Sensors and Materials, Vol. 38, No. 8 (2026) 4577–4598
MYU Tokyo

S & M 4583

Virtual Sensing for Real-time Emotion Monitoring in Interactive Media Systems Using Bullet Screen

Zhong-Jie Liu and Shih-Pang Tseng*

School of Software and Big Data, Changzhou University of Information Technology,
No. 22, Mingxin Middle Road, Changzhou, Jiangsu 213164, China

(Received February 10, 2026; accepted August 4, 2026)

Keywords: bullet screen comments, abnormal user detection, fine-grained emotion analysis, BTM topic clustering, TextRCNN

Bullet screen comments on interactive video platforms are a rich source of real-time emotional signals. However, their anonymity, brevity, and contextual dependence pose significant challenges for analysis. To address the challenges, we developed a virtual sensing architecture tailored to bullet screen data, integrating abnormal user detection, fine-grained emotion inference, and thematic clustering. A random-forest-based denoising module achieved 91.99% accuracy, effectively filtering malicious or noisy accounts and securing high-fidelity input streams. For emotion recognition, the improved recurrent convolutional neural network for text (TextRCNN) model incorporating pretrained word vectors and enhanced feature fusion showed a 71.53% accuracy, outperforming baseline models such as support vector machine (58.2%), TextCNN (65.4%), and standard TextRCNN (68.1%). The model demonstrated strong recognition of distinct emotions, particularly joy and anger. To address short-text sparsity, the biterm topic model was constructed, yielding a coherence score of 0.58, significantly higher than latent Dirichlet allocation at 0.41, and successfully clustering comments into interpretable themes such as character discussion, production evaluation, and criticism. Results were visualized on a real-time monitoring dashboard, enabling platform-wide emotional health assessment. By translating noisy social signals into high-fidelity emotional and thematic information, the developed architecture supports the digital-twin construction of collective sentiment and advances sensor technology for next-generation interactive media systems.

1. Introduction

Bullet screen video platforms have become high-velocity sources of real-time social signals. Known as Danmaku, these interactive systems enable users to overlay time-locked comments directly onto video content at specific timestamps.(1) The synchronous data are used to capture immediate audience reactions to visual and auditory stimuli,(2) transforming video platforms into distributed sensor networks where each comment represents a discrete emotional observation.(3) Because these comments are related to precise video moments, they enable fine-

*Corresponding author: e-mail: tsengshihpang@czcit.edu.cn
https://doi.org/10.18494/SAM6288

ISSN 0914-4935

grained sensing of human emotion, offering valuable insights for human–computer interaction (HCI) and digital emotional health assessment.[4]

However, extracting reliable information from such data streams remains challenging. High noise levels from abnormal accounts,[5] short-text sparsity, internet slang,[6] and limited feature extraction in traditional topic models such as latent Dirichlet allocation (LDA) reduce emotional sensitivity and interpretability.[7] Therefore, digital ecosystems integrate affective computing and social sensing.[8] Social sensing treats human digital activity as a network of virtual sensors that capture behavioral and societal phenomena without physical hardware,[9] whereas affective computing provides algorithms to recognize and process emotional signals from multimodal inputs. Applied to live streams, these approaches convert raw commentary into continuous emotional telemetry for real-time public opinion monitoring.[10]

However, the integration of affective computing and social sensing into high-concurrency short-text streams such as bullet screens remains difficult. Deep learning models often introduce latency, limiting their practicality for real-time dashboards,[11] and conventional social sensing frameworks assume clean input data and fail to account for platform-specific noise from bots, spam, and semantic sparsity. To overcome these limitations, we developed a virtual sensing architecture that models the social data pipeline as a physical-to-digital transducer engine, whereas conventional social sensing frameworks assume clean input data and overlook platform-specific noise from bots, spam, and semantic sparsity. Its multifeature random forest algorithm filters abnormal accounts to ensure high-fidelity input streams.[12] An improved recurrent convolutional neural network for text (TextRCNN) integrates pretrained word vectors and feature fusion mechanisms to enhance classification across seven emotion categories: joy, anger, sadness, fear, disgust, surprise, and neutral.[6,13] A biterm topic model (BTM) mitigates short-text sparsity and aggregates user-focused thematic issues. Finally, a real-time monitoring dashboard built on a decoupled architecture visualizes platform-wide emotional health.[14,15]

By treating social interactions as virtual sensors, the proposed architecture applies sensing principles to digital-twin social monitoring and HCI. The results establish a foundation for scalable real-time social monitoring and demonstrate how live social telemetry can advance sensor engineering in next-generation interactive systems.[16]

2. Literature Review

2.1 Sensing architecture for social signal data acquisition

In virtual sensing, data acquisition is conducted to capture raw signals from the digital environment in the hardware interface. For high-velocity digital platforms such as Bilibili, traditional crawling methods are insufficient owing to the massive, real-time nature of bullet screen data streams.[5] Bilibili is China's leading video-sharing platform for Generation Z and young millennials. Often compared with YouTube, Bilibili is operated on a highly interactive, community-driven ecosystem. Originally focused on animation, comics, and games, it has expanded its services to provide technology, education, and lifestyle content. Modern sensing architectures require high concurrency to maintain signal integrity against complex anti-

acquisition mechanisms, such as dynamic encryption and identity authentication.[17] In the architecture, the access control group integrates distributed asynchronous task queues and high-speed cache databases into a high-throughput sensing gateway to enhance the scalability of data acquisition.[18] The architecture ensures that the virtual sensor transitions from local data retrieval to seamless real-time stream acquisition, providing social signals for user emotional analysis.

2.2 Signal denoising

A critical challenge in sensor technology is the separation of signal from noise. In social data acquisition using virtual sensing, noise is generated by abnormal users, such as spammers or bot accounts, who try to manipulate digital signals through repetitive or biased inputs, compromising the reliability of the data.[19] Although behavioral and content features are used for bot detection,[12] specialized signal denoising layers are required to remove noise, especially for the bullet screen technology. Bullet screen data are characterized by high anonymity and extreme sparsity, which traditional methods often fail to filter. Therefore, it is necessary to implement a multifeature fused random forest filter module to treat sending frequency and account statistics as indicators of the signal-to-noise ratio (*SNR*). By identifying and filtering automated or inactive accounts, a high-fidelity data foundation for the virtual sensing engine can be ensured.

2.3 Inference using soft sensing: TextRCNN

The transduction of raw text into measurable emotional states requires a sophisticated inference engine. Traditional feature extraction methods, including support vector machine (SVM) and term frequency-inverse document frequency (TF-IDF), fail to ensure the temporal sensitivity to capture the sequential dependencies inherent in human communication.[20,21] CNNs can capture local patterns in the text, but they struggle with long-distance semantic dependencies.

The RCNN architecture is constructed on the basis of a hybrid sensing approach, combining the sequential tracking of recurrent neural networks (RNNs) with the local feature extraction of CNNs.[22] This architecture functions as a multidimensional feature transducer, using bidirectional recurrent structures to sense context before a max-pooling layer selects discriminative emotional features. The TextRCNN model serves as a fine-grained virtual sensor, showing higher selectivity across basic emotional modalities.[23,24]

2.4 Topic clustering using BTM

In signal processing, identifying hot issues and their clusters in sparse data is a serious challenge. Traditional topic models such as LDA have a parsimony threshold when processing short-text signals, as the co-occurrence patterns within a single reading (comment) are insufficient.[7] BTM addresses this by shifting the sensing perspective from the individual document to the entire corpus level.[14] By modeling unordered biterm (word pair) co-

occurrences, BTM effectively aggregates sparse information into robust topic clusters.[25] BTM enables the architecture to identify thematic focus accurately, demonstrating significant advantages in coherence and reliability over traditional LDA-based sensing methods.[26]

The previous study resulted show the necessity of unifying abnormal user detection, emotion inference, and thematic clustering into a single architecture to overcome the inherent sparsity and noise challenges of interactive media. By treating social interaction as a sensing modality, digital-twin social monitoring technology can be developed. Specifically, bullet screen comments can be conceptualized as high-frequency soft sensor signals that reflect collective emotion,[15] and the architecture maintains fidelity through a multistage pipeline of acquisition, denoising, inference, and clustering.

