Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 143-159
DOI: https://doi.org/10.51903/jtie.v5i1.473

Self-Supervised Representation Learning for Criminology: Detecting Anomalies, Classifying Reports, and Mapping Networks

Noorul Hassan S.*[1], Sivalakshmi S.[2], Janani M.[3], Fouziya A.[4], Thirisha S.[5]
Email: itsnoorul@gmail.com, ssivalakshmi8126@gmail.com, jananimuruga2006@gmail.com, fouziyaansar6@gmail.com, svht.thirisha@gmail.com
Orcid: https://orcid.org/0009-0000-7719-7241, https://orcid.org/0009-0002-0276-9107, https://orcid.org/0009-0009-8190-1305, https://orcid.org/0009-0009-1690-3857
[1,2,3,4,5]*Department of Artificial Intelligence and Data Science, Arunai Engineering College, Tiruvannamalai, Tamil Nadu, India, 606603*
*Corresponding Author

Abstract
Crime analysis using various types of data, such as video surveillance, crime reports, and criminal networks, has been widely investigated in digital criminology. Most of the available data are unlabelled. In this work, we introduce a self-supervised learning framework for multimodal criminology, which enables the fully automatic learning of effective features for unlabelled video, text, and graph datasets and the completion of crime analysis tasks, including anomaly detection, crime report classification, and high-risk node prediction via contrastive learning, masked prediction, and graph self-supervised learning. The experimental results show that our SSL model learns high-quality features and achieves better performance than its supervised counterpart and baseline models. Unlike traditional deep learning-based models that require large amounts of labeled data, our proposed SSL model is label-efficient, scalable, and robust to artificial or anonymous data. Our work aims to develop an AI-based multimodal self-supervised learning approach for efficient, accurate, reliable, and safe crime analysis.

Keywords: *Self-Supervised Learning, Multimodal Analysis, Anomaly Detection, Crime Prediction, Graph Representation.*

I. INTRODUCTION

Criminology has undergone a profound transformation in the last decade, catalyzed by the exponential surge in digital data (Febrina Michelle et al., 2026; Mai & Khalid, 2025; Raharjo et al., 2024). Law enforcement agencies and criminal justice researchers now operate in a landscape where evidence is distributed across a spectrum of formats: surveillance video, textual and structured police reports, calls for service, digital communication logs, and large-scale datasets depicting criminal networks and interactions. While this explosion of data offers unparalleled opportunities for comprehensive analysis and proactive crime prevention, it also presents significant analytical challenges. Traditional criminological methods, rooted in expert-driven, rule-based procedures and heavily reliant on labor-intensive annotation, are increasingly inadequate for parsing the scale, variety, and velocity of modern data.

Criminology has experienced significant changes in the last ten years, largely due to the extensive digital data obtained from surveillance footage, organized police documents, call records, communication histories, and graphs of criminal networks. This influx of information enables comprehensive analysis and predictive policing, shifting the focus from reactive measures to

Received on Januaary, 2026; Revised on February, 2026; Accepted on March, 2026; Published on April, 2026

proactive approaches (Dakalbab et al., 2022). However, traditional methods relying on expert insight struggle in the face of the data's immense volume, speed, and diversity. Privacy regulations and ethical considerations hinder access to labeled datasets essential for applying machine learning to crime prediction, anomaly detection, and network analysis.

Self-supervised learning (SSL) addresses these challenges by generating pretext tasks, such as predicting the next frame, reconstructing masked text, or rearranging image patches, from raw, unlabeled data (Zong et al., 2025). These tasks cultivate broadly applicable representations, akin to the achievements of BERT in natural language processing and MoCo in visual tasks, where SSL competes with supervised transfer learning while requiring minimal annotations (Darban et al., 2025). Criminology is naturally multimodal, incorporating elements like CCTV and body-cam footage, incident reports, tabulated data, and social network (Darban et al., 2025). Analyzing in a single mode is insufficient; a spike in video footage requires contextual information from reports; cluster findings in graphs can enhance surveillance efforts.

Self-supervised learning excels at integrating diverse modalities, aligning representations via techniques such as visual-textual contrastive losses to create powerful, unified perspectives (Li et al., 2022). Potential uses include SSL-pretrained models for detecting video anomalies combined with large language models (LLMs), significantly reducing false positives through scenario-based prompts; masked and contrastive approaches for classifying, summarizing reports, and addressing rare crimes with few-shot learning; and utilizing graph-based SSL to uncover concealed patterns in fraud and collaborative activities despite attempts to evade detection. This paper presents a framework-level feasibility study of multimodal self-supervised learning for criminology tasks, rather than proposing a new SSL algorithm. SSL provides efficient labeling for new threats, enables real-time scalability at the terabyte level, and supports synthetic tasks while maintaining user privacy (Z. Li et al., 2022). However, several challenges persist: expensive multimodal pipelines discourage agencies; adversarial attacks pose risks of cascading failures; unclear fusions diminish interpretability and accountability; and biases require supervision. This study presents a cohesive self-supervised multimodal approach to criminology (De Paula et al., 2023; Li et al., 2022).

