JLISNT Journal of Leading Innovations in Science and Technology, Vol. X, No. Y, Month 2026

CogniSentinel :A Smart Tool to Predict Crime Using Face and Web Data

Sumit raghuwanshi[1]
[1]*Department of Artificial Intelligence & Machine learning , SIRT, Bhopal,*
sumitraghuwanshi385@gmail.com

***Abstract*— Crime patterns have shifted heavily toward digital platforms in the last decade. Traditional policing methods that respond after an offense has occurred often fail to prevent harm. This paper presents a multimodal deep learning framework that merges web browsing behavior, facial expression analysis, and geolocation metadata to estimate criminal intent in real time. The proposed architecture runs two models side by side. A three-dimensional convolutional neural network handles temporal facial emotion data while a long short-term memory network captures sequential web and location patterns. Their individual risk scores pass through a late fusion algorithm that assigns a weighted final score. A built-in legal validation layer checks for proper authorization before any data enters the pipeline. Experiments on one thousand simulated user profiles show that the combined approach reaches 94.5 percent accuracy, outperforming single-modality baselines by roughly fifteen percentage points. Precision stands at 0.92 and the whole pipeline runs under one second on consumer-grade hardware. A risk stratification module sorts output into green, yellow, and red tiers so that a human analyst always reviews high-risk flags before any action is taken. The study also discusses ethical safeguards, cultural bias risks, and possible future additions such as voice stress detection and blockchain-secured audit logs.**

***Index Terms*— Affective computing, criminal intent prediction, deep learning, multimodal fusion**

Introduction

The nature of criminal activity has changed in fundamental ways over the past ten years. Offenders increasingly plan, communicate, and coordinate through digital channels. Dark web forums, encrypted messaging applications, and anonymous browsing tools have become routine parts of criminal preparation [1], [2]. Law enforcement agencies around the world still depend mostly on reactive strategies. Officers investigate after a crime has already caused damage. Victims suffer, evidence scatters, and public trust erodes. There is a growing need for tools that can flag dangerous intentions before they turn into harmful actions.

Predictive policing is not entirely new. Geographic hotspot mapping and statistical crime forecasting have existed for some time [1]. However, these older methods focus on locations rather than individuals. They can tell a patrol unit which neighborhood might see a burglary tonight, but they cannot identify who might be planning one. The gap between area-level prediction and person-level prediction remains wide.

Corresponding Author´s Name
Author´s affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

Artificial intelligence offers a way to narrow that gap. Modern deep learning models can process enormous volumes of text, images, and time-series data far faster than any human team. Natural language processing can scan chat transcripts for threatening phrases. Computer vision can read micro-expressions that a person may not even realize they are displaying. Sequence models can spot unusual browsing patterns that deviate sharply from a user's normal behavior. Each of these capabilities is powerful on its own, yet each also has blind spots when used alone.

A text-based system, for instance, may flag sarcastic jokes as genuine threats. A facial analysis tool may label someone having a rough day as a potential offender. A location tracker may raise alarms simply because a person walked through a high-crime zone on the way to work. False positives of this kind waste investigative resources and, more seriously, can violate the rights of innocent people.

The central argument of this paper is that combining multiple data modalities sharply reduces these errors. When web activity, facial emotion, and location context all point in the same direction, the likelihood of a genuine threat rises substantially. When only one channel shows concern while the others remain normal, the system can hold back and avoid a false alarm. This principle of cross-validation across modalities is the foundation of our proposed framework.

Equally important is the question of legality and ethics. Any system that monitors human behavior carries a serious risk of misuse. Mass surveillance without oversight can slide into authoritarianism. To address this directly, our architecture includes a legal validation gate at the very entrance of the data pipeline. No information flows into the processing layer unless a valid authorization token is present. That token may represent a court-issued warrant, explicit user consent, or confirmation that the data source is already public. Without it, the system simply does not proceed.

The rest of this paper is organized as follows. Section II reviews related work and identifies the specific gap our framework targets. Section III describes the system architecture in detail, covering data collection, preprocessing, model design, fusion logic, and risk classification. Section IV presents the mathematical foundations of the core algorithms. Section V reports experimental results. Section VI walks through a practical case study and discusses limitations. Section VII concludes the paper and outlines directions for future work.