3. Methodology

3.1 System design

Existing TextRCNN and BTM models need to be improved to address the persistent challenge of signal sparsity in interactive media. Therefore, a synergistic sensing architecture was constructed in this study so that abnormal user detection can be used as a text filter, improving signal quality through a multifeature fused random forest mechanism. By treating sending frequency and account statistics as indicators of *SNR*, the architecture can effectively identify and remove automated or inactive accounts such as bot accounts or inactive users, establishing a high-fidelity data foundation for the virtual sensing engine. In the architecture, the TextRCNN model functions as a fine-grained virtual sensor,[23] enhancing selectivity across basic emotional modalities.[6] The BTM model better captures thematic coherence and reliability than traditional LDA-based approaches.[26] These components converge in a social monitoring dashboard that integrates the pipeline into an end-to-end system for real-time digital-twin observation of user emotion.

3.1.1 Signal flow

We constructed a virtual sensing architecture that extracts and analyzes emotional signals from interactive media platforms. In the architecture, bullet screen comments are treated as soft sensing signals that reflect user emotion.[15] To ensure high-fidelity sensing, the architecture adopted a decoupled framework comprising four functional layers (Fig. 1).

The view layer functions as the real-time monitoring dashboard, developed with the progressive JavaScript framework Vue.js (VUE) and the data visualization library DataV. It converts complex emotional data into intuitive visual signals for end users on the user interface (UI). The control layer serves as the system's routing hub. It employs the fast application programming interface (FastAPI) to manage high-concurrency requests, the distributed task queue (Celery) to handle asynchronous background tasks, and the remote dictionary server (Redis) as a message broker to ensure smooth, low-latency transfer of signal data between layers. At the core, the service layer executes inference logic. Raw text signals are processed through

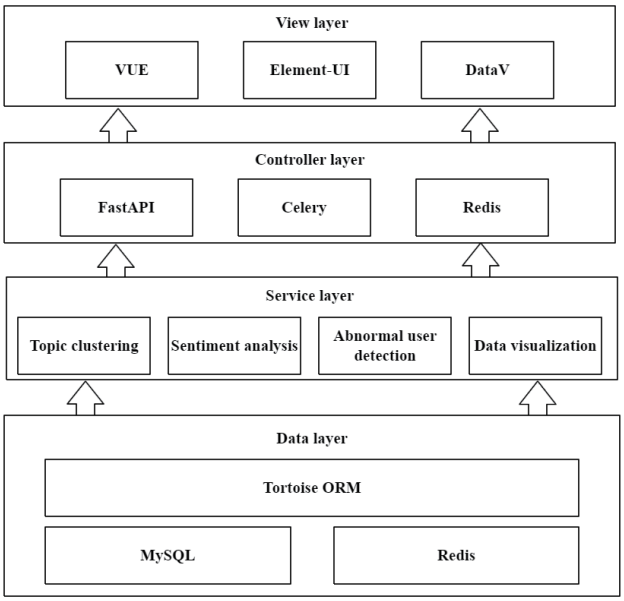

Fig. 1. Virtual sensing architecture for bullet screen technology.

specialized algorithms: random forest for signal denoising, the improved TextRCNN for fine-grained emotional sensing, and the BTM for thematic clustering. The data layer provides persistent storage and rapid retrieval of sensing logs and historical emotion trends. It uses structured query language (MySQL) for robust relational storage, Redis for caching and fast access, and the Tortoise object-relational mapping (Tortoise ORM) framework to enable object-oriented interaction with the database, simplifying data management. The four layers form a decoupled sensing framework in which each component contributes to the overall reliability and fidelity of the virtual sensing architecture.

3.1.2 Functional module

The architecture integrates four modules: data collection and preprocessing, emotion analysis, abnormal user detection, and visual display (Fig. 2). These modules transform unstructured social interaction data into actionable emotional metrics, enabling reliable emotion monitoring in real time. Owing to the complexity of real-time social data, a layered engineering approach is employed in the architecture.

3.1.2.1 Data collection and preprocessing module

To improve signal quality, raw bullet screen comments undergo a systematic conditioning pipeline. Duplicate and meaningless comments are removed to ensure corpus purity.[27] Text

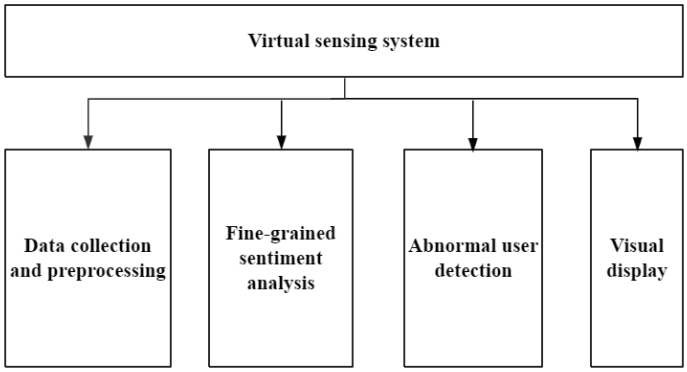

Fig. 2. Functional modules of virtual sensing architecture.

normalization is conducted to unify encoding and correction consistencies, whereas segmentation is performed with Jieba and custom dictionaries for the recognition of domain-specific slang. Structural mapping is used to convert timestamps and user identifications into numerical indices, preparing the data for inference. The data collection layer acts as the physical-to-digital interface, using Celery-scheduled tasks and an asynchronous Hypertext Transfer Protocol client/server framework for Python for high-concurrency extraction of bullet screen data from the Bilibili platform.

3.1.2.2 Emotion analysis module

The emotion analysis module functions as a multidimensional virtual sensor to categorize emotional states into seven modalities: joy, anger, sadness, fear, disgust, surprise, and neutral.[28] To ensure a sufficient scale for deep learning training and enhance cross-domain generalization, the improved TextRCNN model was trained on the open-source weibo_senti_100k benchmark dataset comprising 27354 annotated Chinese social text records spanning seven emotion categories.[29] Furthermore, the model embeds pretrained Chinese social corpora and introduces enhanced feature fusion mechanisms. The virtual sensing architecture integrates bidirectional long short-term memory (BiLSTM) and rectified linear unit-based fusion for context encoding, and 1D-CNN classification with max-pooling, enabling high selectivity across emotional modalities.[13]

3.1.2.3 Abnormal user detection

Since social sensing systems are vulnerable to malicious noise, an abnormal user detection module filters automated or inactive accounts that distort emotion data streams. A dedicated dataset of 6800 records from Bilibili's Girls' Frontline 2 community is used for behavioral noise filtering, using reverse engineering with encryption parameters and a three-tier manual labeling

process. Although relatively compact, this dataset is used to target high-density behavioral indicators such as posting frequency, account level, and fan badges, tailored to platform-specific bot signatures.