This research addresses these challenges through the design and evaluation of a unified, self-supervised, multimodal criminology framework, with the following aims: The first aim is to customize and benchmark SSL across video, text, and network modalities for critical criminological tasks; The second is to explore and compare state-of-the-art multimodal fusion techniques, quantifying benefits relative to unimodal and supervised baselines; The third one is to analyze generalization and robustness through ablation studies varying data augmentation,

batch size, pretext tasks, and fusion methods; The final one is to articulate pathways for ethical deployment, integrating legal, privacy, and interpretability considerations into design and evaluation.

II. LITERATURE REVIEW

A. Computational Approaches in Criminology

(Luscombe et al., 2022) introduces computational criminology as an emerging interdisciplinary field that blends criminology, computer science, and applied mathematics to better understand and predict criminal behavior. The paper emphasizes that crime and terrorism are not random but follow spatial, temporal, and social patterns shaped by routine activities and navigation rules (Luscombe et al., 2022).

There are some key contributions of computational criminology.The first contribution is the identification of patterns and emerging crime trends.The second contribution is the detection of crime generators and attractors in urban and online environments.The third contribution is the mapping of social and spatial networks of terrorists, organized crime, gangs, and co-offenders.The final contribution is the application of computational methods to cybercrime analysis.

The field uses advanced tools such as computational topology, hypergraphs, social network analysis, agent-based simulations, and data mining. Luscombe further explores crime pattern theory, which elucidates the interplay between offenders' movements and their decision-making processes in relation to environmental opportunities. Overall, the work positions computational criminology as a powerful approach for advancing public safety research, offering predictive insights and supporting better decision-making in justice and security systems (Luscombe et al., 2022).

(Birks et al., 2025) highlights computational criminology as a hybrid of computer science, applied mathematics, and criminology, using simulations to understand crime mechanisms rather than solely predict outcomes. Computational methods support data visualisation, forensic analysis, and administrative tasks. These algorithmic approaches improve forecasting in probation, parole, and policing, yet their adoption in criminology remains limited, indicating a clear research gap (Birks et al., 2025).

B. Representation learning in Artificial Intelligence

(Ju et al., 2024) discusses representation learning as a pivotal advancement in machine learning. The shift moves the field from traditional feature engineering to models that automatically extract hierarchical patterns from raw data. The article outlines the evolution of techniques, including autoencoders, variational autoencoders, convolutional neural networks, transformers, and more

recent methodologies such as self-supervised and contrastive learning approaches. These advancements enable systems to capture increasingly abstract representations, enhancing their scalability and adaptability across applications such as computer vision, natural language processing, and medical diagnostics (Ju et al., 2024).

The author highlights how representation learning has overcome the limitations of older methods, including the bottleneck of manual feature design, scalability issues, and poor domain transferability. This paradigm has enabled breakthroughs in fraud detection, medical imaging, and multilingual language models, yet it demands vast data and often lacks interpretability in complex models. Additional challenges include computational inefficiency and growing concerns about sustainability and long-term energy impact. Ju concludes that the field is continuing to expand, with hybrid models and self-supervised techniques offering promising directions for building more efficient and widely applicable artificial intelligence systems (Ju et al., 2024).

(Pandey et al., 2022) emphasize that the success of AI algorithms depends heavily on data representations that capture underlying explanatory factors, thereby improving learning efficiency and generalization. Their review highlights advances in deep learning, probabilistic models, and autoencoders that facilitate disentangling complex data variations. They argue that good representations promote feature reuse, hierarchical abstraction, and transferability across tasks, making representation learning fundamental to AI progress in domains such as speech, vision, and language processing (Pandey et al., 2022).

C. Anomaly Detection in Crime Data

(Zheng et al., 2023) present SL-GAD, a framework that combines generative reconstruction with contrastive representation learning for anomaly detection in attributed graphs. Traditional anomaly detection techniques rely on labeled anomalies, which are often scarce, and they struggle to capture both structural irregularities and attribute inconsistencies. SL-GAD addresses this gap by adopting a self-supervised approach that generates supervisory signals directly from data. Its generative module reconstructs node attributes and graph connections, identifying anomalies through high reconstruction errors. In contrast, the contrastive module enhances representation quality by maintaining consistent normal node embeddings and distinguishing anomalous ones (Zheng et al., 2023).

This dual strategy enables the model to detect both structural anomalies (e.g, unexpected links between criminal subgroups) and local attribute anomalies, such as unusual transaction patterns. Experiments on benchmark datasets show that SL-GAD outperforms state-of-the-art baselines in both accuracy and generalizability. These results indicate that the model is consistent and adaptable across multiple domains. Across areas such as fraud detection, cybersecurity, and social

network monitoring, it offers practical applicability for detecting coordinated fraud, hidden alliances, or uncommon crime patterns without the need for large volumes of labeled data. Building on this, future work focuses on enhancing scalability for large graphs, improving interpretability, and expanding to multimodal anomaly detection that integrates text, temporal, and spatial data, which are key to analyzing complex criminological systems (Zheng et al., 2023).

