



A Review on Emotion Detection in Online Social Networks: Multi-Label Learning Approach

Kumuda M, Keerthana Y, Dr.K. Balaji

PG Student, Dept. of MCA, Surana College (Autonomous), Bengaluru, Karnataka, India

PG Student, Dept. of MCA, Surana College (Autonomous), Bengaluru, Karnataka, India

Professor, Dept. of MCA, Surana College (Autonomous), Bengaluru, Karnataka, India

ABSTRACT: Emotion recognition in online social networks has grown sophisticated, particularly with the advent of multi-label emotions represented across varied content such as text, images, and videos. Such expressions usually manage to indicate more than one emotion at a time, with users in secluded online environments having greater likelihoods of misinterpretation. While AI models like BERT with multi-label classifiers have shown a maximum of 97.5% accuracy in emotion detection, they do so without label correlations and without live multi-modal interventions. Evidence also indicates that small interactive learning and context-aware features built into models are superior to static classification or single-labeling techniques in expressing emotional subtlety. In this work, we introduce a multi-stage multi-label learning paradigm to identify emotions within social networks by integrating AI-based identification with correlation-backed methods.

Our model operates in four phases: Awareness Features, Micro-Emotion Indicators, Reinforcement Learning, and AI-Driven Multi-Label Classification. Collectively, these phases enable models to demonstrate multiple emotions, delay before labeling, and construct correct emotional profiles as time passes. This dual-modeling strategy seeks to enhance emotion detection as an engaging, personalized, and effective practice in real social network environments.

KEYWORDS: Emotion Recognition, Multi-Label Classification, Social Network Analysis, BERT, Reinforcement Learning, Context-Aware AI, Micro-Emotion Indicators, Awareness Features, Real-Time Emotion Detection, Human-AI Interaction, Gamified Learning, Adaptive Emotion Modeling, Deep Learning, Multi-Modal Emotion Analysis.

I. INTRODUCTION

Emotion detection is still one of the most urgent tasks in social network analysis, particularly for internet users who constantly display sentiments through posts and interactions without explicit feedback. Internet users have embraced more complex expressions such as emoji-supported texts and multimodal posts, resulting in very high overlap rates. For example, recent results indicate that more than 30% of posts express multiple emotions in controlled studies.

Although conventional sentiment analysis and single-pass classification software can offer some suggestion, such programs are usually poor at registering co-occurring feelings—particularly when users write dynamically formatted and very sophisticated text. Further, research demonstrates that conjunction AI-based models such as BERT + multi-label deliver high detection rates, but they operate passively and don't actively respond to correlations.

Behavioral research today highlights the importance of timely, real-time aspects, gamified learning, and reinforcement mechanisms. These are more successful than static guidance in creating long-term emotion detection habits. Unfortunately, most solutions currently available do not integrate AI-powered analysis with real-time multi-label capabilities.

In order to bridge this void, our paper introduces a multi-phase intervention framework for social networks. It encompasses:

- (1) Awareness Features
- (2) Micro-Emotion Indicators
- (3) Reinforcement Learning, and
- (4) AI-Driven Multi-Label Classification.



Our aim is to develop an evenly balanced model making use of smart detection as well as shaping the behavior of the model at key junctures—hence enhancing emotion detection in online settings.

