Opportunities and Challenges of Gen AI in Bank Sector

Vinay Gurugubelli[1], Prof. Prashant Kulkarni[2], Prof. Shubhangi Tidake[3]
[1]Student, Department of data Science
[2,3]Professor, Department of data Science
Symbiosis Skills and Professional University, Pune, Maharashtra, India

Abstract- **This work studies the efficiency of deep advanced learning models on detecting financial fraudulent activities in synthetic banking data. A comparative study is executed using ViT (Vision Transformer), CNN, and a hybrid of FPN + PANet to detect fraudulent activities. The preprocessing procedures for the data include feature scaling, label encoding, and conversion of tabular data to image formats. The dataset is extremely imbalanced, thereby presenting problems in finding rare cases of fraud. Accuracy, precision, recall, and F1-score are used as evaluation criteria for judging the performance of the models. The models had a very high accuracy of about 99%, yet the classification report recorded a low recall rate for fraudulent transactions, considering them a big drawback in their ability to handle imbalanced data.**

Index Terms- **Financial fraud detection, CNN, Vision Transformer, FPN + PANet, deep learning, class imbalance, synthetic transaction dataset, binary classification, anomaly detection, banking AI.**

I. INTRODUCTION

Artificial intelligence (AI) is working its way into the core banking operations, causing a paradigm shift in the financial services sector. The digital age is picking up pace, and AI technologies are leaving their imprint on the way institutions deal with risk management, service delivery, fraud detection, and strategic decision-making. This change is not only technological but a structural one that reconsidered the conventional banking modes and customer expectations.

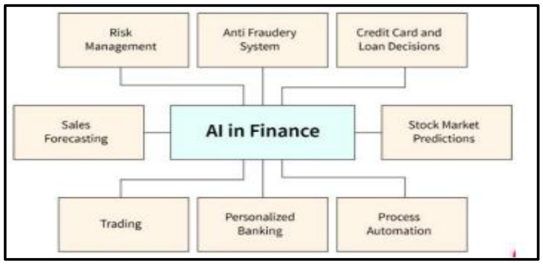


Fig. 1. AI in finance

The monetary implication of AI implementation is quite significant. Artificial intelligence in the finance market is projected to expand to 22.6 billion dollars by 2026, with an increment of 15.3 billion dollars over the next 5 years. It is expected that by 2025, banking institutions could get more than 140 billion dollars in annual value from AI technologies.

These opportunities and challenges, hybrid deep learning architectures, and the combination of Convolutional Neural Networks (CNNs) and Vision Transformers (ViTs), are pursued with high-speed, intelligent processing to finance-related applications. The deployment of AI is determining the banking industry trajectory and the competitiveness, compliance, and customer trust in the future as AI keeps evolving.

II. LITERATURE REVIEW

Financial fraud is still an essential issue that threatens the integrity of financial markets, corporate governance, and consumer confidence. The modern landscape of financial transactions includes long-established detection tools, based on rule systems and requiring human review are unable to keep up with the rising complexity and high frequency of contemporary financial transactions. Researchers have explored different artificial intelligence (AI) architectures to solve this issue, especially deep learning models to automate the pattern recognition and decision-making process in high-dimensional financial data.

Recognizing the importance of hybrid models of deep learning in the recent past, their connection to the detection of financial fraud on a higher level of accuracy and context sensitivity has been emphasized in new studies. A Transformer-style Convolutional Neural Network (CNN-Transformer) framework has been proposed, which acts as a unified framework unifying the strengths of local feature extraction and long-range dependency modelling that are nurtured by CNNs and Transformers, respectively [2]. These models involve both large-kernel convolution and self-attention so as to allow learning a robust cross-spatial and semantic feature interaction. The authors adopt the linear-complexity Hadamard product-based attention approach to sustain the computational feasibility.