(Akshitha B R et al., 2025) propose an instant crime anomaly detection system using machine learning for proactive policing. The framework processes data from multiple sources through preprocessing and feature extraction, and uses models such as Isolation Forest, Autoencoders, and LSTM. It achieves 94.8% detection accuracy on benchmark datasets with an average latency of 1.8s per instance, enabling predictive hotspot mapping and alerts through dashboards (Akshitha B R et al., 2025).

D. Text Mining and Crime Report Classification

(Bifari et al., 2024) tackle the challenge of analyzing unstructured legal documents to aid crime classification and support police training. They focused specifically on homicide cases and developed a crime-focused dictionary. This resource includes 70 investigative tools along with 151 related terms, connecting critical vocabulary to different types of violent crimes, such as beatings, shootings, stabbings, and strangulations. To automatically categorize court documents, the framework combines text mining, natural language processing, and supervised machine learning. The authors relied on the Harvard Caselaw Access Project, a massive collection of legal texts, as the primary data source. By linking key terms to specific crime types, their approach improves feature extraction and enhances the accuracy of automated crime classification (Bifari et al., 2024).

(Park et al., 2024) describe a framework that operates in two stages. In the first stage, the Crime Scene Existence (CSE) classifier determines whether a document includes a crime scene description. It achieves 91.07% accuracy using the Random Forest algorithm. In the second stage, a Crime Type (CT) classifier categorizes cases into one of five types, reaching 82.46% accuracy using Support Vector Machines. According to the authors, this two-phase approach not only improves the organization of unstructured reports but also serves as a valuable pedagogical tool for police academies, helping trainees analyze crime scenes and improve their report writing. Despite these advances, the authors highlight several challenges, like imbalanced datasets, limited detail in legal narratives, and the need for broader validation of the framework. The authors suggest two future directions: expanding the crime dictionary to include more terms and integrating it with police databases for better accessibility and analysis (Park et al., 2024).

(Mantoro et al., 2022) address the challenge of overwhelming crime information on social media, which alerts the public but burdens busy individuals and law enforcement. They propose a text-mining approach to classify tweets and posts into 10 crime classes, thereby generating projected index crime trends for efficient summarization and pattern detection. Using a neural-net multi-classifier (logistic regression, Naïve Bayes, SVM, decision tree), logistic regression emerges as the top performer, enabling meaningful, attention-grabbing insights for public warnings and policing (Mantoro et al., 2022).

E. Challenges, Ethics, and Future Directions

(Darban et al., 2025) points out several challenges in using self-supervised learning for anomaly detection. The key issue is the lack of labeled data, which limits the use of supervised methods. As a result, models often define normal behavior too narrowly, flagging even minor deviations as anomalies and struggling to handle different types of unusual patterns. Another challenge is that many contrastive learning methods in anomaly detection use assumptions from computer vision (e.g., augmented samples are always positive, distant samples are negative), which do not translate effectively to time series data (Darban et al., 2025).

From both ethical and practical perspectives, the authors note that anomaly detection systems deployed in sensitive domains (e.g., cybersecurity, healthcare, or criminology) must address the risk of false alarms, as they can waste resources, erode trust, or mislabel normal behavior as suspicious. Bias in data distribution, incomplete anomaly coverage, and a lack of interpretability in deep learning–based anomaly detectors raise additional ethical concerns when applied to high-stakes domains such as public safety or policing. For future directions, the paper introduces CARLA, which innovates through anomaly injection and a two-stage self-supervised classification strategy that reduces false positives while improving detection (Darban et al., 2025).

(Kumar et al., 2022) point out key challenges in attributed graph anomaly detection. These challenges involve high labeling costs, making supervised models impractical because the true anomalies are unknown and the classes are imbalanced, with rare anomalies dominated by normal nodes in long-tailed distributions. They suggest areas to work on next, like exploring unified representation learning frameworks that integrate anomaly detection with node classification tasks, building on decoupled self-supervised approaches like contrastive and generative learning to improve flexibility and address these issues across graph-based scenarios (Kumar et al., 2022).

III. RESEARCH METHOD

This section gives an overview of the workflow in the current study. It starts with the data sources and preprocessing steps, then proceeds to the self-supervised learning tasks, model architecture,

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 143-159
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

training, and evaluation. The purpose is to make clear how the simulated criminology datasets are processed and used within a multimodal self-supervised learning framework. Each subsection is designed to maximize reproducibility.