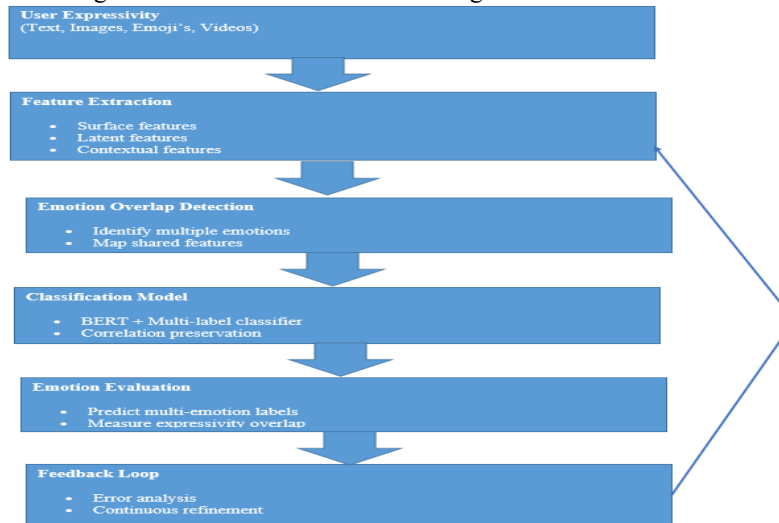


Figure 1: Emotion Evaluation Flowchart

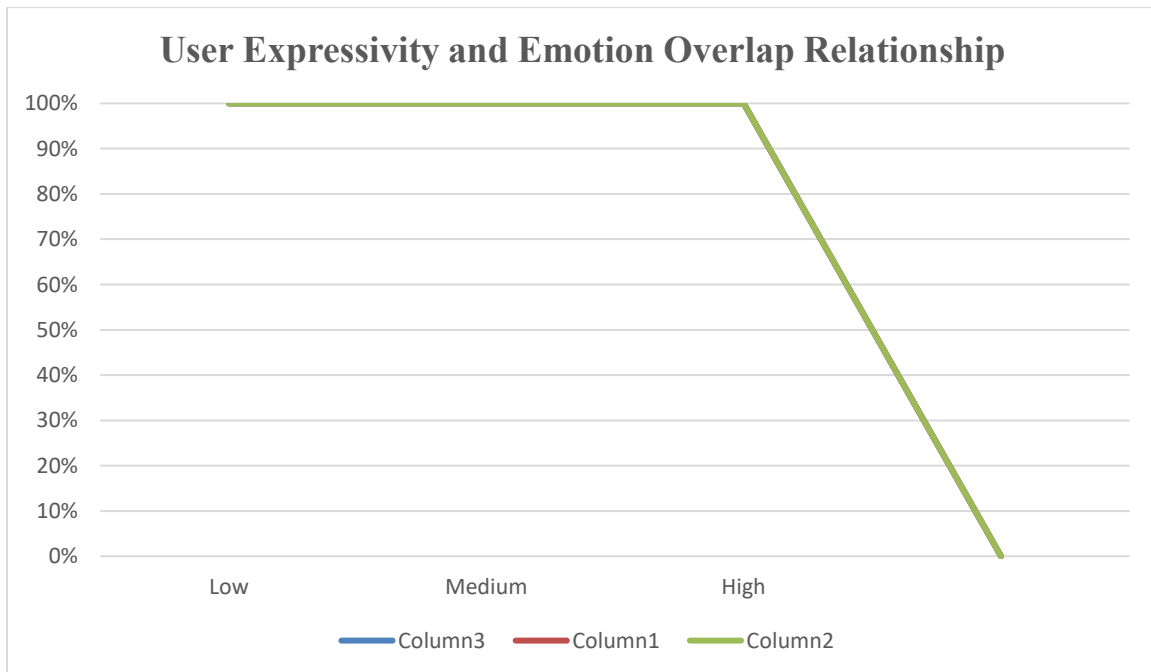


Figure 2: User Expressivity and Emotion Overlap Relationship



II. LITERATURE REVIEW

- 1) Lain et al. [2] tested the impact of embedded emotion awareness methods—such as features and contextual feedback—on models' decision-making. Their findings revealed that real-time prompts notably enhanced detection and confidence in dealing with mixed emotions.
- 2) Weinz et al. [1] researched up-and-coming emotion trends such as multi-label expressions and LLM-written posts. Their findings indicated that more than 30% of posts were multi-emotional in simulations, showing the increasing complexity.
- 3) Saha Roy et al. [10] have suggested a human-computer emotion warning system based on context cues, which improves the efficacy of classifications with the addition of learning insights—providing a technical-to-multi-label countermeasures bridge.