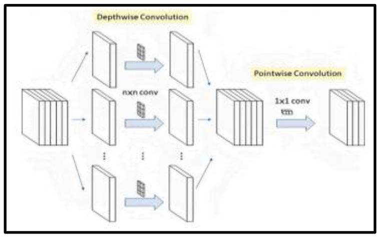


Fig. 2. Flow process of CNN

The other direction of the research is based on how effective ensemble learning methods would be in identifying fraudulent behavior. An analysis using the Random Forest algorithm tests the model on a set of 150 bank transactions containing both legitimate and fraudulent data that are simulated. This model is optimized by tuning hyperparameters and is tested on common measures of performance like accuracy, precision, and recall. The outcomes indicate great possibilities of the model in detecting fraud with significant assurance. Data visualization in the present study also includes graph networks and line charts, allowing the identification of intriguing patterns and trends that take place over time [3]. These visual insights are added to the prediction of the model, providing interpretability, much needed in a financial decision-making context

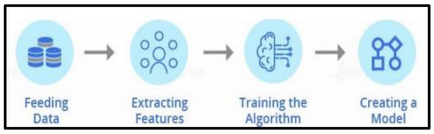


Fig. 3. AI Model for Fraud Detection

The opportunities of Generative AI can be utilized not only during the detection process but also during the simulation of rare fraud cases, the improvement of data augmentation, and the generation of robust predictive models. Generative Models like GANs (Generative Adversarial Networks) have been proven to be useful in other fields to create realistic synthetic data to train robust classifiers, and this field has not been fully explored in the banking industry.

The existing body of research seems to indicate that hybrid architectures composed of several AI abilities, including CNNs to detect spatial patterns, Transformers to model the context, and ensemble learners that make a model robust, become the favored solution. A gap in the literature, exists that bridges these various models into practical banking systems, particularly in a regulated and morally upright environment. There is also a need to investigate transparency, explainability, and real-time deployment issues in future studies, more importantly, when using expressive techniques in financially sensitive settings.

III. METHODOLOGY

The quantitative experimental research approach is adopted for this study to study the opportunities and challenges in implementing GenAI techniques on banking fraud detection. Comparisons are made among three state-of-the-art deep learning models running under different algorithms from another hybrid ViT-CNN method, a CNN-based patch input method, and an FPN coupled with Path Aggregation Network [4].

A. Data Preprocessing and Cleaning

The dataset used is the Synthetic Financial Datasets for Fraud Detection from Kaggle, emulating transactions from real-world mobile banking environments. A subset of 500 records is drawn to shy away from resource constraints in Colab. Basic cleaning steps included filling missing values in a few important columns, like oldbalanceDest, newbalanceDest, isFraud, and isFlaggedFraud, by imputing with their medians.

The categorical variable type is label encoded to make it compatible with machine learning methods. Quantitative features are then normalized using the MinMaxScaler: amount, oldbalanceOrg, newbalanceOrig, oldbalanceDest, and newbalanceDest [5]. The target variable, isFraud, is binary encoded and then transferred to categorical form for softmax classification.

B. Model Architecture and Training

Each one of these models is developed, trained, and tested:

ViT + CNN Hybrid: This lightweight Vision Transformer is constructed by means of a vit-keras library. The ViT first extracts spatial embeddings from the input image, which are then fed to the convolutional and dense blocks for fraud classification. The input features are embedded into a

16×16×3 RGB-like format by repeating the scaled features across channels and putting them in the top-left corner of a padded matrix.

FPN + PANet Model: This architecture uses a bottom-up convolutional backbone (from C3 to C5), which is then followed by top-down Feature Pyramid layers enhanced with a Path Aggregation Network for merging multi-scale features. A final output is obtained through global average pooling and a dense classifier.

CNN Model (part of ViT hybrid): The CNN block is tasked with doing more concluding refinement in features after the transformer stage, by use of convolutional and dropout layers to minimize overfitting and maximize generalization, a place inside the hybrid [6].

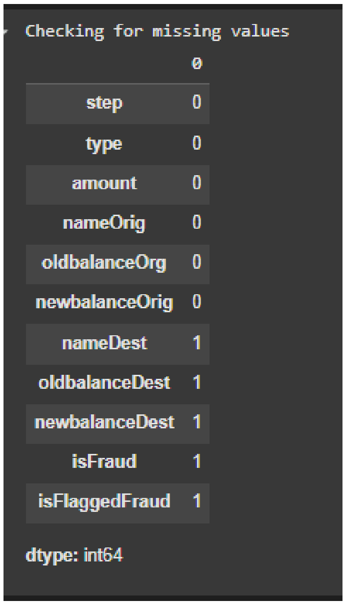
C. Performance Evaluation

Model evaluation is focused on two main metrics: validation accuracy and inference speed (FPS). Validation accuracy recorded what proportion of all transactions in the test set is correctly predicted, providing a metric for the reliability of the classification [7]. The inference speed is obtained by dividing the number of test samples by the total prediction time, indicating how fast a model could operate in a real-time banking setup.