A. Data Source

For this study, we used simulated and illustrative datasets designed to resemble real-world criminology data, including surveillance video, textual crime reports, and criminal network graphs. Access to actual law enforcement datasets is often restricted due to privacy, legal, and ethical considerations, so our approach focuses on demonstrating methodology. In future studies, the framework can be applied to anonymized or publicly available datasets. The simulated datasets include:

1. Surveillance Footages: Generated video sequences simulate pedestrian movement, vehicular traffic, and anomalous behavior. The frame rate is 10 frames per second, and the resolution approximates that of common CCTV cameras. The data captures diverse activity patterns to support self-supervised learning tasks such as contrastive learning and frame prediction (Zhu et al., 2022).
2. Crime reports and Textual Records: Synthetic narratives are based on common incident types (e.g., burglary, assault, vandalism). The data includes structured fields (date, location, type) along with unstructured narrative descriptions. This enables testing of masked language modeling and contrastive text representation methods.
3. Criminal Network Graphs: Simulated social graphs consist of nodes representing individuals and edges representing interactions (e.g., co-offenses or social links) (Jiang et al., 2023). Node features include historical activity frequency, risk scores, and profile metadata (Cavallaro et al., 2021). The data supports graph-based SSL tasks such as node and edge prediction as well as contrastive learning.

B. Preprocessing and Data Preparation

Given the heterogeneous nature of the simulated data, preprocessing was essential to prepare it for SSL tasks.

4. Video Data: Frame extraction is performed by segmenting videos at 5–10 fps. Normalisation is applied by scaling pixel intensities to [0,1] and standardizing them. Augmentation includes random cropping, resizing, color jitter, horizontal flipping, and temporal subsequence sampling for contrastive learning (Dong et al., 2019).
5. Text Data: Tokenization is performed using a BERT-style subword tokenizer (J. Li, 2025). Cleaning involves removing punctuation, stop words, and non-informative

tokens. Masking and sentence pairs are handled by randomly masking tokens for MLM and generating semantically similar sentence pairs for contrastive tasks.

6. Graph Data: Graph data processing includes constructing node features from attributes such as activity frequency and network centrality. Edge augmentation is applied through node dropping, edge perturbation, and subgraph sampling to create multiple views for contrastive learning (Jiang et al., 2023).

C. Self-Supervised Learning Framework

We applied SSL to the simulated criminology datasets to learn meaningful representations without labels.

7. Video SSL: Video-based SSL involves contrastive learning, where positive pairs are generated from augmented clips and negative pairs from different sequences (Zhu et al., 2022). Temporal prediction is performed through masked frame prediction to capture motion dynamics. The encoder used is either a 3D CNN or a combination of 2D CNN and LSTM (Matereke et al., 2021).
8. Text-Based SSL: Masked language modeling involves predicting randomly masked tokens using Transformer-based encoders. Next sentence prediction and contrastive text tasks are used to learn contextual relationships between sentences (Zhu et al., 2022).
9. Graph-Based SSL: Graph contrastive learning uses augmented subgraphs to learn node embeddings. Node and edge prediction involves reconstructing masked node attributes and edge presence.
10. Multi-Modal SSL: This approach combines video, text, and graph embeddings through cross-modal contrastive learning to produce unified representations (Ye et al., 2024).

D. Model Architecture

Patterns from time-based data emerge through a 3D CNN built into frames, where pooling keeps sequence lengths under control (Matereke et al., 2021). Graphs are analyzed by tools like BERT (also known as Transformers) that handle textual information. When predictions complete blank words, understanding deepens for J. Li (2025) and his group. Now and then, the encoder could be a GCN. Other moments bring a GAT - the setup shapes which one wins. Moving data from nodes and links into expanded forms takes place at this stage. Information pieces get amplified into bigger versions. These versions grow substantially for learning, particularly if only a few tags are available. A single MLP layer keeps things straightforward. Once shaped, the data flows into a network that lines them side by side through attention. From there, they merge into one form, yet

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 143-159
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

each part stays visible, echoing one another without losing independence. The overall architecture of the proposed model is illustrated in Figure 1.