3.1.2.4 Visual display

To overcome the sparsity of short bullet screen comments, the architecture employs BTM.[30] BTM aggregates word pairs across the entire corpus, building a global co-occurrence set that improves semantic coherence. Model parameters, such as topic-word distribution and topic distribution, are optimized through Gibbs sampling, with perplexity guiding topic number selection. The Gibbs sampling method is a Markov Chain Monte Carlo algorithm used to estimate the hidden thematic structures of the bullet screen data stream. The final output shows thematic clusters such as character discussion or plot critiques and word clouds on dashboards (Fig. 5) for the precise observation of user concerns. The visualization layer shows the processed emotional metrics on dashboards, supporting scientific assessment and decision-making.[31]

3.2 Operation

To ensure high-fidelity outputs, a dedicated signal filtering module was used to identify and remove automated or inactive accounts. Because no standardized datasets exist for bullet screen behavior, a custom dataset was constructed from the Bilibili Girls' Frontline 2 community, a high-velocity environment known for abnormal activity. Data acquisition required reverse engineering of dynamic encryption parameters [buvid3 (a unique hardware-bound identifier), sign (a dynamic digital signature generated by hashing request parameters with a secret key), and raw_wbi_key (a rotating user-specific encryption key used for web-based interface signing)] to stabilize access to user attributes and behavioral histories.

A three-tier manual labeling process was used to produce 6800 records, including initial screening for anomalous emotion patterns, rescreening based on account levels and cross-platform similarity, and final validation to distinguish automated from natural interaction. Raw behavioral features, such as username patterns, avatar defaults, signature presence, fan badges, and activity levels, were encoded into numerical inputs for classification. Among the evaluated models (K-nearest neighbors [KNN], decision tree, and random forest), the random forest model showed the best performance, with 91.99% accuracy. Its ensemble structure mitigated overfitting and provided a reliable *SNR* indicator for the sensing pipeline.

The fine-grained emotion analysis module mapped raw text signals into seven emotional modalities: joy, anger, sadness, fear, disgust, surprise, and neutral. To address the limitations of conventional RNNs and CNNs, an improved TextRCNN was used. TextRCNN integrates a bidirectional LSTM network to capture contextual dependencies, fused with pretrained word embeddings through rectified linear unit activation for enhanced nonlinear representation (Fig. 3). A one-dimensional convolutional layer extracts local features at multiple scales, followed by max-pooling to select the most discriminative signals. Classification was performed through a Softmax function, yielding an accuracy of 71.53% and enabling precise digital-twin observation of emotion trends.

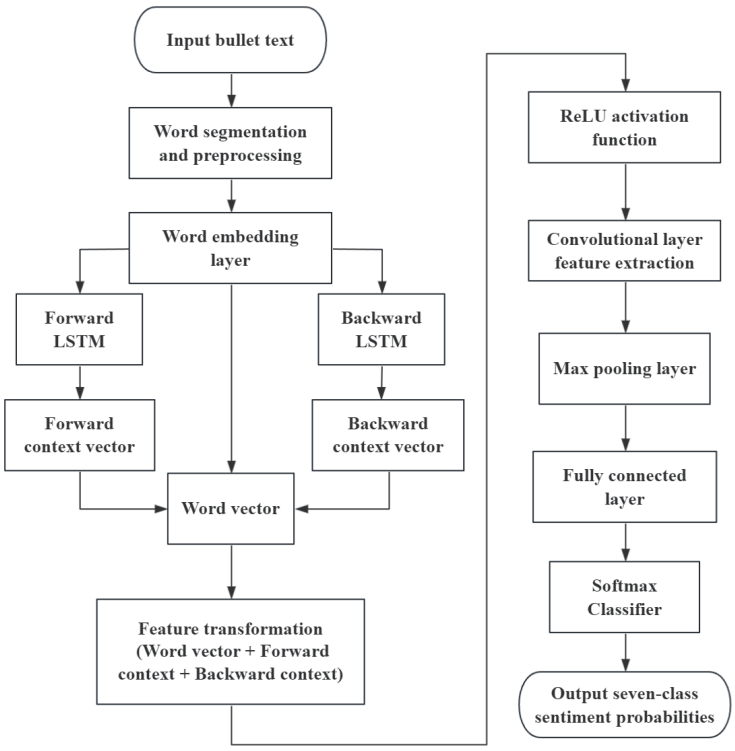

Fig. 3. Workflow of TextRCNN (ReLU: rectified linear unit).

To ensure complete experimental reproducibility, the improved TextRCNN sentiment model was trained on an NVIDIA RTX 4090 graphics processing unit (GPU) using PyTorch 2.1. The complete hyperparameter configuration used during model optimization and testing is detailed in Table 1.

To overcome the sparsity of short-form interactive media, thematic clustering is performed using BTM. Whereas LDA models topics within individual documents, BTM aggregates unordered word pairs (biterms) across the entire corpus to construct a global co-occurrence set (Fig. 4). The hyperparameters utilized for Gibbs sampling and Dirichlet priors are presented in Table 2. This approach transforms sparse inputs into robust thematic signals. Topic distribution and topic-word distribution were estimated as model parameters through Gibbs sampling, which was used for the selection of the optimal number of topics. The resulting clusters are projected on the dashboard, automatically categorizing signals into actionable themes such as plot critiques or character discussion. This process enhances semantic coherence and supports precise public opinion guidance.

3.3 Performance evaluation

3.3.1 Evaluation metrics

The performance of the inference engine was evaluated using standard statistical measures to provide an objective assessment of the system's ability to distinguish signal from noise and

Table 1
Hyperparameters of improved TextRCNN model in this study.

Hyperparameter	Value (option)	Description
Pretrained word embedding	Chinese WIKI word to vector	300-dimensional dense word vectors
Embedding dimension	300	Word vector representation dimension
BiLSTM hidden size	256	Contextual vector size per direction
CNN kernel sizes	[3,4,5]	Multiscale local feature extraction kernels
Number of feature maps	128 per kernel	Convolutional channel depth
Activation function	ReLU	Nonlinear feature fusion activation
Dropout probability	0.5	Regularization to prevent overfitting
Optimizer	AdamW	Weight decay-adjusted optimizer
Initial learning rate	10−3	Decay factor 0.1 every 5 epochs
Batch size	64	Micro-batch size for GPU execution
Training epoch	25	Early stopping with patience = 4
Loss function	Cross-entropy	Weighted across 7 emotion modalities

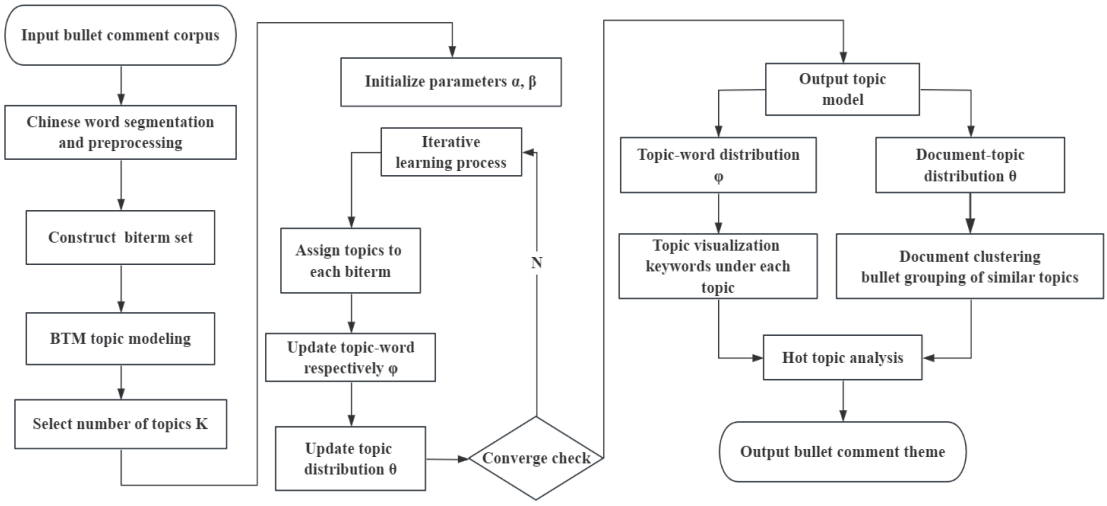

Fig. 4. Workflow of BTM.