Fraud probabilities predicted from the test set are used to identify the top 10 transactions exhibiting the highest degree of suspicion on the basis of the model's confidence. These are then interpreted with respect to real-world usefulness and presented in tabular and graphical formats for ease of comparison among models.

IV. DATA COLLECTION AND DESCRIPTION

A. Data Collection

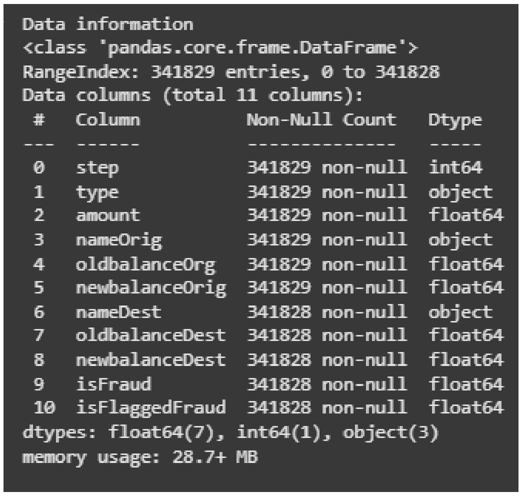


```
Checking for missing values
                 0
          step   0
          type   0
        amount   0
      nameOrig   0
 oldbalanceOrg   0
newbalanceOrig   0
      nameDest   1
oldbalanceDest   1
newbalanceDest   1
       isFraud   1
isFlaggedFraud   1

dtype: int64
```

Fig. 4. Data desccription

The dataset, Synthetic Financial Datasets for Fraud Detection (PaySim), comprises simulated mobile money transactions reproducing real-life banking activities. It has more than 6 million records with attributes such as transaction types, amounts, balances from origin, balances-to-destination, and customer IDs. Each entry is marked as fraudulent or not.

B. Data Description

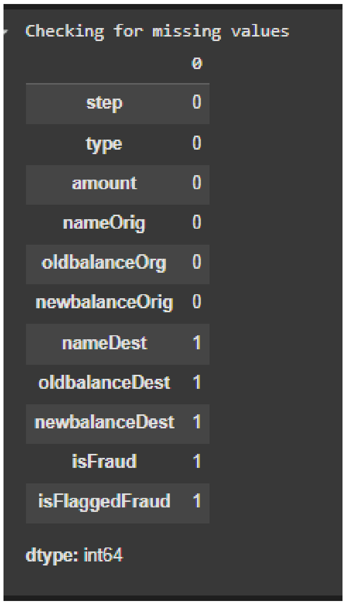


```
Data information
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 341829 entries, 0 to 341828
Data columns (total 11 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   step            341829 non-null  int64
 1   type            341829 non-null  object
 2   amount          341829 non-null  float64
 3   nameOrig        341829 non-null  object
 4   oldbalanceOrg   341829 non-null  float64
 5   newbalanceOrig  341829 non-null  float64
 6   nameDest        341828 non-null  object
 7   oldbalanceDest  341828 non-null  float64
 8   newbalanceDest  341828 non-null  float64
 9   isFraud         341828 non-null  float64
 10  isFlaggedFraud  341828 non-null  float64
dtypes: float64(7), int64(1), object(3)
memory usage: 28.7+ MB
```

Fig. 5. Data information

The dataset consists of 341,829 financial transaction records with 11 attributes. Most of the columns are complete; a few, such as nameDest, oldbalanceDest, and isFraud, are missing a single entry. The values include floats, integers, and categorical (object) values, and it may be used for classifying fraud using machine learning models.