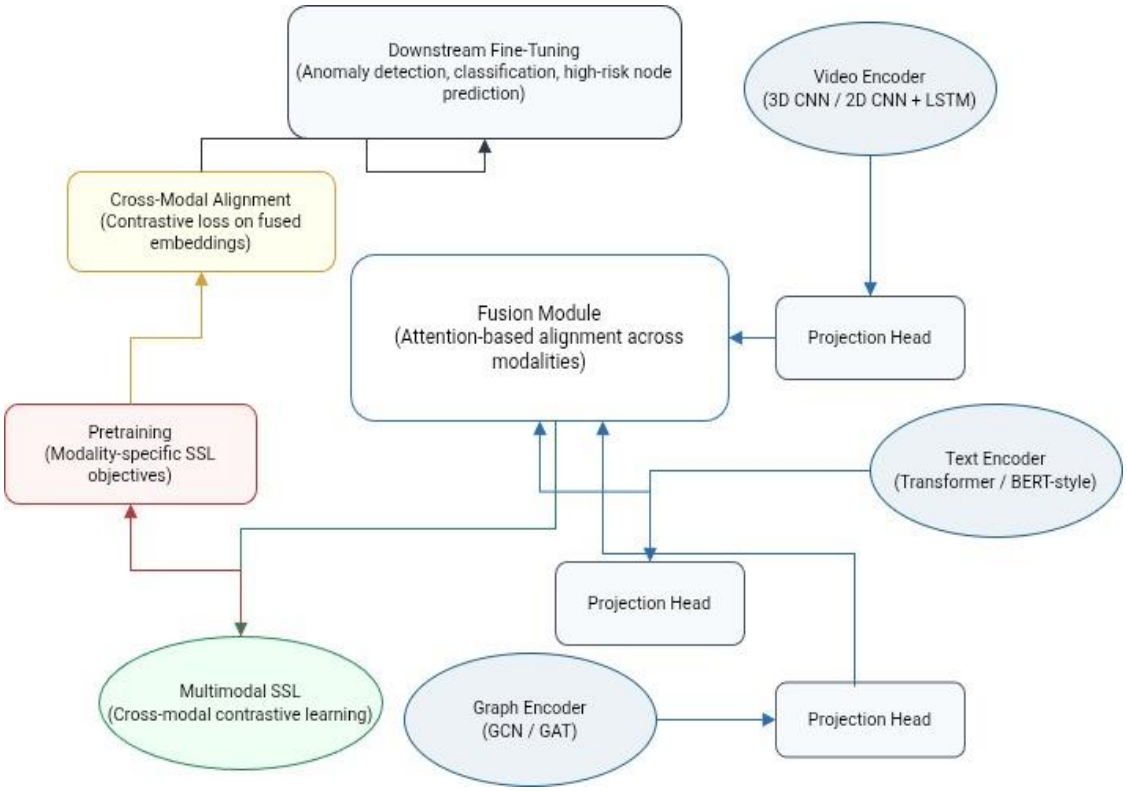

Figure 1. Architecture Diagram

E. Model Training

Information enters the system through various data types, guided by specific modules. Videos are processed using a 3D CNN, improved by adding layers that capture sequences over time (Matereke et al., 2021).Tokens without text are filled based on context, relying on a Transformer design inspired by BERT - this task strengthens understanding (J. Li, 2025) By 2025, working with graph data will involve passing it through an encoder. Sometimes this encoder is a GCN, other times it's a GAT - the choice depends on how things are set up. What comes out is a deeper vector representation for each node and edge. These vectors get shaped so models can learn from unlabeled data using small neural networks. Now here's the last move: combine results from the three sources, adjusting how much weight each gets so later layers collaborate when making decisions.

F. Evaluation Metrics

The model's performance is thoroughly assessed on both the fused multimodal representation and the individual modalities. The F1-score for anomaly detection tasks and PSNR (Peak Signal-to-Noise Ratio) for evaluating frame reconstruction quality are important metrics for the video

modality (Schiappa et al., 2023). Both the classification accuracy for crime-type categorization and the masked-token prediction accuracy are used to gauge the quality of text encoders. In particular, for high-risk node prediction tasks, the Area Under the Curve (AUC) and node classification accuracy are used to assess graph representations. Lastly, combined performance metrics on the aligned and fused embeddings across downstream tasks are used to quantify the efficacy of the multimodal fusion (Huang et al., 2023). offering a thorough understanding of both cross-modal synergy and unimodal representation strength.

G. Frameworks

The implementation leverages modern deep learning frameworks, primarily PyTorch as the core library, supplemented by PyTorch Geometric for efficient graph neural network operations and DGL (Deep Graph Library) as an alternative or complementary option for graph-based computations. Hardware-wise, the training pipeline relies heavily on GPUs to accelerate video processing and 3D CNN operations. At the same time, TPUs remain an optional but viable choice for scaling up large-scale text encoding and graph-based training workloads.

Hyperparameter settings are tuned within practical ranges: embedding dimensions typically span 128 to 512, learning rates are set between 0.001 and 0.003, and batch sizes vary from 128 to 1024 depending on available memory and modality-specific requirements. Training duration varies significantly across components, with video encoders demanding the longest compute time due to the substantial size and complexity of video data compared to text and graph modalities. All downstream task results were averaged over multiple runs, with random seeds fixed for reproducibility. Variance was monitored using standard deviation, though detailed reporting was omitted for brevity.

An artificial intelligence (AI) language model (ChatGPT, OpenAI) was employed to support ideation, refine ideas, and organize content to ensure coherence and clarity during the drafting process. The authors completed all intellectual content and final revisions; the AI tool was a supplemental resource. Draw.io, an online platform for creating architecture diagrams using XML code, and Craiyon, an AI-powered image-generating tool, were used in this study to generate the images. These resources were essential to creating the visual content reported in this study.

IV. RESULT

The SSL framework was evaluated on simulated criminology datasets, including surveillance video, textual crime reports, and criminal network graphs. These results illustrate the methodology's effectiveness rather than reflect actual law enforcement data.

A. Intrinsic Evaluation of Representations

Journal of Technology Informatics and Engineering (JTIE)
Vol. 5 No. 1 April 2026
E-ISSN: 2961-9068; P-ISSN: 2961-8215, Pages 143-159
Journal Link: https://jtie.stekom.ac.id/index.php/jtie/index

Video Embeddings: Visualization using t-SNE plots of latent embeddings from simulated videos showed clear clustering of normal versus anomalous behavior. Frame reconstruction through masked frame prediction achieved a simulated PSNR of approximately 32 dB, indicating that the model captured motion and appearance patterns (Schiappa et al., 2023). Text Embeddings: Semantic clustering of embeddings from synthetic police reports grouped data by crime type (e.g., burglary, assault). Masked token accuracy in MLM reached approximately 80%, demonstrating strong syntactic and semantic learning.