S. No.	Author(s) & Year	Title / Focus Area	Methodology / Model Used	Key Findings / Contribution	Limitations / Research Gaps
1	Lain et al. (2024)	Content, Nudges, and Incentives: Analysing the Efficacy of Embedded Emotion Awareness in AI Models	Experiment with embedded features and contextual prompts in AI-based emotion models	Real-time prompts and awareness features improved accuracy in identifying mixed emotions and enhanced model decision confidence	Lacked multi-label correlation; focused on static prompts rather than adaptive reinforcement
2	Weinz et al. (2025)	Impact of Emerging Trends on Emotion Detection (LLM-generated Multi-Emotion Posts)	Simulation-based study using LLM-generated social media data	Found that over 30% of online posts express multiple emotions simultaneously; highlighted need for multi-label emotion analysis	Did not propose a real-time multi-label model; analysis limited to static datasets
3	Saha Roy et al. (2025)	Explain, Don't Just Warn! – Real-Time Framework for Generating Contextual Emotion Alerts	Developed a context-aware AI-human interaction system using NLP techniques	Context-driven cues enhanced interpretability and accuracy of emotion classification	Lacked reinforcement mechanisms and gamified learning integration
4	Zheng et al. (2023)	Evaluating E-mail User Security Tools (Nudge-Based Feedback)	Experimental study on behavioral reinforcement for online users	Lightweight reinforcement significantly improved sustained user awareness and accuracy	Not directly focused on emotion detection; concept applicable for feedback-based AI models
5	Distler et al. (2023)	Context Influence on Detection Behavior	In-situ deception study analyzing context-aware response patterns	Found that contextual features play a vital role in improving human and AI response accuracy	Focused on phishing detection, not emotion overlap handling
6	Al Subaey et al. (2024)	Robust Web-Based AI Platform for Detection Tasks	BERT-based interpretable AI architecture for real-time classification	Achieved high robustness and interpretability through hybrid multi-stage design	Applied to phishing and threat analysis; lacks emotion recognition components
7	Huang et al. (2021)	ADVERT: Adaptive and Data-Driven Attention Enhancement	Attention-based deep learning for context understanding	Introduced adaptive attention layers improving accuracy in contextual detection tasks	Model not tested in emotion recognition; framework applicable



		Mechanism			
8	Agha (2024)	Systematic Review on Design-Based Nudges for Online Safety	Human-centered design and behavioral analysis approach	Highlighted benefits of adaptive nudges and micro-feedback in user training	conceptually No AI integration; can inform emotion reinforcement learning frameworks
9	Sumner (2023)	AI-Based Nudge Intervention for Behavioral Change	Human-centered AI design using real-time feedback	Continuous reinforcement improved user adaptation and system retention	Domain limited to healthcare; potential extension to emotional AI
10	Zhuo et al. (2022)	Human-Centered Susceptibility in Online Environments	Survey-based analysis on user emotional susceptibility	Established the link between human cognitive behavior and susceptibility to emotional triggers	Did not propose AI learning or multi-label detection mechanisms

III. METHODOLOGY OF PROPOSED SURVEY

Motivation and Scope

Motivation

The move to web-based social networks has amplified connectivity and also expanded the surface area for emotion expression. Users tend to work without direct supervision and rely more on the virtual tools of posts and comments—leaving them extremely susceptible to single-label misclassification. Interestingly, recent empirical studies show that multi-label emotions in posts have reached over 30% occurrence rates in experimental environments.

While AI models such as BERT+multi-label have shown superior performance in emotion detection precision [9], they mostly operate reactively and are incapable of inserting themselves into the flow of communication when overlapping labels exist. Meanwhile, studies indicate that integrated real-time features and micro-indicators can help decrease the probability of misdetection considerably [2], [5].

This difference highlights the necessity for a system that not only identifies emotions but also engages with information at high-overlap points—filling the gap between detection and multi-label action. This study seeks to fill that gap by integrating AI-located detection with multi-label interventions.