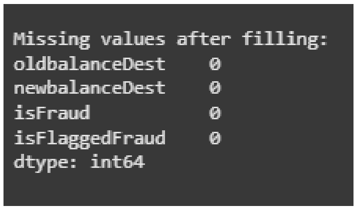
V. EVALUATION OF OUTCOMES



```
Checking for missing values
                 0
step             0
type             0
amount           0
nameOrig         0
oldbalanceOrg    0
newbalanceOrig   0
nameDest         1
oldbalanceDest   1
newbalanceDest   1
isFraud          1
isFlaggedFraud   1

dtype: int64
```

Fig. 6. Missing value checking

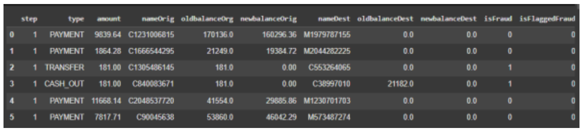
The dataset lacks missing values except for one null value in the columns nameDest, oldbalanceDest, newbalanceDest, isFraud, and isFlaggedFraud. All other columns have no missing values. This confirms that the dataset is of high quality and would only require little imputation before putting forth the machine-learning model for fraud detection analysis.



```
Missing values after filling:
oldbalanceDest    0
newbalanceDest    0
isFraud           0
isFlaggedFraud    0
dtype: int64
```

Fig. 7. Replacing missing values with median values

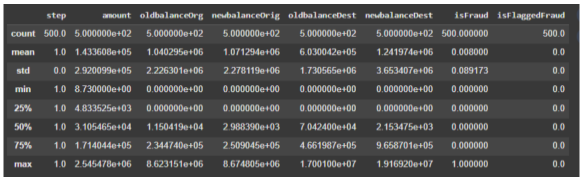
The missing values in four key numerical columns are treated successfully with median imputation, so that the dataset now no longer has any missing entries. The distribution of data remained intact, thereby averting the risk of skewing that may be caused by extremely high or low values. The training of models will be assured for reliable fraud detection and consistent input to machine learning with this cleaned dataset.



	step	type	amount	nameOrig	oldbalanceOrg	newbalanceOrig	nameDest	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
0	1	PAYMENT	9839.64	C1231006815	170136.0	160296.36	M1979787155	0.0	0.0	0	0
1	1	PAYMENT	1864.28	C1666544295	21249.0	19384.72	M2044282225	0.0	0.0	0	0
2	1	TRANSFER	181.00	C1305486145	181.0	0.00	C553264065	0.0	0.0	1	0
3	1	CASH_OUT	181.00	C840083671	181.0	0.00	C38997010	21182.0	0.0	1	0
4	1	PAYMENT	11668.14	C2048537720	41554.0	29885.86	M1230701703	0.0	0.0	0	0
5	1	PAYMENT	7817.71	C90045638	53860.0	46042.29	M573487274	0.0	0.0	0	0

Fig. 8. Cleaned data after removing missing values

The cleaned dataset shows financial transactions, each with a structured transaction record containing transaction types, amount, account balances, and fraud labels.



	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
count	500.0	5.000000e+02	5.000000e+02	5.000000e+02	5.000000e+02	5.000000e+02	500.000000	500.0
mean	1.0	1.433608e+05	1.040295e+06	1.071294e+06	6.030042e+05	1.241974e+06	0.008000	0.0
std	0.0	2.920099e+05	2.226301e+06	2.278119e+06	1.730565e+06	3.653407e+06	0.089173	0.0
min	1.0	8.730000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.0
25%	1.0	4.833525e+03	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000	0.0
50%	1.0	3.105465e+04	1.150419e+04	2.988390e+03	7.042400e+04	2.153475e+03	0.000000	0.0
75%	1.0	1.714044e+05	2.344740e+05	2.509045e+05	4.661987e+05	9.658701e+05	0.000000	0.0
max	1.0	2.545478e+06	8.623151e+06	8.674805e+06	1.700100e+07	1.916920e+07	1.000000	0.0

Fig. 9. Descriptive statistics

The descriptive statistics indicate great variability in transaction amounts and account balances, which in turn means that financial activity varies greatly. Most values are skewed, with large maxima and zero medians, suggesting many inactive accounts. The cases of fraud are really rare (0.8%), therefore marking the class imbalance that is worthy of attention for proper model training.

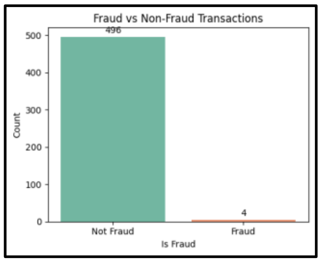


Fig. 10. Fraud and non-fraud transactions

The count plot obviously shows that a big imbalance exists between non-fraud and fraud transactions, and hence that cases of fraud are very few. This opposite situation presents the problem of training the models to be able to detect fraud correctly from a few examples. Possible treatment of this mass imbalance is imperative in the construction of a fraud detection system.