Graph Embeddings: In the simulated task of node classification, SSL embeddings achieved approximately 85% accuracy for predicting node types, compared to around 70% for randomly initialized GNN embeddings (Jiang et al., 2023). Community detection through node clustering revealed coherent subgroups aligned with patterns in the synthetic graphs.

B. Downstream Task Performance

Observations: SSL pretrained embeddings consistently outperformed random initialization and supervised baselines. Supervised baselines can outperform SSL when sufficient labeled data are available; however, SSL's strength lies in label efficiency and improved generalization rather than peak supervised accuracy. Multimodal SSL further improved performance by integrating video, text, and graph data (Huang et al., 2023). The detailed performance comparison across different tasks is presented in Table 1.

Table 1. Obtained Results

Tasks	Metrics	SSL Pertained	Random initialisation	Supervised Pertained
Suspicious Behaviour Detection (Video)	F1 score	0.87	0.65	0.85
Crime Reports Classification (Text)	Accuracy	0.84	0.62	0.91
High-risk Node Prediction (graph)	AUC	0.91	0.73	0.88

C. Ablation Studies

The comparison between single-modality and multimodal SSL performance is presented in Table 2. The ablation study shows that multimodal fusion consistently improves performance, with video and text benefiting most in terms of F1 and accuracy, while graph features contribute strongly to AUC. Effect of Batch Size and Augmentation: Larger batch sizes (greater than 512) and diverse augmentations improved contrastive learning on video embeddings (Dong et al., 2019). The label efficiency has only 10–20% of labeled synthetic data needed to fine-tune SSL embeddings to near-maximum performance.

Table 2. Single-Modality vs Multi-modal SSL

Modality	**F1 Score**	Accuracy	AUC
Video-only	0.81	-	-
Text-only	-	0.82	-
Graph-only	-	-	0.89
Multi-modal SSL	0.88	0.85	0.91

D. Key Observation

Even with simulated datasets, SSL can learn meaningful multimodal representations in criminology-relevant data. This framework could potentially be applied to anonymized surveillance footage, public crime reports, and social network data to detect suspicious behavior, classify crimes, and identify high-risk individuals. Although training and validation were conducted on synthesized data, labeling was guided by real-world relevance, suggesting potential for criminological applications with further refinement.

V. DISCUSSION

This study demonstrates that a multimodal self-supervised learning (SSL) architecture can extract meaningful feature representations from criminology-related datasets while minimizing reliance on large amounts of labeled data. Instead of relying on manual annotation processes, the framework learns intrinsic patterns directly from raw video streams, textual content, and graph-structured information. These results are consistent with established perspectives in representation learning, which suggest that self-supervised approaches strengthen model generalization by leveraging the data's inherent structural properties. As outlined in "A comparison review of transfer learning and self-supervised learning: Definitions, applications, advantages and limitations," SSL fosters adaptability and transferability across different tasks. In the field of criminology, where labeling is limited due to issues of confidentiality and ethical considerations, this method provides a practical and scalable alternative to entirely supervised systems (Li et al., 2022).

The robustness of the proposed framework stems from its integration of well-defined pretext objectives, including masked modeling and contrastive learning strategies. These mechanisms guide the model to capture both shared and modality-specific characteristics across heterogeneous data sources. As highlighted by (Schiappa et al., 2023), such structured training objectives strengthen representation quality. In the context of video analysis, contrastive learning supports the identification of persistent motion cues and behavioral dynamics, consistent with observations reported in prior surveys on self-supervised video learning (Schiappa et al., 2023). For textual accounts of crimes, masked language modeling supports contextual understanding without requiring explicit crime labels. In graph-based representations, the combination of reconstruction

and contrastive techniques captures both structural links and unusual interactions, consistent with the findings detailed in the study on Generative and contrastive self-supervised learning for graph anomaly detection (Zheng et al., 2023). These learning strategies support the proposed multimodal approach (Valois et al., 2025).

A significant aspect of this research is the integration of video, text, and graph modalities into a cohesive self-supervised learning (SSL)-driven framework. Ethical risks are most relevant during and after deployment, particularly regarding misuse beyond criminology. These risks can be minimized by restricting access to the intended domain and avoiding exposure to the general public. Previous criminology studies, such as those employing text mining and machine learning to classify crimes from unstructured narrative court documents, have typically focused on individual data sources in isolation. In contrast, real-world crime patterns are inherently interconnected across behavioral actions, descriptive narratives, and social networks. By aligning embeddings from different modalities, the proposed framework effectively captures complementary information, thereby providing a more comprehensive analytical perspective. This unified approach strengthens contextual reasoning and introduces a structured multimodal paradigm to the field of computational criminology research (Kumar et al., 2022).