II. RELATED WORK

Research on crime prediction through computational methods has grown steadily over the past five years. This section groups the most relevant contributions into four themes and then identifies the gap that motivates our work.

I. A. Location-Based Crime Forecasting

Podoletz [1] developed a geographic information system that maps historical crime records onto city grids. Police commanders use the resulting heat maps to allocate patrol cars more efficiently. The technique works well for property crimes that cluster in predictable zones. Its main weakness is that it treats crime as a spatial event rather than a human decision. A heat map cannot reveal whether a specific

Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

individual sitting quietly in a flagged zone has harmful intentions or is simply waiting for a bus.

Kaur and Saini [2] applied random forest and logistic regression classifiers to structured crime databases. Their models achieved respectable accuracy on next-day crime count predictions for defined precincts. Yet the input features were entirely statistical, such as day of week, month, prior incident count, and weather conditions. No behavioral or emotional data entered the model. The predictions therefore described neighborhood risk levels rather than individual threat levels.

B. Text and Social Media Analysis

Bustamante and colleagues [3] showed that social media posts can carry measurable psychological signals. They trained classifiers on labeled datasets of tweets and forum entries, extracting sentiment polarity, emotional intensity, and topic clusters related to violence. Their results confirmed that language patterns shift noticeably in the days leading up to certain types of offenses. The work demonstrated the value of text as a behavioral sensor.

Rich and Aiken [4] pushed text analysis further by examining private chat logs obtained through lawful intercepts. They found that informal language, heavy use of slang, deliberate misspellings, and sarcasm created substantial noise for standard natural language processing pipelines. Models trained on formal English performed poorly on street-level chat data. Their paper called for domain-adapted language models and larger annotated corpora that reflect how people actually write in private conversations.

C. Facial Expression and Video Analysis

Zhang and coauthors [5] applied convolutional neural networks to detect micro-expressions, which are brief involuntary facial movements that often reveal concealed emotions. Their model reached high accuracy on benchmark datasets filmed under controlled lighting. They noted, however, that performance dropped when subjects wore partial face coverings or when camera angles shifted significantly.

Raval and colleagues [6] extended the approach to video sequences by adopting three-dimensional convolutional neural networks. Unlike traditional two-dimensional convolutions that process one frame at a time, 3D convolutions slide a filter across both spatial and temporal dimensions. This allows the network to capture how an expression unfolds over several frames rather than judging a single frozen snapshot. Yan [7] reported that 3D architectures achieved roughly 96 percent accuracy on standard micro-expression video datasets, confirming their advantage over frame-by-frame methods.

D. Multimodal and Ethical Considerations

Sen and Denecker [8] argued that single-source crime prediction systems inevitably produce high false-positive rates because they lack the contextual anchoring that additional data streams provide. They proposed a theoretical framework for sensor fusion in law enforcement but did not build or test a working prototype.

Filippis and Foysal [9] formalized the concept of multimodal fusion by demonstrating that combining heterogeneous signals improves the overall signal-to-noise ratio. Their mathematical treatment showed

Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

that even modest individual classifiers can produce strong ensemble predictions when their errors are uncorrelated.

