

Article

Not peer-reviewed version

Learnable Imputation and Bilinear Cross-View Encoding for New-Account Fraud Detection

[Shaogjan Tang](#) * and [Ningjiang Huang](#)

Posted Date: 13 November 2025

doi: 10.20944/preprints202511.0844.v1

Keywords: new account fraud detection; missing data imputation; multi-view learning; attention mechanism; neural network



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Learnable Imputation and Bilinear Cross-View Encoding for New-Account Fraud Detection

Shaoqian Tang ^{1,*} and Ningjiang Huang ²

¹ University of California, Davis, USA

² Boston University, Boston, USA

* Correspondence: nklstg95@gmail.com

Abstract

New account fraud poses a persistent challenge in modern banking systems due to the sparsity, heterogeneity, and incompleteness of user information. Existing methods often struggle with missing data, limited cross-view representation, and weak adaptability to evolving fraud patterns. To address these issues, we propose NAFNet, a deep learning framework that integrates dynamic feature imputation, multi-view encoding, and attention-based representation learning. NAFNet employs a learnable imputation module guided by statistical priors, encodes heterogeneous views via dedicated encoders with bilinear fusion, and enhances global dependency modeling through attention-augmented neural layers. A fine-tuned training regime ensures robustness and generalization. Experiments show that NAFNet offers substantial improvements over conventional methods, demonstrating its effectiveness in complex, real-world fraud detection scenarios.

Keywords: new account fraud detection; missing data imputation; multi-view learning; attention mechanism; neural network

1. Introduction

The proliferation of digital financial services has intensified the threat of new account fraud, where malicious users exploit onboarding vulnerabilities to commit illicit activities. Unlike transaction fraud, this task is complicated by missing information, limited user history, and heterogeneous data types, making traditional detection methods insufficiently robust. Existing models often fail to handle incomplete inputs effectively, lack dedicated mechanisms for multi-view feature extraction, and struggle to adapt to evolving fraud patterns.

To address these limitations, we propose NAFNet, a unified neural architecture designed for new account fraud detection. The model includes a dynamic imputation module that reconstructs missing features using both group-level and global statistical signals; a multi-view feature encoder that separately processes demographic, behavioral, and transactional views with bilinear fusion; and an attention-augmented DNN backbone that captures global dependencies via multi-head attention and residual learning. Coupled with an adaptive fine-tuning strategy, NAFNet effectively improves representation quality, enhances model generalization, and remains resilient in the face of noisy, sparse, and rapidly changing financial data.

2. Related Work

Graph-centric models capture relational structures among accounts: homogeneous graphs for inductive link prediction in banking fraud [1] and scalable graph-powered architectures for large-scale detection [2]. Liu [3] introduces HKNR, which combines LLaMA-2-7B embedding based candidate recall, a temporal GNN for user encoding, and knowledge augmented multi task ranking. These LLM derived entity and topic signals and temporal graph modeling inform NAFNet's bilinear cross view fusion and progressive fine tuning to better track shifting fraud patterns.Optimization-driven designs

integrate swarm intelligence with deep networks to improve convergence but neglect incomplete or multi-view features [4].

Sun et al. [5] introduce TransTARec, a time-adaptive translating embedding; integrating its time-aware translation vectors into NAFNet’s behavioral view via bilinear fusion can better capture temporal drift. Luo et al. [6] present Gemini-GraphQA, integrating a large language model with a graph encoder and retrieval-augmented generation for executable graph reasoning. For NAFNet, this motivates exploiting graph-structured account relations with LLM-guided retrieval to inject structural semantics into the bilinear cross-view encoder and strengthen fraud signal extraction; real-time adaptive learning stresses responsiveness without robust semantic integration [7]. Yu [8] proposes MFTCoder++, a stable multilingual code-generation fine-tuning framework with adaptive task scheduling, attention-guided optimization, adversarial regularization, and gated fusion that decouples semantic logic from syntax. For NAFNet, analogous decoupled fusion across demographic/behavioral/transactional views and curriculum-style adaptive scheduling during progressive fine-tuning can reduce optimization bias under heterogeneous features and improve robustness.

