Babylonian Journal of Networking
Vol.2026, **pp**. 26–31
DOI: https://doi.org/10.58496/BJN/2026/003; ISSN: 3006-5372
https://mesopotamian.press/journals/index.php/BJN

Research Article

A Cross-Modal Deep Learning Framework for Real-Time Digital Identity Threat Detection in Intelligent Network Environments

Tatiana G. Kalganova [1,*,]

[1] *Department of Electronic and Electrical Engineering College of Engineering, Design and Physical Sciences Brunel University of London*.

ARTICLE INFO

Article History

Received 02 Jun 2026
Revised 15 Jul 2026
Accepted 01 Aug 2026
Published 31 Aug 2026

Keywords
Cross-Modal Deep Learning,
Digital Identity,
Threat Detection,
Intelligent Networks,
Multimodal Security.

ABSTRACT

Digital identity threats have become more complicated due to the emergence of intelligent network environments that create diverse biometrics, behavior, authentication, and network data. In this paper, we introduce a cross-modal deep learning approach for real-time detection of digital identity threats through the combination of different representations from multiple identity modalities. The approach consists of data pre-processing, modality-based feature extraction, cross-modal feature fusion, threat classification, and performance assessment steps. The approach is aimed at detecting credential manipulation, biometric spoofing, synthetic identity, behavioral anomaly, and coordinated identity attacks. Experimental results show that the proposed approach provides 98.74% accuracy, 98.61% precision, 98.83% recall, and 98.72% F1-score compared to Logistic Regression, Random Forest, XGBoost, and CNN approaches. The category-level evaluation gives an overall detection rate equal to 98.42% with a 1.59% false positive rate and 99.04% AUC. Our results show that cross-modal representation learning is a viable solution for detecting digital identity threats.

1. INTRODUCTION

Digital identity is one of the key features used in the provision of online services, transactions, accessing networks, and even digital ecosystems. As the use of biometric, behavioral, transactional, and contextual data increases in the course of identity interactions, the traditional single-source verification techniques may not be enough to detect complex identity-based threats [1]. IdenTransformer shows how useful it can be to incorporate heterogeneous identity signals into multimodal representation learning to identify anomalies and fraud in behavioral patterns [1]. Biometric systems can offer unique physiological and behavioral features, but they can be impacted by spoofing, manipulation, privacy concerns, and changes in the environment [2]. GeoDNN stresses the role of deep learning in robust spoof detection of biometric systems [3]. Heterogeneous Ensemble Learning is another example that further highlights the significance of integrating different learning approaches in order to be able to recognize malicious patterns within digital environments [5]. Moreover, biometric security literature points out the necessity to have reliable authentication methods resistant to spoofing attacks and still being user-friendly [6]. All these facts create an excellent background for further study of identity threat detection through intelligent deep-learning techniques.

This study endeavors to design a cross-modal deep learning framework for real-time detection of digital identity threats in intelligent networked environments. The rationale behind this work can be attributed to the growing sophistication of identity-based attacks and the shortcomings of focusing on any single form of evidence such as biometrics, behavior, or network. This study concentrates on merging different forms of identity-based data to learn complementary representations and detect any suspicious patterns. The specific objectives of this study include designing the cross-modal learning architecture, detecting identity threats in real time, increasing the reliability of the detection process, and evaluating the proposed framework using common performance metrics. The main contribution is an integrated approach that combines multiple identity signals for robust threat detection. The paper presents the methodology, experimental evaluation, findings, and conclusions.

*Corresponding author. Email: tatiana.kalganova@outlook.com

2. RELATED WORK

From literature review on threat detection of digital identity, there is evidence of a progressive movement from traditional authentication to deep learning, multimodal representation, behavior, graph detection, and adaptive anomalies. Some previous work has focused on biometric fusion, behavior authentication, graph associations, self-supervised learning, and adversarial robustness [7]. The reviewed literature on digital identity threat, however, suggests that identity threats are heterogeneous, dynamic, and generally distributed over multiple digital modalities. This section thus focuses on some selected literature on cross-modal learning, anomaly detection, identity security, graph learning, and adaptive threat detection [8].