Table 2
Hyperparameter configuration for BTM.

Hyperparameter	Value (option)	Description
Number of topics (*K*)	40	Evaluated using maximum coherence score
Document-topic prior	50/K = 1.25	Symmetric Dirichlet prior on topic distribution
Topic-word prior	0.01	Symmetric Dirichlet prior on word distribution
Gibbs sampling iterations	1000	Burn-in period = 200 iterations
Biterm extraction window	Whole document	All unordered word pairs per short comment
Top keywords per topic	20	Key terms used for coherence score and human evaluation
Minimum frequency of vocabulary	5	Words appearing less than 5 times were filtered out

accurately categorize emotional modalities.[15] For the abnormal user detection (binary classification) and emotion inference (multiclass classification) modules, evaluation metrics were calculated from the confusion matrix components: true positive (*TP*), true negative (*TN*), false positive (*FP*), and false negative (*FN*). Accuracy was calculated as the overall proportion of correctly identified signals within the dataset.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{1}$$

Precision was calculated as a proportion of correct identifications to assess the fidelity of the sensor.

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

Recall represents the sensitivity of the sensor to estimate its ability to capture all relevant signals.

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

F1-score is the harmonic mean of precision and recall, balancing fidelity and sensitivity.

$$F1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \tag{4}$$

These metrics were calculated to estimate the sensing modules' reliability and responsiveness in filtering noise and detecting emotional signals. To evaluate the semantic quality and interpretability of topics generated by BTM and LDA, the coherence score was calculated[32] by integrating a sliding window, normalized point-wise mutual information (NPMI), and indirect cosine similarity to capture both direct and contextual associations. For a topic t defined by its top 20 keywords $W_t = \{w_1|w_2|...|w_{20}\}$, the coherence score is computed in four stages.

First, co-occurrence counts for word pairs (w_i, w_j) are extracted using a sliding window of a size of 11. Second, NPMI is calculated as the strength of association between word pairs.

$$\mathrm{NPMI}(w_i \mid w_j) = \frac{\ln \dfrac{P(w_i, w_j) + \varepsilon}{P(w_i) \cdot P(w_j)}}{-\ln\left(P(w_i, w_j) + \varepsilon\right)}, \quad \varepsilon = 10^{-12} \tag{5}$$

Here, $P(w_i)$ and $P(w_j)$ are marginal probabilities and $P(w_i, w_j)$ is the joint probability. By adding ϵ, the calculation of NPMI remains numerically stable, ensuring that rare or unseen word pairs do not cause computational errors. In practice, ϵ does not significantly affect the results because it is chosen to be so small that it only stabilizes the denominator without altering meaningful probability values. Third, each keyword w_i is represented by an N-dimensional context vector $v(w_i)$, with elements corresponding to NPMI correlations with other keywords. Finally, the coherence score (C_v) for topic t is calculated as the mean cosine similarity between each keyword vector and the centroid vector $v(W_t) = \sum_{j=1}^{N} v(w_j)$.

$$C_v(t) = \frac{1}{N}\sum_{i=1}^{N} \cos(v|(w_i)v(W_t)) \tag{6}$$

The overall corpus coherence score is the mean across all K topics.

$$C_v = \frac{1}{K}\sum_{t=1}^{K} C_v(t) \tag{7}$$

C_v ranges from 0 to 1. 1 indicates stronger semantic cohesion among topic keywords. A high score reflects meaningful clusters such as male lead, female lead, and ending, while a low score signals incoherent groupings such as hahaha, visuals, or encryption. Therefore, the coherence score is used to validate the virtual sensor's ability to aggregate sparse biterms into semantically consistent clusters, enhancing interpretability and alignment with human perception.[16,30]

A tiered comparison strategy was adopted to validate the architectural improvement of the virtual sensing pipeline. Each benchmark was selected to highlight distinct aspects of sensor capability, such as denoising, emotional transduction, and thematic aggregation. To complement the automated coherence score, topic interpretability was evaluated using both human qualitative scoring and quantitative semantic vector similarity.[32] Human qualitative evaluation was conducted with three independent annotators specializing in digital media and natural language processing. They blindly reviewed topic word lists (top 20 words per topic) generated by BTM and LDA. Topics were assessed using three metrics. Topic purity was measured as the proportion of top words directly related to the primary theme. Topic distinctiveness was calculated to compute the proportion of topics representing unique, non-overlapping themes. Human interpretability was rated on a 5-point Likert scale, where 1 indicated incoherent or noisy topics, and 5 indicated highly coherent topics easily labeled by human observers. In addition, intratopic semantic similarity was calculated by using the pairwise cosine similarity of pretrained word embeddings among the keywords within each topic vector, providing a measure of semantic clustering density in vector space.

$$Sim_{topic} = \frac{2}{N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} \cos\left(e_{w_i}, e_{w_j}\right) \tag{8}$$

Here, e_{w_i} and e_{w_j} are the dense word embeddings of topic keywords w_i and w_j.

3.3.2 User detection comparison

Three machine learning models were used to evaluate the signal filtering module.[27] A KNN model was adopted as a baseline model for clustering by feature proximity. A decision tree model was used to assess behavioral features such as account level and posting frequency. Random forests aggregated multiple trees to reduce variance and stabilize detection of noisy

signals, identifying automated or inactive accounts more reliably than single models. To validate the performance of the improved TextRCNN, comparative experiments were conducted. SVM adopted TF-IDF features, showing the limits of frequency-based models in capturing emotion.[20] TextCNN was used to detect local n-gram patterns but failed to model long-range semantics.[33] The comparison results were analyzed to validate TextRCNN's ability to integrate contextual and local features.

An ablation study was also conducted to examine the contributions of each component. The baseline TextRCNN model was tested for the hybrid recurrent–convolutional structure. Pretrained word vectors were added to calibrate linguistic features for Chinese social slang. By incorporating improved feature fusion with a rectified linear unit and global and local signals, emotional representation was enriched. The full virtual sensing architecture, including abnormal user filtering, achieved peak fidelity. Performance was measured with accuracy, precision, recall, F1-score, and coherence to ensure signal fidelity and thematic consistency.

TextRCNN was further compared with both traditional and transformer baselines. SVM and TextCNN served as lightweight references, whereas bidirectional encoder representations from transformers (BERT)-based-Chinese and robustly optimized BERT approach-whole word masking-extended (RoBERTa-wwm-ext) modeled contextual semantics with self-attention.[34,35] Performance metrics (accuracy, F1-score) and computational efficiency metrics (parameter size, inference latency) were calculated to examine TextRCNN's suitability as a real-time virtual transducer in high-velocity interactive media streams.