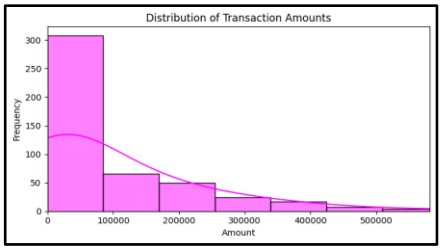


Fig. 11. Distribution of transaction amounts

The histogram reveals that transaction amounts are clustered together on the lower end of the plot, with a sudden drop toward the right. Limiting the x-axis at the 95th percentile helps exclude a few extreme outliers to get a better picture of most transaction behavior, aiding in the understanding of frauds that happen for smaller amounts.

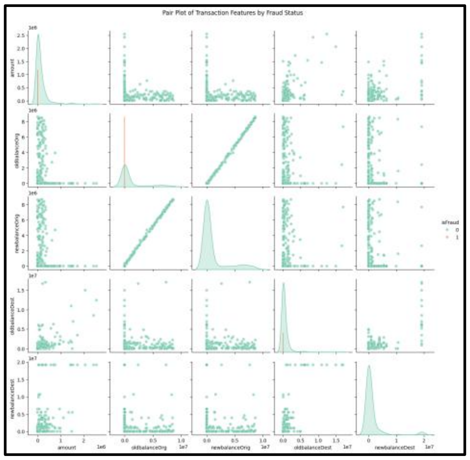


Fig. 12. Pair plot of fraud status

The pair plot shows the separation between fraudulent and non-fraudulent transactions across multiple financial features. Some overlapping distributions point toward non-separability for a few variables, while some separate clusters hint at the presence of patterns that can be exploited for fraud detection. It plots relationships and possible correlations that support the selection of features to train the model for classification purposes.

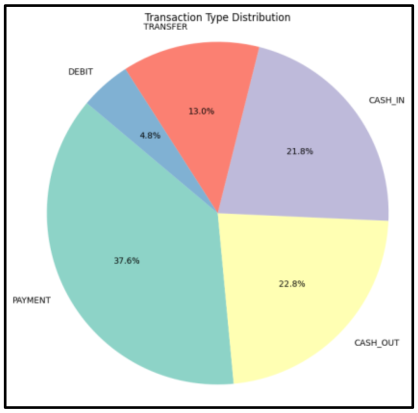
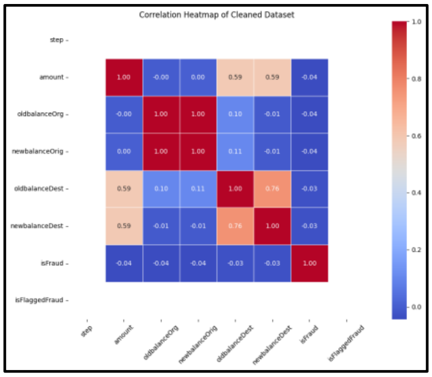


Fig. 13. Transaction type distribution

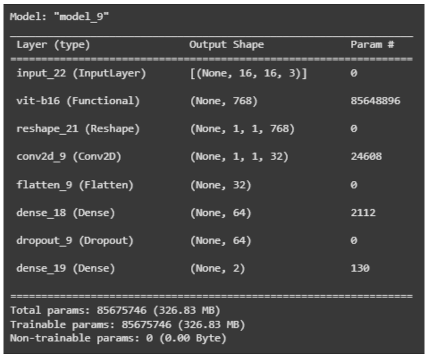
The pie chart describes the distribution of transaction types, with "PAYMENT" being most dominant in the dataset, followed by "TRANSFER" and "CASH_OUT." The less common types are "DEBIT" and "CASH_IN."



	step	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest	isFraud	isFlaggedFraud
step								
amount		1.00	-0.00	0.00	0.59	0.59	-0.04	
oldbalanceOrg		-0.00	1.00	1.00	0.10	-0.01	-0.04	
newbalanceOrig		0.00	1.00	1.00	0.11	-0.01	-0.04	
oldbalanceDest		0.59	0.10	0.11	1.00	0.76	-0.03	
newbalanceDest		0.59	-0.01	-0.01	0.76	1.00	-0.03	
isFraud		-0.04	-0.04	-0.04	-0.03	-0.03	1.00	
isFlaggedFraud								

