [15]. It can be presented in Equation (5):

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \quad (5)$$

2) Precision

The accuracy is calculated by dividing the total number of TP by the sum of FP and TP. It is calculated by dividing the number of positive predictions by the total number of projected positive class values. Equation (6) expressed in precision calculated:

$$Precision = \frac{TP}{TP+FP} \quad (6)$$

3) Recall

Recall is calculated by dividing the total number of TP by the sum of TP and FN. It may be represented in the test data as the proportion of positive class values to positive predictions. Sensitivity or the TPR are other names for it. Its formula expressed in Equation (7).

$$Recall(Rc) = \frac{TP}{TP+FN} \quad (7)$$

4) F1 score

The significance of TP and TN is taken into account by the F-measure, also known as the F1-score. It is the previously established accuracy and recall performance measurements' harmonic mean, which is shown in Equation (8):

$$F1\ score(F1) = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (8)$$

5) Area Under the Receiver Operating Characteristic Curve (AUC)

The area under the receiver operating characteristic curve (AUC) is determined using prediction scores in binary or multi-label classification tasks. The average is then determined using a number of different methods, including weighted, macro, micro, and sample. It may also be computed using sklearn metrics, as shown in Equation (9–10):

$$false - positive - rate = \frac{FP}{(FP+TN)} \quad (9)$$

$$true - positive - rate = \frac{TP}{(TP+FN)} \quad (10)$$

The reliability of the model was evaluated by looking at its predictions on the test dataset using these performance metrics.

IV. Result Analysis And Discussion

This study evaluates the effectiveness of an AlertBERT-based forecasting framework for predicting the evolution of cyber-attacks in the FinTech sector. An NVIDIA Tesla V100 GPU with 16 GB of VRAM was used for training in order to ensure effective handling of large-scale sequential data and deep learning tasks. The tests were carried out using Python 3.9 and PyTorch 2.0. Table II shows that the AlertBERT model performed well on a number of assessment criteria, including F1-score (92.9%), recall (94.1%), accuracy (97.2%), precision (91.8%), and AUC-ROC (98.1%). All of these findings show that the AlertBERT model is quite good at predicting the patterns of cyberattacks, providing reliable detection with high correctness in positive alerts and robust coverage of actual attack instances. This qualifies it as an excellent tool when it comes to active protection and identification of threats within sensitive FinTech scenarios.

TABLE II. Evaluation Metrics for the AlertBERT Model in Forecasting Cyber Attacks in FinTech

Performance Metrics	AlertBERT
Accuracy	97.2
Precision	91.8
Recall	94.1
F1-score	92.9
AUC-ROC	98.1

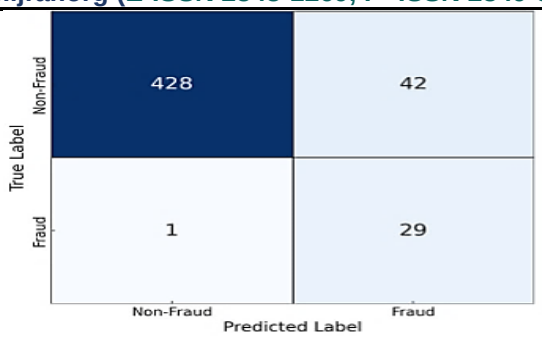

Fig. 5. Confusion Matrix of Alert BERT Model in Forecasting Cyber Attacks in FinTech

Figure 5 displays a confusion matrix evaluating a model's performance in forecasting cyber-attacks in FinTech using the Alert BERT model. It shows the classification results for "Fraud" and "Non-Fraud" categories. The algorithm accurately detected 29 fraudulent instances and 428 non-fraudulent cases, while misclassifying 42 non-fraudulent as fraud and only 1 fraudulent as non-fraudulent.