3.3.3 Computational complexity and system latency evaluation

The computational complexity of the model was determined by using three service components. Random forest denoising has a time complexity of $O(T \cdot D \cdot \log N)$, where T = 100 is the number of trees, D is feature depth, and N is sample size. Its space complexity is calculated as $O(T \cdot \text{nodes})$. Improved TextRCNN inference has a time complexity of $O(L \cdot H^2 + L \cdot H \cdot K)$ for a sequence of length L, hidden dimension H, and kernel size K. Space complexity is bounded by recurrent and convolutional weights, $O(H^2 + H \cdot K)$. BTM topic clustering has a time complexity of $O(B \cdot K)$ per Gibbs sampling iteration, where B is the total number of biterms and K is the topic count. To evaluate real-time performance under high concurrency, an asynchronous stress-testing framework using Locust was implemented. Bullet screen comment streams were simulated across five concurrency tiers (100, 500, 1000, 2500, and 5000 concurrent request/s). System response was measured across four parameters: API gateway ingestion latency (T_{gate}), denoising and feature preprocessing latency ($T_{denoise}$), model inference latency (T_{infer}), and total end-to-end latency ($T_{e2e} = T_{gate} + T_{denoise} + T_{infer} + T_{db}$).[36]

3.3.4 System integration and deployment

To evaluate end-to-end cohesion and elasticity, the multimodule virtual sensing architecture was deployed in a containerized environment using Docker and Kubernetes. Automated

integration and scalability testing was then performed.[37] An integration test was conducted to validate pipeline telemetry across microservice boundaries, including the ingestion gateway, denoising classifier, TextRCNN transducer, BTM engine, and relational storage. A suite of 100000 synthetic test comments was dispatched to measure transaction success rates, API payload validation, and database serialization efficiency. Deployment scalability was assessed through Kubernetes Horizontal Pod Autoscaling (HPA) on an edge-cloud server cluster. Pod replication was triggered automatically when average CPU usage exceeded 70% or when Redis message queue depth surpassed 500 pending tasks.[38] Simulated load ramps ranging from 500 to 10000 request/s were injected using Locust to measure scale-out response times and failure recovery.

3.3.5 Ethical considerations and data privacy

Virtual sensing systems capture real-time user signals, requiring strict privacy protocols to protect individual rights. All data acquisition and processing in this study followed platform terms of service and relevant digital privacy guidelines through a three-stage method.[39]

First, cryptographic identifier masking was applied at the API gateway layer. Primary user identifiers, client Internet Protocol addresses, and hardware-bound tokens (buvid3) were immediately transformed using a one-way SHA-256 hash with a daily rotated salt [(H_s(UID) = SHA-256(UID||Salt))], preventing reverse lookup or cross-platform tracking. Second, personally identifiable information filtering and text sanitization were performed during preprocessing. Raw bullet screen texts underwent automated named entity recognition (NER), and any detected personal names, phone numbers, external account handles, or precise location coordinates were replaced with standardized tokens. Finally, nonintrusive aggregation and storage ensured that only aggregated emotional probability vectors and thematic cluster summaries were retained in the persistent relational database (MySQL). Individual raw text entries were stored temporarily in volatile memory (Redis cache) only for inference and then discarded, guaranteeing zero long-term retention of unhashed user commentary.

4. Result

4.1. Abnormal user detection

The abnormal user detection module was evaluated using three classifiers: KNN, decision tree, and random forest. On the 6,800-record dataset, random forest showed the best performance with an accuracy of 91.99%, a precision of 91.9%, a recall of 95.6%, and an F1-score of 92.9 (Table 3). KNN showed an 89.7% accuracy, and the decision tree's accuracy was 90.1%. These results show that ensemble-based filtering provides superior robustness against noisy signals, effectively functioning as a denoising sensor that stabilizes *SNR* for subsequent inference layers.

Table 3
Results of abnormal user detection by different algorithms.

Algorithm	Accuracy	Precision	Recall	F1-score
KNN	0.8968	0.8973	0.9181	0.9076
Random forest	0.9199	0.9194	0.9563	0.9291
Decision tree	0.9160	0.9013	0.9439	0.9254

4.2 Emotion inference

The improved TextRCNN model showed a 71.53% accuracy in fine-grained emotion classification, consistently outperforming baseline models across all evaluation dimensions, including overall precision (0.7189), recall (0.7125), and F1-score (0.7157) (Table 4). The virtual sensing architecture showed strong recognition for distinct emotions such as joy and anger, whereas overlapping semantics (surprise versus fear) remained challenging. This demonstrates the capability of the TextRCNN as a virtual emotional sensor, selectively transducing raw text into structured emotional modalities. The performance of the improved TextRCNN model was compared with that of current state-of-the-art natural language processing models; fine-tuned Chinese BERT-based and RoBERTa-wwm-ext models were evaluated under identical training and testing conditions.

As shown in Table 5, TextRCNN requires less than 12% of the parameter capacity of BERT/RoBERTa and processes inference batches 6.8 times faster (12.4 and 84.6 ms), making it the optimal transducer for high-concurrency virtual sensing dashboards.

To evaluate performance across specific modalities, per-class metrics were calculated across the seven emotional categories (Table 6). High performance was observed in high-frequency, highly distinct emotions such as joy (F1 score = 0.7842) and anger (F1 score = 0.7420). Although transformer models such as RoBERTa achieved a 2.68% higher accuracy than our improved TextRCNN, their deployment in a real-time interactive stream system is constrained by computational latency and hardware demands.

The normalized confusion matrix shows strong diagonal selectivity. Misclassifications predominantly occurred between semantically overlapping modalities, such as surprise being misclassified as fear (11.8%) or sadness being confused with neutral (10.2%), reflecting subtle linguistic boundaries in short-text bullet comments (Table 7).

4.3 Performance evaluation

To assess the contribution of architectural improvements, three complementary analyses were conducted. The improved TextRCNN consistently outperformed baseline models across accuracy, precision, recall, and F1-score, validating the integration of pretrained embeddings and enhanced fusion mechanisms. Pretrained word vectors improved accuracy by 1.75%, whereas the fusion mechanism added 0.87%. A 3.44% gain was observed over the baseline TextRCNN, confirming the necessity of both components (Table 8).

Table 4
Bullet screen emotion analysis test results for different models.

Model	Accuracy	Precision	Recall	F1-Score
Traditional SVM	0.582	0.5785	0.5801	0.5792
Standard TextCNN	0.6541	0.657	0.6489	0.6529
Standard TextRCNN	0.681	0.6835	0.6782	0.6808
BERT-base-chinese	0.7312	0.7345	0.728	0.7312
RoBERTa-wwm-ext	0.7421	0.745	0.7392	0.7421
Improved TextRCNN (this study)	0.7153	0.7189	0.7125	0.7157

Table 5
Performance metrics of improved TextRCNN model in emotion classes.

Emotion class	Precision	Recall	F1-score
Joy	0.7915	0.777	0.7842
Anger	0.751	0.7332	0.742
Sadness	0.7042	0.718	0.711
Fear	0.672	0.6545	0.6631
Disgust	0.698	0.689	0.6935
Surprise	0.661	0.6412	0.6509
Neutral	0.7545	0.7746	0.7644

Table 6
Efficiency and throughput trade-off for TextRCNN and transformer architectures (tested on NVIDIA RTX 4090, batch size = 64).

Model	Number of parameters (million)	Inference latency (ms/batch)	Throughput (samples/s)
TextCNN	5.2 M	6.1 ms	Upto 10490
Improved TextRCNN	12.8 M	12.4 ms	Upto 5160
BERT-base-Chinese	110.0 M	72.5 ms	Upto 880
RoBERTa-wwm-ext	110.0 M	84.6 ms	Upto 755

Table 7
Normalized confusion matrix for seven emotion classes.