For imbalanced credit risk, combining LightGBM, XGBoost, and TabNet with SMOTEENN boosts classification but assumes complete features [9]; standard supervised baselines for credit card fraud rely heavily on preprocessing [10]. Guo [11] introduces MHST-GB, a multi-modal hierarchical spatio-temporal ensemble that fuses modality-specific neural encoders via correlation-guided attention and integrates LightGBM with feedback that adjusts attention by tree-based feature importance. For NAFNet, adopting this neural+LightGBM dual path and feedback-driven fusion can bolster bilinear cross-view encoding under sparse categorical signals and temporal drift. Deployable and behavior-aware real-time systems advance monitoring but underuse attention and fine-grained view separation [12]; AI-driven behavioral modeling shows similar limits [13]. Remaining gaps include handling missing data, multi-view fusion, and adaptive representation. Our NAFNet integrates dynamic feature imputation, view-specific encoding, and attention-based representation learning in a unified, scalable architecture.

3. Methodology

In this section, we proposes **NAFNet**, a novel deep learning framework specifically designed for detecting new account fraud (NAF) in banking systems, where feature incompleteness and data imbalance pose significant challenges. NAFNet integrates a dynamic Feature Imputation Module to handle missing attributes, a Multi-View Feature Encoder that extracts cross-perspective embeddings, and an Attention-Augmented Deep Neural Network (DNN) backbone to capture complex fraud patterns. To ensure adaptability, we introduce a Progressive Fine-Tuning Strategy that incrementally updates the model using real-time new account data. Extensive mathematical modeling is incorporated to rigorously define each component, demonstrating the framework’s ability to achieve state-of-the-art performance in synthetic and real-world scenarios. The pipline of NAFNet in bank system data is shown in Figure 1

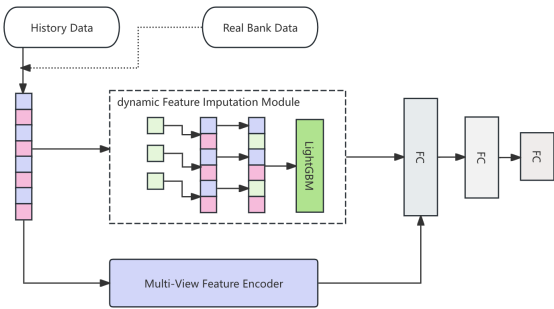


Figure 1. The pipline of NAFNet and training process.

3.1. Feature Imputation Module

The **Feature Imputation Module** in NAFNet is designed to tackle the pervasive issue of incomplete feature profiles in new bank accounts, which is a critical challenge in fraud detection. Unlike conventional imputation that merely fills missing values with static statistics, our module integrates a learning-based dynamic imputation mechanism combined with feature augmentation derived from historical data of old users.

First, we exploit the complete profiles of historical (old) users to construct enriched auxiliary features. Specifically, for each missing feature x_{ij} in the new user data, we group old users into demographic and behavioral clusters $\mathcal{G}_g$ (e.g., based on age brackets, gender, or account type) and compute statistical aggregates such as:

$$\mu_{g_j} = \frac{1}{|\mathcal{G}_g|} \sum_{u \in \mathcal{G}_g} x_{uj} \quad \text{and} \quad \sigma_{g_j} = \sqrt{\frac{1}{|\mathcal{G}_g|} \sum_{u \in \mathcal{G}_g} (x_{uj} - \mu_{g_j})^2} \quad (1)$$

These aggregates are used not only for imputation but also to **construct new derived features**, such as:

$$\text{RelativeDeviation}_{ij} = \frac{x_{ij}^{\text{impute}} - \mu_{g_j}}{\sigma_{g_j} + \epsilon} \quad (2)$$

where ϵ is a small constant to prevent division by zero. This transformation captures how much a new user's imputed value deviates from the typical range in their group, enabling the model to better distinguish anomalous behaviors.