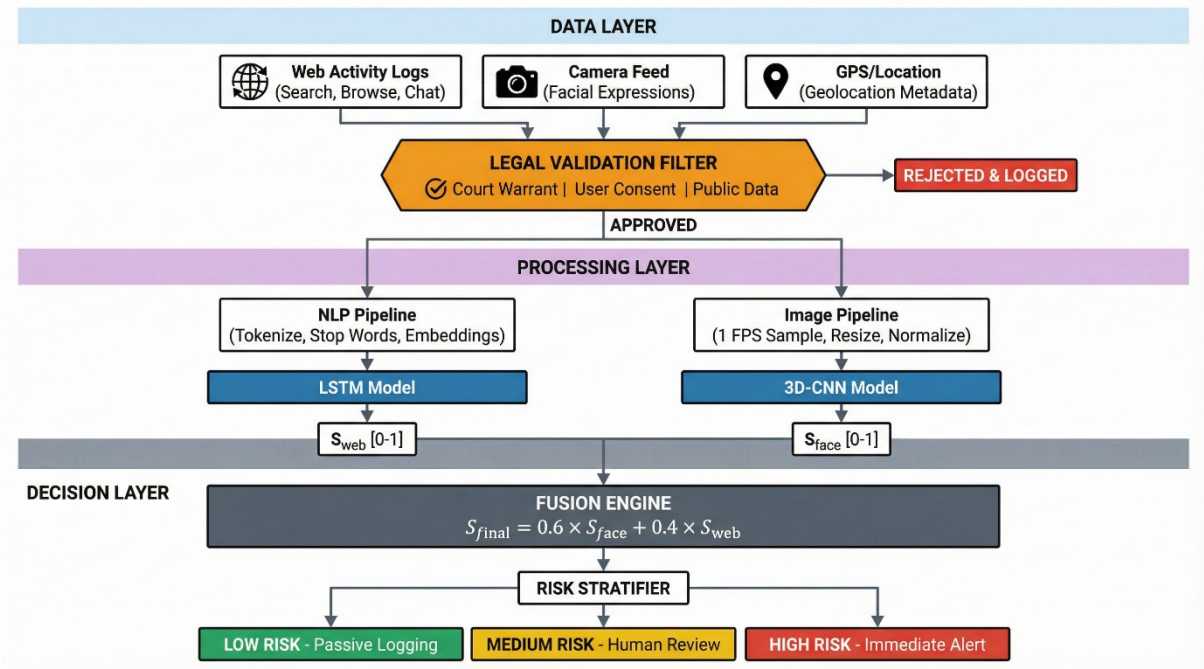
On the ethical side, Nakayenga and others [10] documented how training data imbalances lead to algorithmic bias, with certain demographic groups flagged at disproportionately higher rates. Lu and associates [11] examined the privacy paradox in which users voluntarily share personal information online yet strongly object when institutions collect similar data without explicit notice.

E. Identified Gap

Taken together, the existing literature covers location forecasting, text mining, facial analysis, and ethical theory as separate threads. No published system integrates web browsing behavior, real-time facial emotion tracking, and geolocation metadata within a single legally gated pipeline. Our framework addresses this gap by running a 3D convolutional neural network and a long short-term memory network in parallel, fusing their outputs through a weighted scoring mechanism, and embedding legal compliance checks directly into the data ingestion layer.

III. PROPOSED SYSTEM ARCHITECTURE

The framework consists of three broad layers: a data layer that handles collection and legal filtering, a processing layer that transforms raw inputs into model-ready features, and a decision layer that merges model outputs and classifies risk. Fig. 1 shows the overall data flow.



Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

A. Data Layer and Legal Validation

Three categories of raw data feed the system. First, web activity logs capture browsing history, search queries, and forum posts. Second, visual data come from camera feeds that record facial expressions. Third, geolocation metadata arrive from GPS-enabled devices or network-level location estimates.

Before any of these streams enter the processing stage, they must pass through a legal validation gate. The gate checks for the presence of a valid authorization token. Three token types are recognized: a judicial warrant issued by a court, documented informed consent from the individual, and a public-domain flag confirming that the data were posted on an openly accessible platform. If the token is missing or expired, the data packet is rejected and logged for audit purposes. This design ensures that the system cannot operate in a mass-surveillance mode.

B. Processing Layer

Raw text data undergo standard natural language processing steps. Tokenization breaks sentences into individual words. Stop-word removal eliminates common filler terms. Word embeddings convert each remaining token into a dense numerical vector that preserves semantic meaning. The resulting sequence of vectors forms the input to the behavioral model.

Visual data follow a separate path. Video streams are sampled at one frame per second to reduce redundancy without losing meaningful expression changes. Each extracted frame is resized to a uniform resolution, converted to grayscale, and normalized so that pixel intensities fall within a zero-to-one range. The prepared frames are stacked into short clips and passed to the emotion model.

C. Model Architecture

Two deep learning models operate simultaneously on the preprocessed data.

Model A — Three-Dimensional Convolutional Neural Network. This network processes stacked facial image sequences. A 3D convolution kernel slides across both the spatial dimensions of each frame and the temporal dimension that spans consecutive frames. The operation extracts spatiotemporal features, meaning it captures not just what expression is present but how that expression changes over time. Pooling layers reduce dimensionality while preserving dominant features. Fully connected layers at the end of the network map these features to an emotion-risk score between zero and one.