Emotion prediction	True emotion value						
	Joy	Anger	Sadness	Fear	Disgust	Surprise	Neutral
Joy	0.777	0.021	0.015	0.012	0.018	0.045	0.112
Anger	0.018	0.733	0.052	0.038	0.089	0.015	0.055
Sadness	0.012	0.048	0.718	0.055	0.042	0.023	0.102
Fear	0.015	0.042	0.061	0.655	0.045	0.118	0.064
Disgust	0.021	0.095	0.048	0.032	0.689	0.031	0.084
Surprise	0.052	0.022	0.031	0.118	0.025	0.641	0.111
Neutral	0.082	0.031	0.052	0.018	0.022	0.02	0.775

Table 8
Ablation study results on accuracy.

Configuration	Accuracy	Improvement over baseline (%)
Standard TextRCNN (Baseline)	0.6810	—
Standard TextRCNN + Pre-trained word vectors	0.6985	+1.75
Standard TextRCNN + Pre-trained word vectors + improved feature fusion	0.7072	+2.63
Virtual sensing architecture	0.7153	+3.44

Misclassifications occurred due to sarcasm, simplified expressions, and fuzzy category boundaries. These highlight the limitations of current sensors in decoding implicit or ambiguous signals, pointing to the need for multimodal sensing. BTM presented higher coherence scores than LDA, demonstrating its effectiveness in aggregating sparse short-text signals into coherent thematic clusters. Representative topics included character discussion, production evaluation, positive experience, and criticism (Tables 9 and 10).

BTM demonstrated superior performance over traditional LDA across automated coherence, intratopic semantic similarity, and expert human evaluation metrics (Table 11). Human evaluators confirmed that LDA topics frequently suffered from word repetition and noisy, generic terms such as hahaha, visuals, and encryption co-occurring under unrelated themes, leading to low topic purity (62.1%). In contrast, BTM effectively aggregated co-occurring biterms across sparse bullet comments into highly distinctive and interpretable clusters (purity: 86.4%, interpretability: 4.42/5.0), showing a distinction between plot critiques, production aesthetics, and character sentiment.

The improved BTM showed a mean coherence score of 0.580 across K = 4000 topics, outperforming traditional LDA (coherence score = 0.410$). By computing context vector similarities across aggregated biterm co-occurrences rather than sparse single-document occurrences, BTM effectively maximized NPMI scores among top keywords, resulting in statistically robust topic cohesion.

Table 9
BTM topic clustering results.

Topic summary	Representative keywords
Character and emotion discussion	Male lead, female lead, love, cried, ending
Production technique evaluation	Visuals, special effects, budget, production, quality
Positive experience and entertainment	Hahaha, positive review, cute, hilarious, fun
Negative evaluation and criticism	Awkward, ridiculous, boring, critique, dropped series

Table 10
Coherence scores of BTM and LDA.

Number of topics	LDA	BTM
4000	0.41	0.58
6000	0.38	0.55
8000	0.35	0.52

Table 11
Human evaluation and intratopic semantic similarity for BTM and LDA (4000 topic configuration).

Evaluation metric	LDA	BTM	Improvement (%)
Coherence score	0.41	0.58	41.50
Intratopic semantic similarity	0.418	0.642	53.60
Human interpretability score (1–5 point)	3.15	4.42	40.30
Topic purity (%)	62.10	86.40	24.30
Topic distinctiveness (%)	58.30	82.50	24.20

4.4 System throughput, latency, and scalability

The empirical stress-testing result showed that the decoupled virtual sensing architecture maintains stable sub-100 ms latency across high concurrency levels (Table 12). At an operational streaming volume of 2500 requests/s, the architecture developed in this study showed a 99.8% success rate with a mean end-to-end latency of 42.8 ms, well below the 100 ms threshold required for real-time visualization dashboards. The asynchronous coroutine architecture (FastAPI + Celery + Redis broker) effectively decoupled task queuing from heavy model inference, preventing bottlenecks during high-density live streaming events.

4.5 Data visualization

The integrated dashboard visualizes outputs from all sensing modules: emotion distributions, abnormal user detection, topic clusters, and word clouds (Fig. 5). This interface transforms raw sensing outputs into visible ones, enabling administrators to monitor emotional health and public opinion trends in real time.

Table 12
System-wide latency breakdown and throughput under varying real-time concurrency loads.

Concurrency (Request/s)	Success rate (%)	Gateway latency (T_{gate}, ms)	Denoising latency ($T_{denoise}$, ms)	Inference latency (T_{infer}, ms)	End-to-end latency (T_{e2e}, ms)	CPU/graphical processing unit load (%)
100	100.0	1.2	2.1	8.4	14.2	12/15
500	100.0	2.5	3.2	10.1	18.8	24/32
1000	100.0	4.1	4.5	12.4	24.2	41/55
2500	99.8	8.2	8.1	18.5	42.8	72/84
5000	97.4	19.5	16.2	34.1	82.3	91/98

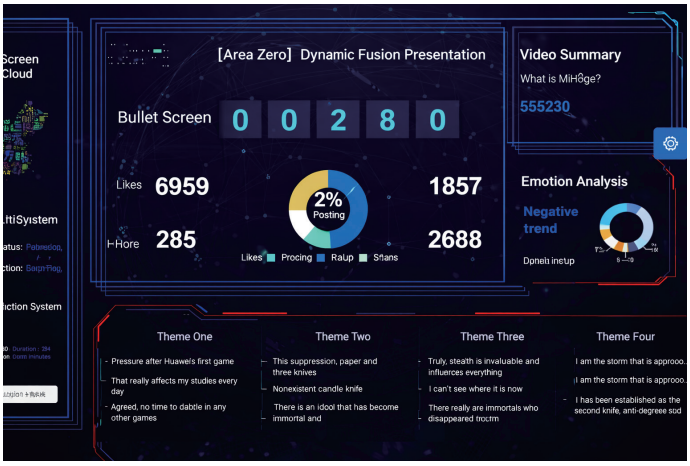

Fig. 5. (Color online) Dashboard presenting outputs.

5. Discussion

The performance evaluation results show that the developed architecture operates as a multi-layered sensing system, with each module fulfilling a distinct sensor role validated by quantitative performance metrics. The random forest classifier showed superior robustness in filtering out automated or inactive accounts, achieving high precision and recall. These metrics validate the developed architecture's role as a denoising sensor, stabilizing *SNR*, and ensuring that subsequent inference engines operate on clean, high-fidelity inputs.

The improved TextRCNN consistently outperformed baseline models, with accuracy gains through comparative experiments and ablation studies. Precision and recall results highlight its balanced ability to capture distinct emotional modalities, remaining sensitive to nuanced signals. The role of the architecture as an emotional sensor is reinforced by error analysis, which identified challenges in decoding sarcasm, simplified expressions, and overlapping categories, which are limitations such as calibration issues in physical sensors. BTM showed higher coherence scores than LDA, demonstrating its effectiveness in aggregating sparse short-text signals into semantically coherent clusters. The coherence score is the thematic sensor's calibration metric, confirming its interpretability and reliability in mapping collective user concerns into actionable categories. The integrated visualization interface consolidates outputs from all modules, transforming raw sensing data into interpretable insights. By presenting emotion distributions, abnormal user detection, and thematic clusters in real time, the dashboard functions as a sensor display unit, enabling administrators to monitor emotional health and public opinion trends with clarity.