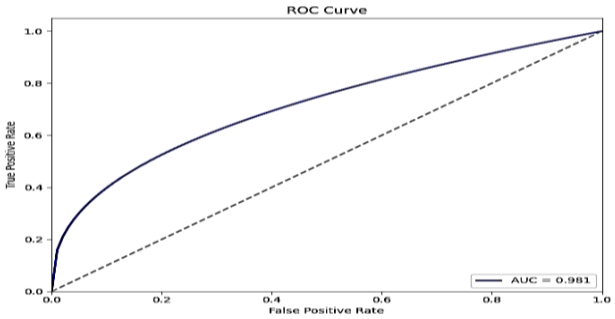

Fig. 6. ROC Curve of Alert BERT Model in Forecasting Cyber Attacks in FinTech

Figure 6 is the Receiver Operating Characteristic (ROC) curve, which is a very important parameter to assess the Alert BERT model in predicting FinTech cyberattacks. The blue graph displays the ratio of the false positive rate to the genuine positive model rate. The superior discriminative power of the model, or its ability to distinguish between positive and negative classes, is demonstrated by its's under the curve (AUC) of 0.981.

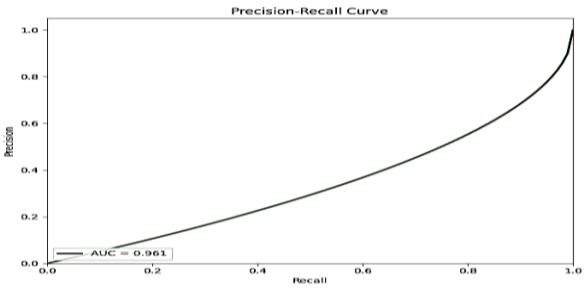

Fig. 7. Precision-Recall Curve of AlertBERT Model in Forecasting Cyber Attacks in FinTech

Precision-Recall Curve representing the performance of Alert BERT model in predicting cyberattacks in FinTech represented in Figure 7. The accuracy versus recall graph for various threshold settings is represented by the green curve. With an AUC of 0.961, the model performs well and offers a decent trade-off between irrelevant versus relevant instances (high accuracy) and relevant against irrelevant instances (high recall).

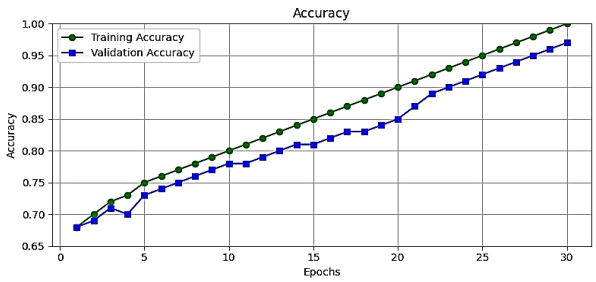

Fig. 8. Training and validation Accuracy of Alert BERT Model in Forecasting Cyber Attacks in FinTech

Figure 8 displays the AlertBERT model's training and validation accuracy for forecasting cyberattacks in the FinTech industry across 30 epochs. The plot shows a linear dynamic growth of both training and validation accuracy, which displays a stable learning process and successful generalization. The similarity in the two plots indicates that there are few chances of overfitting and reveals the fact that the model is strong in handling both the observed and unseen data. In the last epoch, they can see that the training accuracy is approaching 100% and validation accuracy is approaching 97.2 indicating the predictive potential of the model and its consistency during the training.

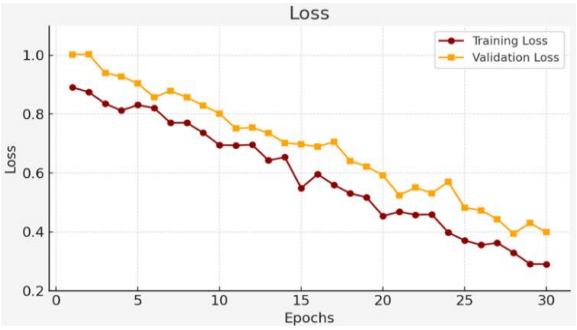

Fig. 9. Training and validation Loss of AlertBERT Model in Forecasting Cyber Attacks in FinTech

The training and validation loss curves of the Alert BERT model throughout the 30 epochs depicted in Figure 9 are essential for predicting FinTech risks. The two graphs show a clear decrease over time illustrating that the model is learning and making fewer and fewer errors on the training set and the validation data. For there is this convergence, the model has been effectively optimized to lessen the discrepancy between expected and actual outcomes for identifying cyberattacks.