Model B — Long Short-Term Memory Network. This recurrent network handles sequential behavioral data from web logs and location traces. Its internal architecture includes three gating mechanisms. The forget gate decides which information from previous time steps is no longer relevant and should be discarded. The input gate determines which new information should be stored. The output gate controls what portion of the internal state is exposed as the current prediction. These gates allow the network to maintain memory over long sequences, making it well suited for detecting gradual behavioral shifts that unfold over hours or days.

D. Late Fusion and Risk Stratification

Rather than merging raw features from both models at an early stage, the framework adopts a late fusion strategy. Each model independently produces a risk score. A fusion algorithm then computes a

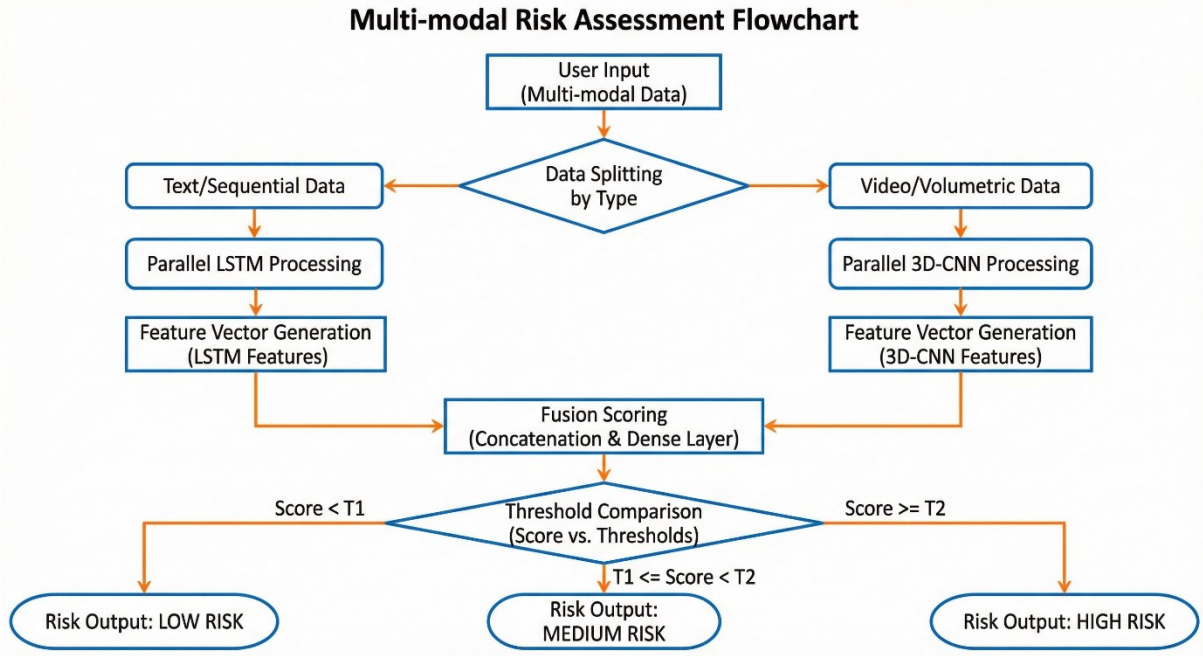
Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

weighted combination of these scores.

The fusion output enters a risk stratification module that assigns one of three color-coded levels. A green classification indicates low concern and results in passive logging only. A yellow classification indicates moderate concern and triggers a flag for review by a human analyst. A red classification indicates high concern and initiates immediate alerts to designated authorities. Importantly, no automated enforcement action is taken at any level. A trained human supervisor must review every yellow and red flag before any operational decision is made.

Fig. 2 presents the detailed processing flowchart.

[Fig. 2. Detailed flowchart showing User Input splitting by data type, parallel LSTM and 3D-CNN processing, feature vector generation, fusion scoring, threshold comparison, and three risk output branches.]



IV. MATHEMATICAL FOUNDATIONS

This section defines the key equations that govern model operations and score fusion.