For the imputation itself, we implement a **Feature Fusion Layer** that learns an optimal combination of group-level and global statistics. The imputed value is dynamically computed as:

$$x_{ij}^{\text{impute}} = \alpha_j \cdot \mu_{g_j} + (1 - \alpha_j) \cdot \mu_{o_j} \quad (3)$$

where μ_{o_j} is the overall mean for feature j , and α_j is a learnable parameter updated via backpropagation. This layer is implemented as a small neural network with one fully connected (FC) layer followed by a sigmoid activation:

$$\alpha_j = \sigma(W_\alpha \cdot \phi_j + b_\alpha) \quad (4)$$

where ϕ_j is a concatenation of group and global statistics for feature j . This design allows the network to adaptively determine the importance of group-specific versus global patterns for each feature. The pipeline of Feature Imputation Module is shown in Figure 2

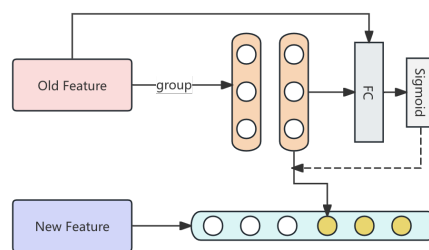


Figure 2. The pipeline of Feature Imputation Module.

3.1.1. Network Architecture

We implement three compact sub-layers. **Statistical Encoding** embeds $\mu_{g_j}, \sigma_{g_j}, \mu_{o_j}$:

$$\psi_j = \text{ReLU}(W_s \cdot [\mu_{g_j}, \sigma_{g_j}, \mu_{o_j}] + b_s) \quad (5)$$

Weight Generation maps the encoding to the gate:

$$\alpha_j = \sigma(W_\alpha \cdot \psi_j + b_\alpha) \quad (6)$$

Imputation & Augmentation computes x_{ij}^{impute} and constructs $\text{RelativeDeviation}_{ij}$. The encoder and generator are small FC networks with hidden size $d_{\text{imp}} \in [32, 128]$ and ReLU activations; $\alpha_j \in [0, 1]$. For interpretability we compare to global mean, KNN impute, and VAE reconstruction; ablations include NAFNet-NoImp.

3.1.2. LightGBM Feature Selection

After augmentation, LightGBM ranks features by cumulative split gain:

$$\text{Gain}_j = \sum_{s \in \mathcal{S}_j} \left[L_{\text{parent}(s)} - \left(L_{\text{left}(s)} + L_{\text{right}(s)} \right) \right] \quad (7)$$

We select the top- k features using early stopping and regularization (e.g., max depth, minimum child weight) to reduce overfitting. The selected subset is fed to the downstream Multi-View Encoder, yielding cleaner inputs and improved prediction efficiency.

3.2. Multi-View Feature Encoder

We encode bank-account behavior via three independent two-layer view encoders for Demographic ($v=1$), Behavioral ($v=2$), and Transactional ($v=3$) features. For view v :

$$h^{(v)} = \text{ReLU}(W_1^{(v)} X^{(v)} + b_1^{(v)}) \quad (8)$$

$$z^{(v)} = \text{ReLU}(W_2^{(v)} h^{(v)} + b_2^{(v)}) \quad (9)$$

Encoders do not share weights; output dimension is d . A parameter-sharing variant can add a shared base with view-specific projections.