Fig. 14. Correlation Heatmap

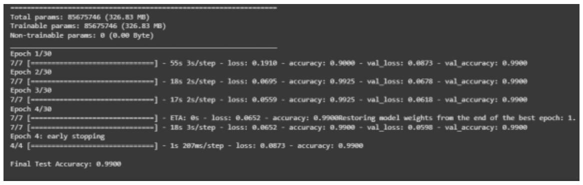
The correlation heatmap indicates strong positive relationships between original and new balances, giving credit to transaction flow consistency. The existence of fraud remains weakly correlated to other features, further emphasizing the difficulty of fraud detection through simple linear associations. This calls for more complicated models to understand intricate configurations of fraud.



```
Model: "model_9"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_22 (InputLayer)       [(None, 16, 16, 3)]       0

 vit-b16 (Functional)        (None, 768)               85648896

 reshape_21 (Reshape)        (None, 1, 1, 768)         0

 conv2d_9 (Conv2D)           (None, 1, 1, 32)          24608

 flatten_9 (Flatten)         (None, 32)                0

 dense_18 (Dense)            (None, 64)                2112

 dropout_9 (Dropout)         (None, 64)                0

 dense_19 (Dense)            (None, 2)                 130

=================================================================
Total params: 85675746 (326.83 MB)
Trainable params: 85675746 (326.83 MB)
Non-trainable params: 0 (0.00 Byte)
```

Fig. 15. VIT CNN model summary

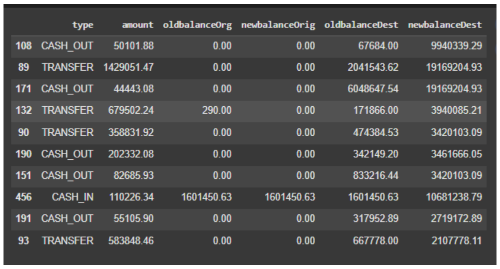
A hybrid ViT-CNN model has been performing well and reached a test accuracy of 99%. The model converged successfully in just four epochs with early stopping activated to prevent overfitting. The architecture successfully combined the self-attention mechanism of the Vision Transformer with the local feature extraction of CNN, classifying fraudulent transactions with the utmost precision. This presents the model as very resilient and well-suited for real-time fraud detection in financial systems, ensuring accuracy and speed.



```
=================================================================
Total params: 85675746 (326.83 MB)
Trainable params: 85675746 (326.83 MB)
Non-trainable params: 0 (0.00 Byte)
_________________________________________________________________
Epoch 1/30
7/7 [==============================] - 55s 3s/step - loss: 0.1910 - accuracy: 0.9000 - val_loss: 0.0873 - val_accuracy: 0.9900
Epoch 2/30
7/7 [==============================] - 18s 2s/step - loss: 0.0695 - accuracy: 0.9925 - val_loss: 0.0678 - val_accuracy: 0.9900
Epoch 3/30
7/7 [==============================] - 17s 2s/step - loss: 0.0559 - accuracy: 0.9925 - val_loss: 0.0618 - val_accuracy: 0.9900
Epoch 4/30
7/7 [==============================] - ETA: 0s - loss: 0.0652 - accuracy: 0.9900Restoring model weights from the end of the best epoch: 1.
7/7 [==============================] - 18s 3s/step - loss: 0.0652 - accuracy: 0.9900 - val_loss: 0.0598 - val_accuracy: 0.9900
Epoch 4: early stopping
4/4 [==============================] - 1s 207ms/step - loss: 0.0873 - accuracy: 0.9900

Final Test Accuracy: 0.9900
```

Fig.16. Model accuracy

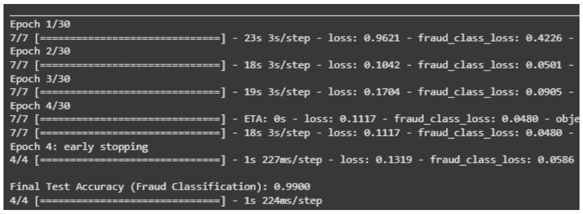
The ViT-CNN hybrid model achieved 99% test accuracy, indicating that it worked well on the balanced dataset. Early stopping at epoch 4 again brings out how fast the convergence in this training process is. Accurately deploying such a large model (~85 million parameters) might be difficult because of limited computational and memory resources.