A. Comparative Study

In this section, a comparison of ML models applicable to predicting the trends of cyber-attacks in FinTech systems is provided. Table III summarizes different algorithms' accuracy in predicting emerging cyber threats with historical attack telemetry data. It was found that AlertBERT had the best accuracy (97.2%) and thus it performed well in recognizing sophisticated cyclical and contextual trends in cyber-attack sequences. Deep Neural Networks (DNN) came next with an accuracy of 95.3% with great strength in its ability to model non-linear relations in the data. The accuracy of Gradient Boosting Decision Trees (GBDT) was also commendable with 93.5 %. GBDT is a good and explainable method in forecasting threat evolution. Random Forest achieved an accuracy of only 88%, which is the lowest of all compared models, but can still be helpful in situations when interpretability of the model and speed of training are valued most of all. These results suggest that AlertBERT is an excellent candidate to be deployed in FinTech cybersecurity frameworks, where the pace and accuracy of the forecast are crucial to the proactive defence measures.

TABLE III. ACCURACY COMPARISON OF MACHINE LEARNING MODELS FOR FORECASTING CYBER ATTACKS IN FINTECH

Model	Accuracy
AlertBERT	97.2
DNN[16]	95.3
GBDT[17]	93.5
Random forest[18]	88

The comparison analysis demonstrates that, with a maximum accuracy of 97.2%, the proposed AlertBERT model fared better than the other models being examined. DNN came next with 95.3% showing good learning ability but it was not at par with AlertBERT in terms of context. The GBDT provided a balance between performance and interpretability, with an accuracy of 93.5%, and the Random Forest demonstrated the minimum accuracy, 88%. These outcomes prove that AlertBERT is the most appropriate tool to predict cyber-attacks in FinTech settings due to its ability to identify the complexities of patterns.

V. CONCLUSION AND FUTURE SCOPE

In the era of increasing sophistication of cyber threats in the FinTech environment, effective preventive security is essential to maintain trust and the existence of a service. This paper proves the utility of AlertBERT, which is a transformer-based forecasting model of time series, to detect and predict patterns of cyberattacks with a high degree of accuracy. The flexibility of the model in terms of representing intricate time relationships and context dependencies allows identifying the fraudulent behavior in time. With an accuracy of 97.2%, a precision of 91.8%, and an F1 score of 92.9%, AlertBERT outstrips traditional deep learning and ensemble models in both accuracy in detection and flexibility. The covert of SMOTE overcomes the problem of class imbalance, making the predictions fairer on minority fraud classes. Furthermore, the model has been able to demonstrate a sustained performance that has been attested through learning curves as well as ROC analysis, which means that it can be used in real-time in FinTech infrastructures.

Future research will be aimed at identifying multi-modal data sources like API logs, user behavior, threat intelligence feeds to be used in enriching prediction contexts. The further work will be investigating explainable AI elements to improve the transparency of the model and trust of the stakeholders. The further development of real-time deployment capabilities and collaboration of the model with the detection of new attack vectors will add to its usefulness in dynamic and high-risk financial environments.

REFERENCES

[1] S. Anyfantaki, “The Evolution of Financial Technology (FINTECH),” *Bank Greece (Economic Bulletin.*, no. December, pp. 47–62, 2016.

[2] R. R. Suryono, I. Budi, and B. Purwandari, “Challenges and Trends of Financial Technology (Fintech): A Systematic Literature Review,” *Information*, vol. 11, no. 12, 2020, doi: 10.3390/info11120590.

[3] M. Williams, M. F. Yussuf, and A. O. Olukoya, “Machine Learning For Proactive Cybersecurity Risk Analysis And Fraud Prevention In Digital Finance Ecosystems,” no. 125, pp. 160–177, 2021.

[4] K. Najaf, M. I. Mostafiz, and R. Najaf, “Fintech firms and banks' sustainability: Why cybersecurity risk matters?,” *Int. J. Financ. Eng.*, vol. 08, no. 02, Jun. 2021, doi: 10.1142/S2424786321500195.

[5] S. Mehrban *et al.*, “Towards Secure FinTech: A Survey, Taxonomy, and Open Research Challenges,” *IEEE Access*, vol. 8, pp. 23391–23406, 2020, doi: 10.1109/ACCESS.2020.2970430.