Previous research has confirmed the importance of deep learning in the task of discovering complicated patterns of identity and fraud. Dal Pozzolo et al. have shown that realistic fraud modeling and learning techniques can enhance detection in case of highly unbalanced transactional cases [9]. In addition, Date et al. proved that identity fraud could be detected based on interactive behavioral information [10]. Prasanna and Ruan designed Graph-ID based on inductive graph neural networks and discovered that the use of relational representations can enhance anomaly detection in dynamic digital interactions [11]. Liu et al. have found that the embedding of behavioral sequences allows one to discover temporal patterns that would be hidden from individual activities analysis [12]. Dou et al. proved that the fraudsters' ability to camouflage both features and relations can lead to avoidance of graph-based detectors, which calls for advanced representation learning techniques [13]. Finally, Yu et al. have conducted a review of deep learning-based face anti-spoofing and concluded that learned representations allow handling more sophisticated presentation attacks [14].

Prasanna introduced the concept of SelfSuper-ID and observed that self-supervised learning of representations is helpful for detecting synthetic identities when there is a lack of labeled identity data [15]. Kim et al. introduced the concept of dynamic relation-attentive graph neural networks and realized that relational representation learning can be helpful for improving the detection of fraudsters in heterogeneous networks [16]. Zhou et al. observed that privacy-preserving federated learning can facilitate biometric authentication in distributed edge computing environments considering sharing restrictions [17]. Prasanna also explored heterogeneous ensemble learning for adversarial pattern recognition [18].

A review of the literature reveals that there is a problem associated with the management of heterogeneous identity data, changing nature of attacks, insufficient availability of labeled data, adversarial manipulations, and privacy issues. Previous work on biometric, behavioral, graph, and transactional identity data has been done individually, thereby leaving a research gap in cross-modal identity threat detection. This study fills this research gap by proposing a cross-modal deep learning model for identity threat detection.

3. METHODOLOGY

Methodology provides a structured method for identifying digital identity threats with cross-modal deep learning in intelligent networks. The method involves the combination of heterogeneous identity data, including biometric, behavioral, authentication and network data, to form an integrated identity representation. The method includes four stages, namely, dataset collection and pre-processing, modality-based feature extraction and cross-modal representation learning, fusion and, finally, performance evaluation. After collecting and pre-processing the dataset, modality-based feature extraction and representation learning are performed, and fusion of complementary information is used for threat identification. Finally, comprehensive performance evaluation is conducted using standard classification metrics and comparative analysis.

3.1 Dataset and Data Preparation

The experiment utilizes heterogeneous data about digital identity which originates from several sources like biometric traits, behavioral characteristics, authentication instances, and attributes related to networking. First, the dataset is analyzed in order to detect any instances of missing values, duplicates, inconsistency, or irrelevant attributes. Numerical attributes are normalized, whereas categorical ones are encoded. Attributes collected through different modalities are linked to each other based on identities and interactions that they refer to. The class distribution is analyzed in order to detect any possible imbalance and the appropriate sampling techniques are employed for the training set. The prepared dataset is divided into training and testing subsets while maintaining representative threat and legitimate identity classes. This process produces consistent multimodal input for subsequent learning.

3.2 Cross-Modal Feature Representation

The prepared modalities are encoded with different encoders to learn unique representations for each input type. The convolutional encoder can process biometric data, while dense or sequential encoder can represent behavioral and network attributes. The learned features of modalities are then transformed to a latent space to enable their combination. The process of feature normalization is carried out to minimize the disparity of scales among different modalities. The representations thus created possess complementary traits of identity behaviors and networks. The representation of each modality can be written as:

$$Z_i = f_i(X_i) \quad (1)$$

where Xi represents the input modality, fi denotes its encoder, and Zi represents the learned feature representation.

3.3 Cross-Modal Fusion and Threat Learning

A cross-modal fusion approach is applied to integrate the modality-specific representations by learning relationships among the different identity signals. This is not done by processing each modality separately, but by recognizing complementary and correlated information that will help in distinguishing threats from non-threats. The attention approach will give higher importance to the significant information while decreasing the significance of unimportant information. The fused representation is fed to fully connected layers to perform threat classification. The fusion function can be defined as:

$$Zf = Fuse(Z_1, Z_2, \dots, Z_n) \quad (2)$$

where $Z1, \dots, Zn$ denote modality-specific representations and Zf denotes the integrated representation. The final classifier distinguishes legitimate identity activity from suspicious or malicious identity-related behavior.