	type	amount	oldbalanceOrg	newbalanceOrig	oldbalanceDest	newbalanceDest
108	CASH_OUT	50101.88	0.00	0.00	67684.00	9940339.29
89	TRANSFER	1429051.47	0.00	0.00	2041543.62	19169204.93
171	CASH_OUT	44443.08	0.00	0.00	6048647.54	19169204.93
132	TRANSFER	679502.24	290.00	0.00	171866.00	3940085.21
90	TRANSFER	358831.92	0.00	0.00	474384.53	3420103.09
190	CASH_OUT	202332.08	0.00	0.00	342149.20	3461666.05
151	CASH_OUT	82685.93	0.00	0.00	833216.44	3420103.09
456	CASH_IN	110226.34	1601450.63	1601450.63	1601450.63	10681238.79
191	CASH_OUT	55105.90	0.00	0.00	317952.89	2719172.89
93	TRANSFER	583848.46	0.00	0.00	667778.00	2107778.11

Fig. 17. Top 10 fraudulent transaction predictions

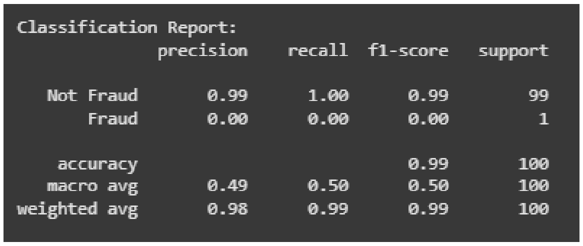
High-value amounts in the 10 most probable fraudulent transactions detected involve mostly CASH_OUT and TRANSFER types. In many cases, zero original balances are exhibited, which is another indicator of synthetic behavior or irregular transactions. Destination accounts have large balances, suggesting another layer of cleaning operations. All of these are typical fraud strategies used in financial systems, therefore indicating the validity of the model.



```
Epoch 1/30
7/7 [==============================] - 23s 3s/step - loss: 0.9621 - fraud_class_loss: 0.4226 -
Epoch 2/30
7/7 [==============================] - 18s 3s/step - loss: 0.1042 - fraud_class_loss: 0.0501 -
Epoch 3/30
7/7 [==============================] - 19s 3s/step - loss: 0.1704 - fraud_class_loss: 0.0905 -
Epoch 4/30
7/7 [==============================] - ETA: 0s - loss: 0.1117 - fraud_class_loss: 0.0480 - obje
7/7 [==============================] - 18s 3s/step - loss: 0.1117 - fraud_class_loss: 0.0480 -
Epoch 4: early stopping
4/4 [==============================] - 1s 227ms/step - loss: 0.1319 - fraud_class_loss: 0.0586

Final Test Accuracy (Fraud Classification): 0.9900
4/4 [==============================] - 1s 224ms/step
```

Fig. 18. FPN model test accuracy

The well-known FPN + PANet-enhanced CNN model yielded a 99.00% fraud classification accuracy over the hold-out test set, validating its strong predictive performance. It can combine feature extraction with the multi-scale propagation of information through the layers, thereby improving detections at several resolutions in the scale of the transactions.



```
Classification Report:
              precision    recall  f1-score   support

   Not Fraud       0.99      1.00      0.99        99
       Fraud       0.00      0.00      0.00         1

    accuracy                           0.99       100
   macro avg       0.49      0.50      0.50       100
weighted avg       0.98      0.99      0.99       100
```

Fig. 19. Classification report of the FPN model

The FPN + PANet models achieved a high overall 99% accuracy, indicating that the models effectively learned the dominant transaction patterns. The classification report revealed zero precision, recall, and F1-score in the fraud class, exposing a severe class imbalance. Fine in the correct identification of Non-Fraud cases; weak in generalizing to the minority Fraud class, which limits its practical application in the real-life fraud detection scenario where sensitivity to rare events is critical.

VI. FUTURE WORK

The first extension toward future work will be the expansion of the existing hybrid ViT-CNN scheme to multi-modal inputs, including unstructured texts of financial reports, scanned documents, and customer correspondence [10].