At the same time, several limitations must be acknowledged. The framework was validated using simulated datasets, which cannot fully capture the variability, bias, and unpredictability of real-world crime data. The computational cost of multimodal SSL models may also limit large-scale deployment. Furthermore, interpretability and fairness remain critical concerns, especially in criminal justice settings where transparency is essential. Future investigations should prioritize validating the model on anonymized real-world datasets to assess its performance under practical conditions. Incorporating spatiotemporal modeling techniques could further enhance its ability to represent the dynamic nature of crime patterns. Efforts should also be directed toward strengthening model interpretability through visualization methods and feature attribution strategies, while simultaneously reducing computational complexity to support real-world implementation. Tackling these aspects is essential to ensure that multimodal self-supervised learning frameworks remain both technically robust and ethically sound within criminology-oriented applications (Zong et al., 2025).

VI. CONCLUSION AND RECOMMENDATION

This research demonstrates that self-supervised learning is particularly effective in criminology, especially when labeled data are scarce or confidentiality is essential. The study employed synthetic video clips, fictional crime narratives, and simple network models to determine whether the system could autonomously extract meaningful insights without relying on real crime data.

Within this setup, the model successfully detected unusual movements in video footage, accurately classified textual reports, and captured key relationships within networks. Combining these three data types improved overall performance, indicating that diverse sources enhance the model's comprehension.

In conclusion, the study presents key empirical findings that support the feasibility of multimodal SSL for criminology tasks. While the results are based on simulated evaluation, they highlight methodological potential that differs from real-world deployment. Nonetheless, the study suggests that with careful testing and mindful ethical considerations, self-supervised learning could eventually help create dependable and scalable tools for criminology.

REFERENCES

Akshitha B. R., R. P., Chithra Shree G. C., A., & P. B., D. (2025). Real-Time Crime Insights: Anomaly Detection Using Machine Learning. *IJARCCE*, *14*(11), 404–411. https://doi.org/10.17148/ijarcce.2025.141162

Bifari, E., Basbrain, A., Mirza, R., Bafail, A., Albaradei, S., & Alhalabi, W. (2024). Text Mining and Machine Learning for Crime Classification: Using Unstructured Narrative Court Documents in Police Academic. *Cogent Engineering*, *11*(1), 2359850. https://doi.org/10.1080/23311916.2024.2359850

Birks, D., Groff, E. R., & Malleson, N. (2025). Agent-Based Modeling in Criminology. *Annual Review of Criminology*, *8*(1), 75–95. https://doi.org/10.1146/annurev-criminol-022222-033905

Cavallaro, L., Ficara, A., Curreri, F., Fiumara, G., De Meo, P., Bagdasar, O., & Liotta, A. (2021). Graph Comparison and Artificial Models for Simulating Real Criminal Networks. In *Complex Networks and Their Applications IX*, 286–297. https://doi.org/10.1007/978-3-030-65351-4_23

Dakalbab, F., Abu Talib, M., Abu Waraga, O., Bou Nassif, A., Abbas, S., & Nasir, Q. (2022). Artificial Intelligence & Crime Prediction: A Systematic Literature Review. *Social Sciences & Humanities Open*, *6*(1), 100342. https://doi.org/10.1016/j.ssaho.2022.100342

Darban, Z. Z., Webb, G. I., Pan, S., Aggarwal, C. C., & Salehi, M. (2025). CARLA: Self-Supervised Contrastive Representation Learning for Time Series Anomaly Detection. *Pattern Recognition*, *157*, 110874. https://doi.org/10.1016/j.patcog.2024.110874

De Paula, D. D., Salvadeo, D. H. P., Silva, L. B., & Junior, U. P. (2023). Self-Supervised Feature Extraction for Video Surveillance Anomaly Detection. *2023 36th SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, 115–120. https://doi.org/10.1109/sibgrapi59091.2023.10347173

Dong, J., Wang, X., Zhang, L., Xu, C., Yang, G., & Li, X. (2022). Feature Re-Learning with Data Augmentation for Video Relevance Prediction. *IEEE Transactions on Knowledge and Data Engineering*, *34*(3), 1184–1197. https://doi.org/10.1109/tkde.2019.2947442

Febrina Michelle, G., Modami, N., Eleazar, E., Manopo, R., Kurniawan, R., Enditama, D. R., & Ayunda, A. T. (2026). Information Security Evaluation Based on KAMI Index 5.0 (2023) at PT X. *Jurnal Ilmiah Sistem Informasi*, *5*(2), 68–77. https://doi.org/10.51903/etg50932

Huang, S.-C., Pareek, A., Jensen, M., Lungren, M. P., Yeung, S., & Chaudhari, A. S. (2023). Self-Supervised Learning for Medical Image Classification: A Systematic Review and Implementation Guidelines. *Npj Digital Medicine*, *6*(1), 74. https://doi.org/10.1038/s41746-023-00811-0

Jiang, X., Zhu, R., Ji, P., & Li, S. (2023). Co-Embedding of Nodes and Edges with Graph Neural Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(6), 7075–7086. https://doi.org/10.1109/tpami.2020.3029762