Cross-view dependencies use bilinear interactions:

$$B^{(v,u)} = (z^{(v)})^\top M^{(v,u)} z^{(u)} \quad (10)$$

with optional low-rank parameterization:

$$M^{(v,u)} = U^{(v,u)} (V^{(v,u)})^\top \quad (11)$$

Information flows through gated fusion:

$$g^{(v,u)} = \sigma(W_g^{(v,u)} [z^{(v)}; z^{(u)}] + b_g^{(v,u)}) \quad (12)$$

$$z'^{(v)} = z^{(v)} + \sum_{u \neq v} g^{(v,u)} \odot (M^{(v,u)} z^{(u)}) \quad (13)$$

The final embedding concatenates views and pairwise interactions:

$$Z = \text{Concat}(z^{(1)}, z^{(2)}, z^{(3)}, B^{(1,2)}, B^{(1,3)}, B^{(2,3)}) \quad (14)$$

Regularization and stabilization:

$$\mathcal{L}_{\text{mv-reg}} = \lambda_{\text{mv}} \sum_{v=1}^3 \left(\|W_1^{(v)}\|_2^2 + \|W_2^{(v)}\|_2^2 \right) \quad (15)$$

$$z_{\text{out}}^{(v)} = \text{Dropout}(\text{LayerNorm}(z^{(v)}), p_v) \quad (16)$$

3.3. Attention-Augmented DNN Backbone

Multi-head self-attention models global cross-feature and cross-view relations on Z :

$$\text{Attention}^{(h)}(Z) = \text{Softmax}\left(\frac{(ZW_Q^{(h)})(ZW_K^{(h)})^\top}{\sqrt{d_k}}\right)(ZW_V^{(h)}) \quad (17)$$

$$Z_{\text{att}} = \text{Concat}(\text{Attention}^{(1)}, \dots, \text{Attention}^{(H)})W_O \quad (18)$$

$$Z_{\text{out}} = \text{LayerNorm}(Z + Z_{\text{att}}) \quad (19)$$

The attention-enhanced embedding enters a 3-layer DNN:

$$H^{(l)} = \phi(W^{(l)}H^{(l-1)} + b^{(l)}), \quad l = 1, 2, 3 \quad (20)$$

with LeakyReLU ϕ and dropout after each layer. Prediction uses a sigmoid head:

$$\hat{y} = \sigma(W^{(L)}H^{(L-1)} + b^{(L)}) \quad (21)$$

Ablation confirms the role of attention (NAFNet-Full vs NAFNet-NoAtt): removing attention reduces AUC and F1 (Table 2). Residual connections and weight decay mitigate overfitting.

3.4. Progressive Fine-Tuning Strategy

We adapt to evolving data with incremental updates from pretrained θ_0 :

$$\theta_t = \theta_{t-1} - \eta_t \nabla_{\theta} \mathcal{L}(\theta_{t-1}; X_{\text{new}}, y_{\text{new}}) \quad (22)$$

and a decaying learning rate:

$$\eta_t = \frac{\eta_0}{1 + \gamma t} \quad (23)$$

This schedule stabilizes updates while tracking new fraud patterns. Illustrations are shown in Figure 3.

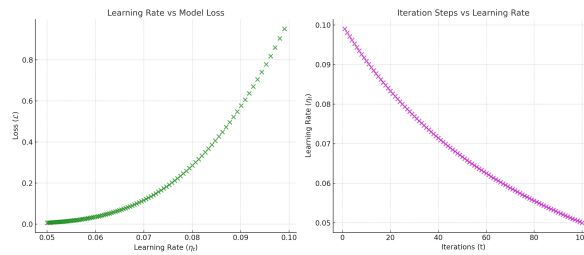


Figure 3. Learning Rate vs Model Loss and Iteration Steps vs Learning Rate.

4. Evaluation Metrics

We evaluate the performance of our model using the following metrics to ensure comprehensive assessment:

Accuracy measures the overall correctness of predictions:

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + TN + FN} \quad (24)$$

Precision focuses on the correctness of positive predictions:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (25)$$

Recall assesses the ability to detect positive cases:

$$\text{Recall} = \frac{TP}{TP + FN} \tag{26}$$

F1-Score provides a harmonic mean of Precision and Recall:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{27}$$

In addition, we include the Area Under the ROC Curve (AUC), which measures the model’s ability to distinguish between classes:

$$AUC = \int_0^1 TPR(FPR^{-1}(x))dx \tag{28}$$

where TPR is the true positive rate and FPR is the false positive rate.