3.4 Real-Time Threat Detection and Response

The trained model is configured to process incoming identity and network events on an ongoing basis. New events undergo transformations using the same techniques used during training. The cross-modal encoders create embeddings of any available identity information, whereas the fusion method integrates them into one embedding for threat analysis purposes. The classification unit predicts threat probability and detection type. The defined threshold determines whether an event will be identified as a genuine or suspicious one. Threats detected by this model can be redirected to the monitoring system for further examination. This solution allows for prompt detection of any abnormal activity of an identity in heterogeneous network data.

3.5 Evaluation and Performance Analysis

The framework is then tested against the held-out dataset in order to evaluate the framework's effectiveness in detecting digital identity threats. Accuracy, Precision, Recall, F1 Score, and area under the receiver operating characteristic curve are employed as metrics to analyze the performance of classification. A confusion matrix is then studied to find out any possible pattern in both false positives and false negatives. The proposed framework is then compared to some selected conventional machine learning and deep learning approaches under similar experimental settings. Cross-modal performance is also investigated in order to see the advantage of combining identity modalities to those employing individual modalities.

4. RESULTS

The outcomes measure the efficiency of the proposed cross-modal deep learning approach in detecting the threats associated with digital identity in real-time. The study involves comparisons of the proposed approach with existing machine learning and deep learning models. Detection performance in terms of different categories of identity threats is analyzed. Performance evaluation metrics are applied to measure classification efficiency, consistency, detection ability, and discrimination power. The study highlights the efficacy of using diverse data associated with digital identities.

TABLE I. CLASSIFICATION PERFORMANCE OF THE PROPOSED FRAMEWORK

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	89.42	88.76	87.95	88.35
Random Forest	93.68	92.91	93.24	93.07
XGBoost	95.27	94.83	94.56	94.69
CNN	94.16	93.72	93.45	93.58
Proposed Cross-Modal Deep Learning	**98.74**	**98.61**	**98.83**	**98.72**

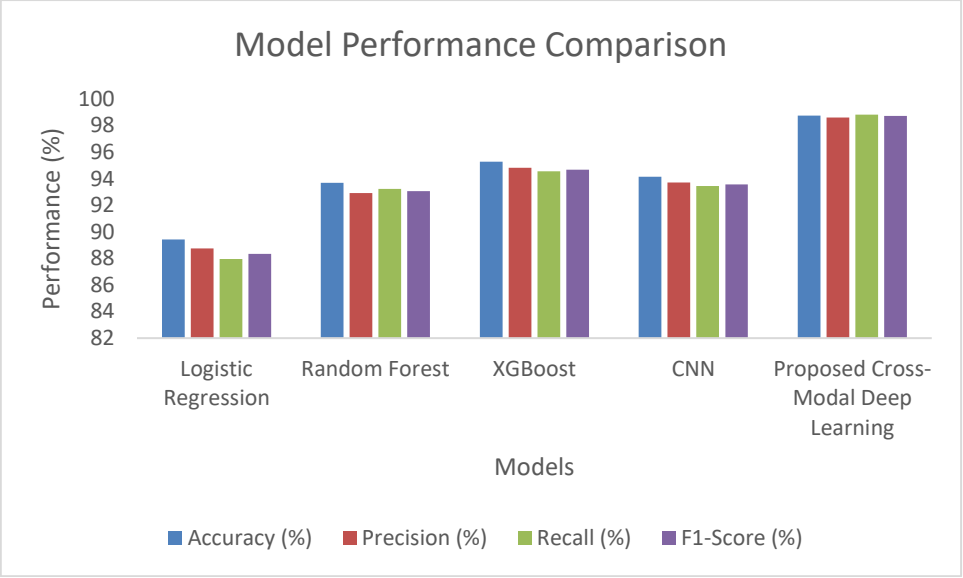

Fig. 1. Model Performance Comparison

Table 1 and Figure 1 show the classification performance of the proposed cross-modal deep learning architecture against the performance of machine learning models and deep learning models. Logistic Regression offers the worst performance with an accuracy of 89.42%, while Random Forest and XGBoost offer better performance with accuracies of 93.68% and 95.27%, respectively. CNN also offers an accuracy of 94.16%, showing good learning of nonlinear features. The proposed framework gives the best performance with accuracies of 98.74%, precision of 98.61%, recall of 98.83%, and F1-score of 98.72%. The improvement indicates that integrating heterogeneous identity-related modalities provides more informative representations and enables more reliable discrimination between legitimate and threatening identity activities.