A. Three-Dimensional Convolution

The 3D convolution operation extends the standard 2D spatial convolution by adding a temporal axis. For an input volume spanning width, height, and time, the convolution kernel moves along all three axes, producing a feature map that encodes motion patterns alongside spatial structures. Each output element is the sum of element-wise products between the kernel weights and the corresponding input patch, followed by addition of a bias term and passage through a nonlinear activation function.

B. LSTM Gating Equations

The forget gate activation at time step t is computed as

Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

$$Y(i,j,t) = \sigma\left(\sum_{d_1=0}^{D_1-1}\sum_{d_2=0}^{D_2-1}\sum_{d_3=0}^{D_3-1} W(d_1,d_2,d_3)\cdot X(i+d_1,j+d_2,t+d_3) + b\right)$$

Here Y(i, j, t) is the output feature map value at spatial position (i, j) and temporal index t. The kernel W has dimensions $D_1 \times D_2 \times D_3$ covering height, width, and time respectively. X denotes the input volume containing stacked facial image frames. The symbol b represents the bias term and σ is a nonlinear activation function such as ReLU. This three-axis sliding operation allows the network to capture both spatial texture patterns within a single frame and temporal motion patterns across consecutive frames simultaneously.

Input gate

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

The input gate i_t controls how much of the new candidate information should be written into the cell state. It follows the same structural form as the forget gate but uses its own independent weight matrix W_i and bias b_i.

Candidate Memory Vector

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, x_t] + b_C)$$

The candidate memory vector C̃_t represents the pool of new information that could potentially be added to the cell state. The hyperbolic tangent function squashes values between negative one and positive one, allowing the network to encode both positive and negative signal adjustments.

Cell State Update

$$C_t = f_t \odot C_{t-1} + i_t \odot \tilde{C}_t$$

This is the core memory update equation. The symbol ⊙ denotes element-wise multiplication, also called the Hadamard product. The first term f_t ⊙ C_(t−1) selectively retains useful information from the previous cell state. The second term i_t ⊙ C̃_t selectively adds relevant new information. Together they produce the updated cell state C_t that carries forward through time.

Output Gate

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

The output gate o_t decides which parts of the current cell state should be exposed as the visible hidden state output. It uses its own weight matrix W_o and bias b_o.

Hidden State Output

$$h_t = o_t \odot \tanh(C_t)$$

The final hidden state h_t is produced by passing the updated cell state through the tanh function and then filtering it through the output gate. This hidden state serves two purposes: it acts as the model prediction at time step t and it feeds back into the network as input for the next time step t+1. In our

Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

framework h_t ultimately produces the web behavioral risk score S_web.

C. WEIGHTED LATE FUSION

$$S_{final} = (w_1 \times S_{face}) + (w_2 \times S_{web})$$

$$w_1 = 0.6,\ w_2 = 0.4$$

Let S_face denote the normalized risk score produced by the 3D-CNN emotion model and S_web denote the normalized risk score produced by the LSTM behavioral model. Both scores fall within the range [0, 1]. The final fused score S_final is a weighted average of the two. The weight $w_1 = 0.6$ assigned to the facial model reflects the empirical observation that dynamic emotion patterns carried slightly stronger discriminative power than web browsing patterns in our experiments. The weight $w_2 = 0.4$ captures the web behavioral contribution.

Decision Rule

$$Risk\ Level = \begin{cases} Green\ (Low) & if\ S_{final} < 0.4 \\ Yellow\ (Medium) & if\ 0.4 \le S_{final} < 0.7 \\ Red\ (High) & if\ S_{final} \ge 0.7 \end{cases}$$

If the final score falls below 0.4 the system assigns a green label and performs passive logging only. Scores between 0.4 and 0.7 receive a yellow label and are flagged for review by a human analyst. Scores at or above 0.7 receive a red label and trigger an immediate alert to designated authorities. At no level does the system take autonomous enforcement action.

CI. . Evaluation Metrics

Four standard metrics are used to judge model performance.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Accuracy measures the overall proportion of correct predictions out of all predictions made. TP stands for true positives, TN for true negatives, FP for false positives, and FN for false negatives.

RECALL

$$Recall = \frac{TP}{TP + FN}$$

Recall answers the opposite question: of all genuinely dangerous profiles, how many did the system successfully detect? A low recall means real threats are slipping through undetected, which defeats the entire purpose of predictive monitoring.