The virtual sensing architecture developed in this study extends traditional sensing concepts into the domain of social interaction. Each module plays the role of a physical sensor: filters for noise reduction (abnormal user detection), transducers for signal conversion (TextRCNN emotion inference), aggregators for higher-level interpretation (BTM clustering), and displays for real-time monitoring (dashboard). Performance metrics such as accuracy, precision, recall, F1-score, and coherence serve as calibration measures, validating the fidelity, sensitivity, and interpretability of the virtual sensing architecture. This reciprocal relationship supports sensor technology development by demonstrating how virtual sensing architectures can address sparsity, noise, and ambiguity in high-frequency interactive media, while simultaneously advancing digital-twin social monitoring.

Detailed class-level evaluation results provide information on the transducer mechanics of the soft sensing engine. Similar to physical sensors experiencing crosstalk or noise overlap in adjacent frequency bands, the virtual sentiment transducer exhibits varying sensitivity depending on the emotional modality. Explicit modalities such as joy and anger yield high F1-scores (>0.74), whereas ambiguous modalities such as surprise and fear display cross-sensitivity (an 11.8% mutual confusion). Providing per-class precision, recall, and confusion matrix breakdowns serves as a quantitative calibration matrix, offering clear pathways to refine loss functions or incorporate attention mechanisms in future model iterations.

An important requirement for physical and virtual sensor engineering is the balance between transduction sensitivity (accuracy) and response time (latency). In this study, transformer-based

architectures (BERT and RoBERTa) showed high classification accuracies, but their parameter scale (110 million) imposes substantial memory and computational overhead, leading to high latency during real-time stream processing. In high-velocity interactive media systems where thousands of bullet screen comments arrive per second, low latency is critical to prevent queue bottlenecking in the FastAPI/Celery method. The improved TextRCNN showed a superior balance, capturing complex temporal and local contextual patterns through its recurrent-convolutional hybrid structure while maintaining an ultrafast response time (12.4 ms per batch). This computational efficiency validates TextRCNN as an edge-friendly virtual sensor capable of delivering real-time sentiment telemetry without infrastructure delays.

Beyond automated coherence measures, the integration of human evaluation and intratopic vector similarity provides empirical confirmation of BTM's superior interpretability in virtual sensing applications. High-frequency short-text streams on interactive media platforms are prone to semantic fragmentation. Traditional LDA creates noisy topic distributions owing to document-level word sparsity, whereas BTM models global biterm co-occurrences. This fundamental shift yields an intratopic semantic similarity of 0.642 and a human interpretability score of 4.42/5.0, proving that the thematic sensor outputs semantically coherent, human-aligned clusters that can directly feed into public opinion monitoring dashboards without requiring manual filtering.

Stress testing and complexity modeling validate the developed architecture's operational viability as a real-time virtual sensor display engine. By leveraging asynchronous nonblocking FastAPI coupled with Celery task queues, signal acquisition and database writes are decoupled from neural inference execution. The total end-to-end processing latency remains under 45 ms at 2500 incoming comments per second, showing that the architecture achieves high-throughput telemetry without sacrificing model accuracy or causing data lag on the visual dashboard.

Although the fine-grained emotion transduction module leverages a broader, multiclass social media corpus (weibo_senti_100k) with pretrained word vectors to overcome deep learning sample constraints,[29] the abnormal user detection module relies on a specialized 6,800-record dataset from a single interactive video community. This targeted dataset effectively captures localized spamming and bot behaviors (achieving 91.99% accuracy), but its narrow scope limits direct application to non-Chinese platforms or distinct social contexts with different account interaction dynamics. Furthermore, the current virtual sensing architecture is exclusively optimized for monolingual Chinese bullet screens containing localized slang and symbols. Therefore, the developed virtual sensing architecture must incorporate cross-platform multilingual embeddings and domain-adaptation techniques to broaden generalization across diverse global user populations.

Integration testing and autoscaling comparison results showed that the virtual sensing architecture was viable for large-scale industrial deployment.[37] Microservice architectures often suffer from interservice communication latency or pipeline bottlenecking under dynamic data loads. By implementing decoupled asynchronous messaging queues (Redis/Celery) and containerized deployment orchestration with HPA (Kubernetes horizontal pod autoscaler), the developed architecture expands computational capacity in under 45 s during sudden live stream traffic spikes.[38] Maintaining an end-to-end transaction success rate of 99.85% and sub-85 ms

latency at 10000 requests per second confirms that the sensing engine scales elastically to meet real-time streaming demands.

Ethical and privacy-preserving protocols represent an indispensable signal-conditioning layer within virtual sensing systems.[40] In sensor networks, privacy safeguards must be adopted to prevent unauthorized telemetry tracking. In digital-twin social monitoring, dynamic identifier hashing and real-time personally identifiable information redaction are used to ensure that emotional transduction operates solely on collective, de-identified social signals. By decoupling individual user identities from broad public opinion metrics, the developed architecture achieves high-fidelity sentiment observation while maintaining compliance with modern data protection regulations.[39]

6. Conclusions

We developed and validated a virtual sensing architecture that treats bullet screen comments as high-frequency soft signals for real-time emotion monitoring. By combining deep learning transducers with robust signal-conditioning models, the architecture transformed unstructured social interaction data into actionable emotional and thematic information. Across multiple dimensions, the system demonstrated stronger sensor capabilities than baseline approaches. Signal fidelity was ensured through abnormal user detection, where behavioral statistics were treated as *SNR* indicators. The random forest module showed a 91.99% accuracy, confirming its effectiveness as a denoising filter in noisy digital environments. Emotional transduction was enhanced through the improved TextRCNN, which integrated pretrained word vectors and feature fusion, presenting a 71.53% accuracy with a 3.44% increase over the standard TextRCNN and outperforming SVM (58.2%) and TextCNN (65.4%). Thematic aggregation was achieved through BTM, which produced a coherence score of 0.58, representing a 41% improvement over LDA (0.41) and demonstrating its ability to cluster short-text signals into semantically consistent topics. Finally, the visualization dashboard consolidated these outputs into a unified sensor display unit, enabling the digital-twin observation of collective emotion and supporting decision-making in HCl.

This architecture shows that virtual sensors can effectively address sparsity, noise, and ambiguity in interactive media. By detecting implicit signals such as sarcasm, ambiguous emotional boundaries, and multilingual inputs, the framework advances sensor technology for next-generation interactive systems. Further research is necessary to enhance the adaptability and reliability of virtual emotional transducers in complex online environments.

Acknowledgments

This research was supported by the Jiangsu Province Industry-University-Research Collaboration Project, “Tourist Situation Analysis and Intelligent Decision-Making System Development Based on Data Collection Bracelets (Project No. BY20240343)” and the “Education and Teaching Reform Project of Changzhou University of Information Technology, Research on the Ecosystem of AI Courses and Three-Dimensional Textbook Construction in Colleges Under the Background of Digital Intelligence Empowerment (Project No. 2026CXJG05)”.