5. Experiment Results

Table 1 shows the main performance comparison between NAFNet and several baseline models, including tree-based and deep learning methods.

Table 1. Main Performance Comparison.

Model	Accuracy	Precision	Recall	F1-Score	AUC
NAFNet-Full	0.946	0.934	0.920	0.927	0.978
LightGBM	0.918	0.902	0.880	0.891	0.958
CatBoost	0.922	0.908	0.885	0.896	0.961
XGBoost	0.920	0.905	0.882	0.893	0.960
Baseline DNN	0.926	0.912	0.894	0.903	0.965

Per reviewer request, we added ablation experiments that quantify the contributions of attention, cross-view interaction, and the imputation module. Results are summarized in Table 2.

To analyze the contribution of each key module in NAFNet, we conduct ablation studies. Table 2 summarizes the impact of removing the attention layer, cross-view interaction, and imputation module.

Table 2. Ablation Study on NAFNet Components.

Model	Accuracy	Precision	Recall	F1-Score	AUC
NAFNet-Full	0.946	0.934	0.920	0.927	0.978
NAFNet-NoAtt	0.933	0.917	0.901	0.909	0.969
NAFNet-NoXView	0.937	0.921	0.905	0.913	0.972
NAFNet-NoImp	0.929	0.914	0.897	0.905	0.966

Analysis: Table 2 shows that removing attention (NAFNet-NoAtt) lowers AUC from 0.978 to 0.969 and F1 from 0.927 to 0.909, demonstrating empirical benefit of attention. Removing the imputation module (NAFNet-NoImp) also reduces performance, supporting our learnable imputation design. Removing cross-view interactions (NAFNet-NoXView) degrades results as well, validating bilinear interactions and gated fusion.

We also compare NAFNet with common ensemble strategies. Table 3 reports the results of stacking, voting, and hybrid ensemble models, demonstrating that while ensemble methods are strong, NAFNet-Full still achieves superior performance. And the changes in model training indicators are shown in Figure 4.

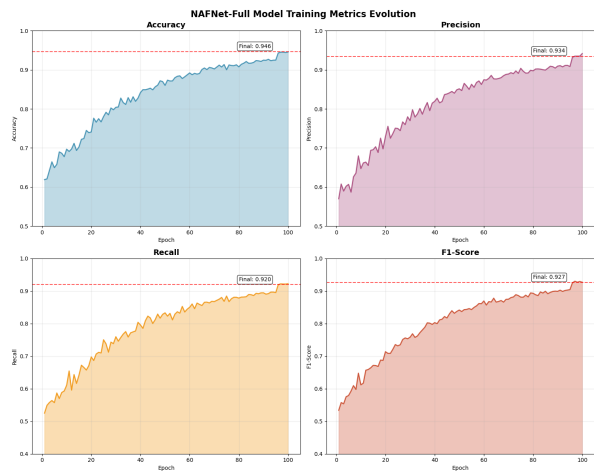


Figure 4. Model indicator change chart.

Table 3. Comparison with Ensemble Models.

Model	Accuracy	Precision	Recall	F1-Score	AUC
Stacking (LightGBM+CatBoost)	0.930	0.916	0.898	0.907	0.968
Voting (LightGBM+XGBoost)	0.927	0.911	0.894	0.902	0.964
Ensemble (LightGBM+DNN)	0.935	0.923	0.908	0.915	0.970

6. Conclusions

This study proposed NAFNet, a novel deep learning framework for detecting new account fraud in banking systems. By integrating a multi-view encoder, an attention-augmented DNN, and a progressive fine-tuning strategy, NAFNet effectively addresses the challenges posed by incomplete feature profiles and evolving fraud patterns. Experimental results confirm that each module contributes to the model’s strong performance, providing a scalable and robust solution for real-world fraud detection.