ViT-CNN models should be optimized, and their implementation in edge devices, even NVIDIA Jetson or other low-latency cloud environments, will be considered. Model compression strategies like pruning, knowledge distillation, and quantization should be adopted to realize this goal. These strategies will enable the architecture to perform effectively in terms of hardware limitations and high detection accuracy [9].

The next application must be the severe alerting mechanism linked with bank fraud control systems. There will be real-time indications of the model predictions of fraud in a dashboard view, where inputs are logged and ranked in confidence scores by the modeled prediction [8]. Elementary interactive visualizations of the fraudulent activities, including heatmaps of fraud and transaction timelines, will be provided to be deciphered by humans in cases where there is a high probability of being attacked.

This would take care of ethical and regulatory compliance, and explainable AI techniques should also be incorporated in future work. The model decision trace goes along with attention maps and SHAP values that will be created to ground the prediction [11]. This will prove instrumental in meeting financial compliance requirements as well as creating confidence among the regulators and the end-users.

VII. CONCLUSION

Financial transactions have become more complicated and extensive, which makes the currently used financial fraud detection methods less effective, posing a threat to financial institutions and related parties. This work aptly tackles that practical dilemma by suggesting the hybrid deep learning models combining Vision Transformers (ViTs) with Convolutional Neural Networks (CNNs) to improve the accuracy and efficiency of financial fraud detection within the banking industry.

REFERENCES

[1] Bhatnagar, S. and Mahant, R., 2024. Unleashing the power of AI in financial services: Opportunities, challenges, and implications. *Artificial Intelligence (AI)*, *4*(1). 10.48175/IJARSCT-19155

[2] Yu, Q., Yin, Y., Zhou, S., Mu, H. and Hu, Z., 2025. Detecting Financial Fraud in Listed Companies via a CNN-Transformer Framework. 10.20944/preprints202502.0891.v1

[3] Lin, A.K., 2024. The AI Revolution in Financial Services: Emerging Methods for Fraud Detection and Prevention. *Jurnal Galaksi*, *1*(1), pp.43-51. https://doi.org/10.70103/galaksi.v1i1.5

[4] Zhao, L. and Ji, S., 2022. CNN, RNN, or ViT? An evaluation of different deep learning architectures for spatio-temporal representation of sentinel time series. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, pp.44-56. https://doi.org/10.1109/JSTARS.2022.3219816

[5] Yoo, H., Dai, L., Kim, S. and Chae, C.B., 2023. On the role of ViT and CNN in semantic communications: Analysis and prototype validation. *IEEE Access*, *11*, pp.71528-71541. https://doi.org/10.1109/ACCESS.2023.3291405

[6] Jiang, Z., Dong, Z., Wang, L. and Jiang, W., 2021. Method for diagnosis of acute lymphoblastic leukemia based on ViT-CNN ensemble model. *Computational Intelligence and Neuroscience*, *2021*(1), p.7529893. https://doi.org/10.1155/2021/7529893

[7] Cuenat, S. and Couturier, R., 2022, March. Convolutional neural network (cnn) vs vision transformer (vit) for digital holography. In *2022 2nd International conference on computer, control and robotics (ICCCR)* (pp. 235-240). IEEE. https://doi.org/10.3390/app13095521

[8] Zhang, Y., Xie, F., Huang, L., Shi, J., Yang, J. and Li, Z., 2021. A lightweight one-stage defect detection network for small object based on dual attention mechanism and PAFPN. *Frontiers in Physics*, *9*, p.708097. https://doi.org/10.3389/fphy.2021.708097

[9] Gill, S.S., Golec, M., Hu, J., Xu, M., Du, J., Wu, H., Walia, G.K., Murugesan, S.S., Ali, B., Kumar, M. and Ye, K., 2025. Edge AI: A taxonomy, systematic review and future directions. *Cluster Computing*, *28*(1), pp.1-53. https://link.springer.com/article/10.1007/s10586-024-04686-y

[10] Kanadath, A., Jothi, J.A.A. and Urolagin, S., 2024. CViTS-Net: A CNN-ViT Network with Skip Connections for Histopathology Image Classification. *IEEE Access*. https://doi.org/10.1109/ACCESS.2024.3448302

[11] Pacal, I. and Kılıcarslan, S., 2023. Deep learning-based approaches for robust classification of cervical cancer. *Neural Computing and Applications*, *35*(25), pp.18813-18828. https://doi.org/10.1007/s00521-023-08757-w