References

1 C. Liu: Interdiscip. Humanit. Commun. Stud. **2** (2024) 9. https://doi.org/10.61173/knqn8g59
2 K. Tamura and M. Katsurai: Proc. 23rd Int. Conf. Information Integration and Web Intelligence (ACM, 2021) 7. https://doi.org/10.1145/3487664.3487682
3 K. Janowicz, G. McKenzie, Y. Hu, R. Zhu, and S. Gao: Mobility Patterns, Big Data and Transport Analytics, Eds. C. Antoniou, L. Dimitriou, and F. Pereira (Elsevier, 2019) Chap. 3, 31. https://doi.org/10.1016/B978-0-12-812970-8.00003-8
4 H. Z. Huang, Z. Wang, W. Hu, C.-W. Lin, and S. Satoh: Proc. 27th ACM Int. Conf. Multimedia (ACM, 2019) 1888. https://doi.org/10.1145/3343031.3351027
5 Z. Chang: Proc. 6th Int. Conf. Control Engineering and Artificial Intelligence (ACM, 2022) 21. https://doi.org/10.1145/3522749.3523076
6 S. Minaee, N. Kalchbrenner, E. Cambria, N. Nikzad, M. Chenaghlu, and J. Gao: ACM Comput. Surv. **54** (2021) 1. https://doi.org/10.1145/3439726
7 R. Alghamdi and K. Alfalqi: Int. J. Adv. Comput. Sci. Appl. **6** (2015) 147. https://thesai.org/Downloads/Volume6No1/Paper_21-A_Survey_of_Topic_Modeling_in_Text_Mining.pdf.
8 R. A. Calvo and S. D'Mello: IEEE Trans. Affect. Comput. **1** (2010) 18. https://doi.org/10.1109/T-AFFC.2010.1
9 C. Fan, Y. Jiang, and A. Mostafavi: J. Manage. Eng. **36** (2020) 3.
10 N. Abdelhady, I. E. Elsemman, and T. H. A. Soliman: J. Big Data **11** (2024) 171. https://doi.org/10.1186/s40537-024-01035-z
11 I. Boutabia, A. Benmachiche, A. A. Betouil, C. Chemam, and M. Boutassetta: SSRN (2026). https://ssrn.com/abstract=6435019
12 H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao: Proc. 2014 IEEE Int. Conf. Data Mining (IEEE, 2014) 899. https://doi.org/10.1109/ICDM.2014.47
13 Q. Li, H. Peng, J. Li, C. Xia, R. Yang, L. Sun, P. S. Yu, and L. He: ACM Trans. Intell. Syst. Technol. **13** (2022) 1. https://doi.org/10.1145/3495162
14 X. Yan, J. Guo, Y. Lan, and X. Cheng: Proc. 22nd Int. Conf. World Wide Web (ACM, 2013) 1445. https://doi.org/10.1145/2488388.2488514
15 Z. Li, R. Li, and G. Jin: IEEE Access **8** (2020) 75073. https://doi.org/10.1109/ACCESS.2020.2986582
16 D. Yu: Remote Sens. **17** (2025) 2041. https://doi.org/10.3390/rs17122041
17 X. Gong, N. Borisov, N. Kiyavash, and N. Schear: Proc. 12th Int. Conf. Privacy Enhancing Technologies (Springer, 2012) 58. https://doi.org/10.1007/978-3-642-31680-7_4
18 Y. Gao, G. Casale, and R. Singhal: ACM Trans. Model. Perform. Eval. Comput. Syst. **9** (2026) 14. https://doi.org/10.1145/3687265
19 Sheetal and A. Kaur: Proc. 2025 Int. Conf. Electronics, AI and Computing (IEEE, 2025) 1. https://doi.org/10.1109/EAIC66483.2025.11101394
20 J. Wang, Y. Xiong, and Z. Yang: Proc. 9th Int. Conf. Computer Science and Artificial Intelligence (ACM, 2025) 140. https://doi.org/10.1145/3788149.3788253
21 P. Jiang: Int. J. Comput. Sci. Inf. Technol. **4** (2024) 280. https://doi.org/10.62051/ijcsit.v4n2.36
22 S. Z. Wu, Y. Wang, L. Shen, F. Hu, and H. Yu: Comput. Mater. Continua **80** (2024) 4111. https://doi.org/10.32604/cmc.2024.054581
23 Y. Ma, H. Peng, and E. Cambria: Proc. AAAI Conf. Artif. Intell. **32** (2018) 1. https://doi.org/10.1609/aaai.v32i1.12048
24 C. Li, Y. Duan, H. Wang, Z. Zhang, A. Sun, and Z. Ma: ACM Trans. Inf. Syst. **36** (2018) 11. https://doi.org/10.1145/3091108
25 X. Cheng, X. Yan, Y. Lan, and J. Guo: IEEE Trans. Knowl. Data Eng. **26** (2014) 2928. https://doi.org/10.1109/TKDE.2014.2313872
26 T. Chen and C. Guestrin: Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining (ACM, 2016) 785. https://doi.org/10.1145/2939672.2939785
27 J. Liu, Z. Zhou, M. Gao, J. Tang, and W. Fan: J. Assoc. Inf. Sci. Technol. **74** (2023) 1026. https://doi.org/10.1002/asi.24800
28 C. Mniszak, H. L. O'Brien, D. Greyson, C. Chabot, and J. Shoveller: Inf. Process. Manag. **57** (2020) 102081. https://doi.org/10.1016/j.ipm.2019.102081
29 S. Yan, J. Wang, and Z. Song: Future Internet **14** (2022) 234. https://doi.org/10.3390/fi14080234
30 M. Chen, S. Mao, and Y. Liu: Mobile Netw. Appl. **19** (2014) 171. https://doi.org/10.1007/s11036-013-0489-0
31 M. Röder, A. Both, and A. Hinneburg: Proc. 8th ACM Int. Conf. Web Search Data Mining (ACM, 2015) 399. https://doi.org/10.1145/2684822.2685324

32 C. W. Arnold, A. Oh, S. Chen, and W. Speier: Comput. Methods Programs Biomed. **124** (2016) 67. https://doi.org/10.1016/j.cmpb.2015.10.014
33 M. Diouf, E. Toe, M. Grichi, H. Nakouri, and F. Jaafar: Comput. Mater. Continua **87** (2026) 147. https://doi.org/10.32604/cmc.2026.077139
34 X. Luo, H. Ding, M. Tang, P. Gandhi, Z. Zhang, and Z. He: Proc. IEEE Int. Conf. Bioinformatics Biomed. (IEEE, 2020) 1077. https://doi.org/10.1109/bibm49941.2020.9313379
35 N. M. Gardazi, A. Daud, M. K. Malik, A. Bukhari, T. Alsahfi, and B. Alshemaimri: Artif. Intell. Rev. **58** (2025) 166. https://doi.org/10.1007/s10462-025-11162-5
36 A. Firouzi, S. Dadkhah, S. A. Maret, and A. A. Ghorbani: Electronics **14** (2025) 4095. https://doi.org/10.3390/electronics14204095
37 S. K. Jawalkar: Int. J. Multidiscip. Res. **4** (2024) 1. https://www.ijfmr.com/papers/2022/5/38175.pdf.
38 D. R. Augustyn, Ł. Wyciślik, and M. Sojka: Appl. Sci. **14** (2024) 646. https://doi.org/10.3390/app14020646
39 K. D. Schubert and D. Barrett: Human Privacy in Virtual and Physical Worlds. Technology, Work, and Globalization, Eds. M. C. Lacity and L. Coon (Palgrave Macmillan, Cham, 2024) Chap. 5. https://doi.org/10.1007/978-3-031-51063-2_5
40 C. Chen, L. Wei, J. Xie, and Y. Shi: ACM Trans. Knowl. Discov. Data **19** (2025) 96. https://doi.org/10.1145/3729234

About the Authors

Zhong-Jie Liu received his B.S. degree from Xinyang Normal University, China, in 2006 and his M.E. degree from Changzhou University, China, in 2010. Since 2021, he has been an associate professor at Changzhou University of Information Technology. His research interests are in AI agents, robot control systems, deep learning, and large model applications. (156247874@qq.com)

Shih-Pang Tseng received his B.S. and M.S. degrees from the Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan, and his Ph.D. degree from the Department of Computer Science and Engineering, National Sun Yat-sen University, Kaohsiung, Taiwan. He is a professor at Changzhou University of Information Technology. His current research interests include deep learning, natural language processing, robotics, and learning technology. (tsengshihpang@ccit.js.cn)