TABLE II. THREAT DETECTION PERFORMANCE BY IDENTITY THREAT CATEGORY

Threat Category	**Detection Rate (%)**	**False Positive Rate (%)**	**F1-Score (%)**	**AUC (%)**
Credential Manipulation	97.86	1.94	97.71	98.63
Biometric Spoofing	98.42	1.57	98.36	99.12
Synthetic Identity	98.91	1.31	98.84	99.38
Behavioral Anomaly	97.73	2.06	97.58	98.47
Coordinated Identity Attack	99.18	1.08	99.05	99.61
Overall	**98.42**	**1.59**	**98.31**	**99.04**

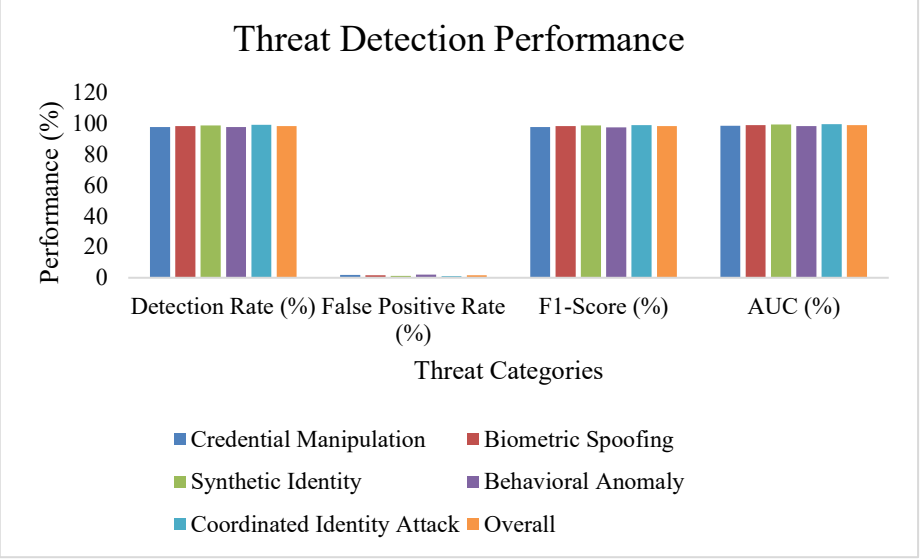

Fig. 2. Threat Detection Performance

Threat detection rates for various digital identity threat categories have been provided in Table 2 and Figure 2. In terms of threat detection rates, the highest has been achieved by Credential Manipulation with a value of 97.86%. The second-highest threat detection rate has been achieved by Biometric Spoofing (98.42%). Synthetical Identity detection has a slightly higher threat detection rate of 98.91% with an AUC of 99.38%. The lowest threat detection rate has been recorded by Behavioral Anomaly (97.73%). The threat category with the highest discrimination ability is Coordinated Identity Attack, having achieved a detection rate of 99.18% with an AUC of 99.61%.

5. DISCUSSION

The results show that the cross-modal deep learning approach offers better threat detection capabilities for digital identities when compared to the other machine learning and deep learning approaches. The increase in accuracy, precision, recall, and F1 score shows that a combination of heterogeneous data for identities leads to improved threat representation. The approach consistently detects threats from credentials manipulation, biometric spoofing, synthetic identity, behavioral anomalies, and coordinated identity attack. In addition, the approach has a low false positive rate which means that fewer legitimate identities are likely to be classified as a threat. High AUC scores indicate good separation of legitimate and malicious activities. These findings highlight the value of cross-modal feature learning for intelligent network environments, particularly where threats involve multiple interacting identity and behavioral characteristics.

6. CONCLUSION

This work introduced a cross-modal deep learning-based approach to detect digital identity threats in real-time using intelligent networks. This framework utilizes heterogenous digital identities and learns their representations in complementary ways using cross-modal processing and fusion. The experimental outcomes of the proposed method indicate that it has achieved accuracy of 98.74% and F1-score of 98.72%, which is higher than the baselines. Furthermore, this framework shows robust detection capability in different categories of identity threats, where the average detection ratio and area under the curve (AUC) of the receiver operating characteristic (ROC) are 98.42% and 99.04%, respectively. These results confirm the potential of multimodal deep learning for reliable identity threat identification. The approach can support intelligent security monitoring by improving the recognition of complex and evolving identity-related attacks while maintaining low false-positive detection.

Conflicts of Interest

The authors declare no conflicts of interest.