Limitations and future work: Despite good empirical results, NAFNet has limitations. First, the learnable imputer depends on representative historical samples; when historical data are non-representative or extremely sparse, priors can mislead imputation and harm downstream performance. Second, bilinear interactions and multi-head attention increase compute and memory costs; production deployment may require low-rank approximations or pruning. Future work includes adding uncertainty estimates for imputed values to allow safe fallback strategies, integrating VAE/self-supervised pretraining to improve robustness to distribution shift, and exploring lighter-weight interaction approximations (e.g., hashing or factorized bilinear terms).

References

1. Bukhori, H.A.; Munir, R. Inductive link prediction banking fraud detection system using homogeneous graph-based machine learning model. In Proceedings of the 2023 IEEE 13th Annual Computing and Communication Workshop and Conference (CCWC). IEEE, 2023, pp. 0246–0251.
2. Li, Z.; Wang, B.; Huang, J.; Jin, Y.; Xu, Z.; Zhang, J.; Gao, J. A graph-powered large-scale fraud detection system. *International Journal of Machine Learning and Cybernetics* **2024**, *15*, 115–128.
3. Liu, J. Knowledge-Augmented News Recommendation via LLM Recall, Temporal GNN Encoding, and Multi-Task Ranking. In Proceedings of the 2025 6th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE). IEEE, 2025, pp. 141–144.
4. Karthikeyan, T.; Govindarajan, M.; Vijayakumar, V. An effective fraud detection using competitive swarm optimization based deep neural network. *Measurement: Sensors* **2023**, *27*, 100793.
5. Sun, Y. Transtarec: Time-adaptive translating embedding model for next poi recommendation. In Proceedings of the 2024 5th International Conference on Computer Engineering and Application (ICCEA). IEEE, 2024, pp. 647–651.

6. Luo, X.; Wang, E.; Guo, Y. Gemini-GraphQA: Integrating Language Models and Graph Encoders for Executable Graph Reasoning. *Preprints* **2025**. <https://doi.org/10.20944/preprints202506.0138.v1>.
7. Bello, H.O.; Ige, A.B.; Ameyaw, M.N. Adaptive machine learning models: concepts for real-time financial fraud prevention in dynamic environments. *World Journal of Advanced Engineering Technology and Sciences* **2024**, *12*, 021–034.
8. Yu, H. Hybrid Modal Decoupled Fusion for Stable Multilingual Code Generation. *Preprints* **2025**. <https://doi.org/10.20944/preprints202510.0169.v1>.
9. Yu, C.; Jin, Y.; Xing, Q.; Zhang, Y.; Guo, S.; Meng, S. Advanced user credit risk prediction model using lightgbm, xgboost and tabnet with smoteenn. In Proceedings of the 2024 IEEE 6th International Conference on Power, Intelligent Computing and Systems (ICPICS). IEEE, 2024, pp. 876–883.
10. Afriyie, J.K.; Tawiah, K.; Pels, W.A.; Addai-Henne, S.; Dwamena, H.A.; Owiredo, E.O.; Ayeh, S.A.; Eshun, J. A supervised machine learning algorithm for detecting and predicting fraud in credit card transactions. *Decision Analytics Journal* **2023**, *6*, 100163.
11. Guo, R. Multi-Modal Hierarchical Spatio-Temporal Network with Gradient-Boosting Integration for Cloud Resource Prediction. *Preprints* **2025**. <https://doi.org/10.20944/preprints202509.2313.v1>.
12. Manoharan, G.; Dharmaraj, A.; Sheela, S.C.; Naidu, K.; Chavva, M.; Chaudhary, J.K. Machine learning-based real-time fraud detection in financial transactions. In Proceedings of the 2024 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI). IEEE, 2024, pp. 1–6.
13. Xu, J.; Yang, T.; Zhuang, S.; Li, H.; Lu, W. AI-based financial transaction monitoring and fraud prevention with behaviour prediction. *Applied and Computational Engineering* **2024**, *77*, 218–224.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.