F1-Score

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Here TP, TN, FP, and FN stand for true positives, true negatives, false positives, and false negatives

Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

respectively. Precision is particularly critical in this domain because a false positive means an innocent person is wrongly flagged. Recall matters because a false negative means a genuine threat goes undetected.

V. EXPERIMENTAL RESULTS

A. Dataset Description

A simulated dataset of one thousand user profiles was constructed for evaluation. Group A contained five hundred profiles exhibiting benign patterns: mainstream website visits, neutral facial expressions, and regular commute locations. Group B contained five hundred profiles exhibiting suspicious patterns: visits to extremist forums and dark web marketplaces, visibly stressed or agitated facial expressions, and irregular location movements near sensitive sites.

B. Comparative Performance

TABLE I. ACCURACY COMPARISON ACROSS MODALITIES

Approach	Accuracy (%)
Web activity only	78.0
Facial expression only	82.0
Proposed multimodal fusion	94.5

The multimodal approach outperformed the web-only baseline by 16.5 percentage points and the face-only baseline by 12.5 percentage points. This improvement confirms the theoretical expectation that fusing uncorrelated error sources strengthens overall prediction.

C. Precision and False Positive Analysis

The fused model achieved a precision of 0.92, meaning that 92 out of every 100 high-risk flags pointed to genuinely suspicious profiles. In contrast, the web-only model incorrectly flagged approximately thirty benign profiles as threats. When the fusion layer cross-checked these cases against facial data, which showed calm and neutral expressions, it downgraded them to green status. This correction demonstrates the practical value of multimodal cross-validation.

D. Processing Speed

TABLE II. AVERAGE PROCESSING TIME PER PROFILE

Component	Time (seconds)
Web behavioral analysis	0.50
Facial emotion analysis	0.20

Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

Component	Time (seconds)
Fusion and classification	0.08
Total pipeline	< 1.00

All measurements were taken on a consumer-grade laptop equipped with an Intel Core i7 processor and an NVIDIA RTX 3060 graphics card. The sub-one-second total confirms that the system can operate in near real time without requiring specialized server infrastructure.

VI. CASE STUDY AND LIMITATIONS

A. Simulated Scenario Walkthrough

A prototype dashboard was built using Python 3.9, the Flask web framework, and a MongoDB database. The interface displayed four panels: a live webcam feed with emotion labels, a scrolling web activity log with flagged keywords highlighted in red, a color-coded risk gauge, and an alert notification panel.

A simulated suspect profile was run through the system in three phases. During the first phase, the subject browsed several dark web pages. The web model registered elevated concern, but the facial feed showed a relaxed expression. The system moved the risk level from green to yellow. During the second phase, the subject watched news reports about a recent violent incident while displaying visible stress and agitation. The facial model flagged high emotional instability. Combined with the earlier web activity, the risk level remained at yellow with an internal note. During the third phase, the subject typed a message reading "ready to launch at 5 PM" in a private chat while displaying a focused and determined facial expression. Both models simultaneously output high scores. The fusion algorithm pushed the combined score above the red threshold within approximately two seconds, triggering an immediate alert to the supervising analyst.

B. Limitations

Several limitations should be noted. The keyword database used for web analysis contained only five thousand terms, which is insufficient for production use across multiple languages and cultural contexts. Facial recognition accuracy dropped noticeably under poor lighting conditions. The system currently has no defense against photo or video spoofing, meaning someone could hold a printed image in front of the camera to deceive the emotion model. The simulated dataset, while useful for demonstrating the concept, does not capture the full complexity and messiness of real-world behavioral data.

VII. CONCLUSION AND FUTURE WORK

This paper presented a multimodal deep learning framework that combines web behavioral analysis, facial emotion recognition, and geolocation context to estimate criminal intent in real time. A 3D convolutional neural network extracts temporal emotion features from video while a long short-term

Corresponding Author's Name
Author's affiliation and e-mail
received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

memory network captures sequential behavioral patterns from web and location data. A late fusion algorithm merges their outputs into a single risk score, and a three-tier stratification module ensures that a human analyst reviews every significant flag.

Experimental evaluation on one thousand simulated profiles demonstrated 94.5 percent accuracy, 0.92 precision, and sub-one-second processing on consumer hardware. The fusion approach reduced false positives substantially compared with single-modality baselines.