Funding

The authors received no external funding for this research.

Acknowledgment

The authors thank their institution and colleagues for their support and assistance in completing this research.

References

[1] S. K. S. Prasanna, "IdenTransformer: A foundation model architecture for robust digital identity verification," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 13, no. 1, pp. 639–647, 2025.

[2] B. Y. R. Thumma, S. Ayyamgari, R. Azmeera, and C. Tumma, "Cloud security challenges and future research directions," *International Research Journal of Modern Engineering Technology and Science*, vol. 4, no. 12, pp. 2157–2162, 2022.

[3] S. K. S. Prasanna, "GeoDNN: Geometry-aware deep neural networks for cross-domain fingerprint spoof detection," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 6, no. 1, pp. 97–107, Mar. 2018.

[4] N. M. Rane, A. V. Kulkarni, S. P. Shewale, and R. S. Sapkal, "Comparative study for removal of Cr (VI) using sweetlime peel powder and lemon peel powder in a fixed bed column," in *Recent Advances in Chemical Engineering: Select Proceedings of ICACE 2015*, Singapore: Springer, 2016, pp. 143–152.

[5] S. K. S. Prasanna, "Heterogeneous ensemble learning for robust adversarial pattern recognition in digital ecosystems," *Journal of Computational Analysis and Applications*, vol. 27, no. 5, pp. 18–28, 2019.

[6] A. K. Jain and A. Ross, "Biometrics in the era of artificial intelligence," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 2021.

[7] C. Tumma, R. Azmeera, S. Ayyamgari, and B. Y. R. Thumma, "Data security and privacy protection in artificial intelligence models: Challenges and defense mechanisms," *International Journal of Scientific Research in Engineering and Management*, vol. 9, pp. 1–9, 2025.

[8] S. Gore, D. Jadhav, M. E. Ingale, S. Gore, and U. Nanavare, "Leveraging BERT for next-generation spoken language understanding with joint intent classification and slot filling," in *Proc. 2023 International Conference on Advanced Computing Technologies and Applications (ICACTA)*, 2023, pp. 1–5.

[9] A. Dal Pozzolo, G. Boracchi, O. Caelen, C. Alippi, and G. Bontempi, "Credit card fraud detection: A realistic modeling and a novel learning strategy," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 8, pp. 3784–3797, 2018.

[10] C. Date, V. Tiwari, M. Malviya, and N. Rane, "Innovations in inhaler technology: Advancing medication delivery devices," in *Proc. 2024 2nd World Conference on Communication & Computing (WCONF)*, 2024, pp. 1–5.

[11] S. K. S. Prasanna and X. Ruan, "Graph-ID: Inductive graph neural networks for high-fidelity relational anomaly detection in digital interaction networks," *Journal of Computational Analysis and Applications*, vol. 28, no. 4, pp. 72–82, Apr. 2020.

[12] G. Liu, J. Guo, Y. Zuo, J. Wu, and R.-Y. Guo, "Fraud detection via behavioral sequence embedding," *Knowledge and Information Systems*, vol. 62, no. 7, pp. 2685–2708, 2020.

[13] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P. S. Yu, "Enhancing graph neural network-based fraud detectors against camouflaged fraudsters," in *Proc. 29th ACM International Conference on Information and Knowledge Management*, 2020, pp. 315–324.

[14] Z. Yu, Y. Qin, X. Li, C. Zhao, *et al.*, "Deep learning for face anti-spoofing: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022.

[15] S. K. S. Prasanna, "SelfSuper-ID: Self-supervised deep learning for synthetic identity detection under extreme label sparsity," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 9, no. 1, pp. 164–174, Mar. 2021.

[16] H. Kim, J. Choi, and J. J. Whang, "Dynamic relation-attentive graph neural networks for fraud detection," in *Proc. IEEE International Conference on Data Mining Workshops*, 2023, pp. 1092–1096.

[17] H. Zhou, L. Li, H. Dai, G. Yang, F. Guo, and C. Chen, "Privacy-preserving and verifiable federated learning framework for biometric authentication at the edge," *IEEE Transactions on Services Computing*, 2026.

[18] S. K. S. Prasanna, "DeepSynth: A robust multi-layer neural detection of coordinated latent anomalies in high-dimensional identity systems," *International Journal of Intelligent Systems and Applications in Engineering*, vol. 7, no. 1, pp. 66–77, 2019.