[6] N. Kandregula, "Leveraging Artificial Intelligence for Real-Time Fraud Detection in Financial Transactions: A Fintech Perspective," *World J. Adv. Res. Rev.*, vol. 3, no. 3, pp. 115–127, 2019.

[7] D. Narsina, "The Integration of Cybersecurity, IoT, and Fintech: Establishing a Secure Future for Digital Banking," *NEXG AI Rev. Am.*, vol. 1, no. 1, pp. 119–134, 2020.

[8] M. Qasaimeh, R. A. Hammour, M. B. Yassein, R. S. Al-Qassas, J. A. L. Torralbo, and D. Lizcano, "Advanced security testing using a cyber-attack forecasting model: A case study of financial institutions," *J. Softw. Evol. Process*, vol. 34, no. 11, Nov. 2022, doi: 10.1002/smr 2489.

[9] R. Akintoye, O. Ogunode, M. Ajayi, and A. A. Joshua, "Cyber Security and Financial Innovation of Selected Deposit Money Banks in Nigeria," *Univers. J. Account. Financ.*, vol. 10, no. 3, pp. 643–652, May 2022, doi: 10.13189/ujaf.2022.100302.

[10] F. Lacruz and J. Saniie, "Applications of Machine Learning in Fintech Credit Card Fraud Detection," in *2021 IEEE International Conference on Electro Information Technology (EIT)*, 2021, pp. 1–6. doi: 10.1109/EIT51626.2021.9491903.

[11] R. Ochoa, D. Ticse, E. Herrera, and J. Vargas, "Ransomware scenario oriented financial quantification model for the financial sector," in *2021 IEEE Sciences and Humanities International Research Conference (SHIRCON)*, 2021, pp. 1–4. doi: 10.1109/SHIRCON53068.2021.9652252.

[12] H. M. K. Al Duhaidahawi, J. Zhang, M. S. Abdulreda, M. Sebai, and S. Harjan, "Analysing the effects of FinTech variables on cybersecurity: Evidence form Iraqi Banks," *Int. J. Res. Bus. Soc. Sci. (2147- 4478)*, vol. 9, no. 6, pp. 123–133, 2021, doi: 10.20525/ijrbs.v9i6.914.

[13] A. Pattabhi, "Deep Learning for Cyber Risk Management in Financial Services: A Case Study of Data Breach Prediction," *Int. J. Emerg. Res. Eng. Technol.*, vol. 1, no. 2, pp. 37–45, 2020, doi: 10.63282/3050-922x.ijeret-v1i2p105.

[14] A. Makolo and T. Adeboye, "Credit Card Fraud Detection System Using Machine Learning," *Int. J. Inf. Technol. Comput. Sci.*, vol. 13, no. 4, pp. 24–37, 2021, doi: 10.5815/ijitcs 2021.04.03.

[15] H. Najadat, O. Altiti, A. A. Aqouleh, and M. Younes, "Credit Card Fraud Detection Based on Machine and Deep Learning," in *2020 11th International Conference on Information and Communication Systems (ICICS)*, 2020, pp. 204–208. doi: 10.1109/ICICS49469.2020.239524.

[16] K. I. Alkhatib, A. I. Al-Aiad, M. H. Almahmoud, and O. N. Elayan, "Credit Card Fraud Detection Based on Deep Neural Network Approach," *2021 12th Int. Conf. Inf. Commun. Syst. ICICS 2021*, no. June, pp. 153–156, 2021, doi: 10.1109/ICICS52457.2021.9464555.

[17] S. Lei, K. Xu, Y. Huang, and X. Sha, "An XGBoost-based system for financial fraud detection," *E3S Web Conf.*, vol. 214, pp. 1–4, 2020, doi: 10.1051/e3sconf/202021402042.

[18] U. Noor, Z. Anwar, T. Amjad, and K.-K. R. Choo, "A machine learning-based FinTech cyber threat attribution framework using high-level indicators of compromise," *Futur. Gener. Comput. Syst.*, vol. 96, pp. 227–242, Jul. 2019, doi: 10.1016/j.future.2019.02.013.