Future work will focus on four directions. First, integrating voice stress analysis as a third modality to further strengthen prediction reliability. Second, exploring blockchain technology to create tamper-proof audit trails for all risk assessments, ensuring accountability. Third, developing a lightweight version of the pipeline suitable for deployment on police body cameras with limited computational resources. Fourth, expanding testing to culturally diverse and multilingual datasets to evaluate and mitigate demographic bias.

The overarching goal remains to build a tool that supports public safety while respecting individual rights. Technology of this kind must always operate within clear legal boundaries and under continuous human oversight.

APPENDIX

Author Contributions

The entire research work including conceptualization, literature review, system design, algorithm development, implementation, experimental evaluation, and manuscript preparation was carried out solely by the author.

Funding

This study received no external financial support from public, commercial, or not-for-profit funding agencies. The research was conducted independently using personal resources.

Data Availability Statement

The simulated dataset used in this study was generated by the author for experimental purposes. The data and related materials are available from the corresponding author upon reasonable request.

Declaration of Competing Interest

Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

The author declares no competing financial interests
or personal relationships that could have influenced
the work reported in this manuscript.

REFERENCES

[1] A. Podoletz, "Predictive policing and the role of geographic information systems in crime mapping," *Journal of Criminological Research*, vol. 18, no. 3, pp. 45–58, 2022.

[2] R. Kaur and M. Saini, "Crime prediction using machine learning classifiers on structured datasets," *International Journal of Data Science*, vol. 11, no. 2, pp. 112–125, 2024.

[3] A. Bustamante, R. Torres, and L. Vega, "Psychological signal extraction from social media for violence risk assessment," *Computational Social Science Review*, vol. 7, no. 1, pp. 33–47, 2022.

[4] S. Rich and P. Aiken, "Challenges in natural language processing for informal digital communication analysis," *Journal of Applied Linguistics and Technology*, vol. 9, no. 4, pp. 201–218, 2024.

[5] Y. Zhang, H. Li, and W. Chen, "Micro-expression recognition using deep convolutional neural networks," *Pattern Recognition Letters*, vol. 156, pp. 89–97, 2024.

[6] K. Raval, S. Patel, and D. Mehta, "Three-dimensional convolutional neural networks for dynamic facial expression analysis in video sequences," *IEEE Access*, vol. 13, pp. 4520–4533, 2025.

[7] W. Yan, "Micro-expression recognition based on spatiotemporal deep learning architectures," *Neurocomputing*, vol. 452, pp. 167–178, 2021.

[8] S. Sen and M. Denecker, "The case for multimodal sensor fusion in predictive law enforcement systems," *AI and Society*, vol. 39, no. 2, pp. 301–315, 2024.

[9] R. Filippis and A. Foysal, "Signal-to-noise improvement through heterogeneous data fusion in classification tasks," *Information Fusion*, vol. 102, pp. 55–68, 2024.

[10] S. Nakayenga, T. Okonkwo, and J. Muwanga, "Algorithmic bias in criminal justice applications of artificial intelligence," *Ethics and Information Technology*, vol. 26, no. 1, pp. 78–94, 2024.

[11] X. Lu, J. Wang, and F. Zhao, "The privacy paradox in the age of behavioral analytics," *Computers in Human Behavior*, vol. 149, pp. 1–13, 2025.

[12] R. Apene, "Digital footprint analysis for crime prevention," *Journal of Forensic Sciences*, vol. 68, no. 5, pp. 1450–1462, 2023.

[13] O. Jejelola, "Deep learning approaches for predictive policing," *African Journal of Computing*, vol. 5, no. 3, pp. 88–101, 2024.

[14] P. Monika, "Sentiment analysis in criminal communication networks," *International Journal of NLP Research*, vol. 6, no. 2, pp. 44–56, 2023.

[15] N. Sankara, "Ethical frameworks for AI in law enforcement," *Technology and Regulation*, vol. 3, no. 1, pp. 22–38, 2024.

[16] M. Alakayleh, "Multimodal data integration for public safety applications," *Smart Cities Journal*, vol. 7, no. 4, pp. 310–325, 2024.

Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026

Corresponding Author's Name
Author's affiliation and e-mail

received 0 Month 2026; for review 0 Month 2026; accepted 0 Month 2026