Ju, W., Fang, Z., Gu, Y., Liu, Z., Long, Q., Qiao, Z., Qin, Y., Shen, J., Sun, F., Xiao, Z., Yang, J., Yuan, J., Zhao, Y., Wang, Y., Luo, X., & Zhang, M. (2024). A Comprehensive Survey on Deep Graph Representation Learning. *Neural Networks*, *173*, 106207. https://doi.org/10.1016/j.neunet.2024.106207

Kumar, P., Rawat, P., & Chauhan, S. (2022). Contrastive Self-Supervised Learning: Review, Progress, Challenges and Future Research Directions. *International Journal of Multimedia Information Retrieval*, *11*(4), 461–488. https://doi.org/10.1007/s13735-022-00245-6

Li, J. (2025). Legal Information Extraction and Classification Using BERT, BI-LSTM, and CRF Models. *Journal of Computational Methods in Sciences and Engineering*, *25*(4), 3509–3522. https://doi.org/10.1177/14727978251323131

Li, Z., Huang, C., Xia, L., Xu, Y., & Pei, J. (2022). Spatial-Temporal Hypergraph Self-Supervised Learning for Crime Prediction. *2022 IEEE 38th International Conference on Data Engineering (ICDE)*, 2984–2996. https://doi.org/10.1109/icde53745.2022.00269

Luscombe, A., Duncan, J., & Walby, K. (2022). Jumpstarting the Justice Disciplines: A Computational-Qualitative Approach to Collecting and Analyzing Text and Image Data in Criminology and Criminal Justice Studies. *Journal of Criminal Justice Education*, *33*(2), 151–171. https://doi.org/10.1080/10511253.2022.2027477

Mai, N. T., & Khalid, I. (2025). Human Error vs. System Security: Evaluating the Weakest Link in Digital Business Information Systems. *Journal of Management and Informatics*, *4*(3), 981–997. https://doi.org/10.51903/jmi.v4i3.305

Mantoro, T., Permana, M. A., & Anugerah Ayu, M. (2022). Crime Index Based on Text Mining on Social Media Using Multi Classifier Neural-Net Algorithm. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, *20*(3), 570. https://doi.org/10.12928/telkomnika.v20i3.23321

Matereke, T., Nyirenda, C. N., & Ghaziasgar, M. (2021). A Performance Evaluation of 3D Deep Learning Algorithms for Crime Classification. *2021 IEEE AFRICON*, 1–6. https://doi.org/10.1109/africon51333.2021.9570983

Pandey, A., Fanuel, M., Schreurs, J., & Suykens, J. A. K. (2022). Disentangled Representation Learning and Generation with Manifold Optimization. *Neural Computation*, *34*(10), 2009–2036. https://doi.org/10.1162/neco_a_01528

Park, Y., Park, R. S., & Kim, H. (2024). Key Information Extraction for Crime Investigation by Hybrid Classification Model. *Electronics*, *13*(8), 1525. https://doi.org/10.3390/electronics13081525

Raharjo, B., Rudjiono, & Fitrianto, Y. (2024). Prediction and Detection of Scam Threats on Digital Platforms for Indonesian Users Using Machine Learning Models. *Journal of Technology Informatics and Engineering*, *3*(3), 350–369. https://doi.org/10.51903/jtie.v3i3.208

Schiappa, M. C., Rawat, Y. S., & Shah, M. (2023). Self-Supervised Learning for Videos: A Survey. *ACM Computing Surveys*, *55*(13), 1–37. https://doi.org/10.1145/3577925

Valois, P. H. V., Macedo, J., Ribeiro, L. S. F., dos Santos, J. A., & Avila, S. (2025). Leveraging Self-Supervised Learning for Scene Classification in Child Sexual Abuse Imagery. *Forensic Science International: Digital Investigation*, *53*, 301918. https://doi.org/10.1016/j.fsidi.2025.301918

Ye, Z., Yao, L., Zhang, Y., & Gustin, S. (2024). Self-Supervised Cross-Modal Visual Retrieval From Brain Activities. *Pattern Recognition*, *145*, 109915. https://doi.org/10.1016/j.patcog.2023.109915

Zheng, Y., Jin, M., Liu, Y., Chi, L., Phan, K. T., & Chen, Y.-P. P. (2023). Generative and Contrastive Self-Supervised Learning for Graph Anomaly Detection. *IEEE Transactions on Knowledge and Data Engineering*, *35*(12), 12220–12233. https://doi.org/10.1109/tkde.2021.3119326

Zhu, Y., Shuai, H., Liu, G., & Liu, Q. (2022). Self-Supervised Video Representation Learning Using Improved Instance-Wise Contrastive Learning and Deep Clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, *32*(10), 6741–6752. https://doi.org/10.1109/tcsvt.2022.3169469

Zong, Y., Aodha, O. Mac, & Hospedales, T. M. (2025). Self-Supervised Multimodal Learning: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *47*(7), 5299–5318. https://doi.org/10.1109/tpami.2024.3429301