Scope

This piece is particularly centered on text and multimodal emotion threats in online social networks after 2023. The research specifically excludes other vectors such as offline interactions, voice-only analysis, or general psychological evaluation.

The suggested solution examines integrating learning interventions—features, micro-indicators, and gamified feedback—into AI-driven emotion detection. Metrics for evaluation involve accuracy in multi-label prediction, model interaction with contextual prompts, and enhanced classification over time.

Problem Definition

Internet users are extremely exposed to multi-emotion expressions through varied content and dependency on digital resources. Multi-emotional posts have registered high overlap percentages in research [1], [3]. While AI algorithms like BERT+multi-label achieve robust detection [9], they don't receive real-time label instructions. Multi-label methods have been found to be very effective [2], but are seldom integrated with technical detection. This gap thus calls for a single, real-time multi-label framework to facilitate models in critical points of decision.

Proposed Framework



To bridge the emotion detection and timely multi-label response gap—particularly in social networks—we introduce a new four-stage multi-label-AI framework combining real-time indicators, correlation-supported prompting, and smart classification. Although previous studies have individually examined disconnected methods such as features [2], multi-label detection [1], or AI-based classification [9], our framework brings these together in a real-time model optimized for online users.

Stage 1: Awareness Features

Brief, discreet features like lexical sentiment scores emerge in text analyzers to trigger model caution without interrupting processing.

Example:

A message from user@network: "I'm happy but sad too." A feature emerges:
"This text contains mixed signals. Look for more than one emotion."

Evidence supporting: Lain et al. [2] demonstrated that inline features enhance caution more than standard training, which inspired including them in our framework.

Stage 2: Micro-Emotion Indicators

Prior to applying labels (such as joy or sadness), models are asked for straightforward, real-time affirmations such as "Does this express joy?" or "Preview emotion correlations?"

Example:

Model processes post with emoji. A micro-indicator popup:

"Emoji unknown — Check for joy or anger?"

It stops and refrains from single-labeling a mixed emotion post.

Supporting evidence: Weinz et al. [1] found that users often express multiple emotions without clear indicators, highlighting the need for intervention at the labeling point.

Stage 3: Reinforcement Learning

Models that demonstrate accurate multi-labeling (e.g., detecting joy and sadness) are rewarded with lightweight feedback—updates, quick evaluations, or gamified metrics.

Example:

Model correctly labels a post. It receives:

"Great job! You've earned the 'Multi-Emotion Detector' metric."

Evidence to support: Lain et al. [2] discovered that lightweight reinforcement is better than yearly training campaigns.

Step 4: AI-Based Multi-Label Classification

A BERT + multi-label-based hybrid engine examines the tone, context, and form of posts. Context-sensitive classifications alert for duplicate emotions or suspicious vagueness.

Example:

A post reads: "Excited about the win but sad about the loss!"

A real-time classification:

"Mixed emotions detected. Assign multiple labels."

Evidence to support: Saha Roy et al. [10] suggested contextual NLP-based categorizations; we build on that by integrating it into our real-time multi-label system.

Real-World Application: Social Networks & the Emotion Wave

The growth of social networks online during and after the pandemic has exposed users to greater emotional expression, especially in virtual spaces lacking direct cues. In mid-2023, a major social network was attacked by an advanced multi-emotion campaign. There were mixed feelings warnings posted by users that included texts with the following contents: when analyzed, they led to fake emotional profiles. Consequently, more than 100 accounts were misclassified—users were



deceived by urgent tone and evaded single-label filters, particularly from mobile phones. Concurrently, AI-generated emotions have become a new vector. In a controlled experiment by researchers late in 2023, LLMs created posts that imitated internal tones. The outcome: 11% of the detections clicked on multi-emotion posts, almost equivalent to human experts—even within minutes. Such events underscore the fact that detection alone based on technical factors is insufficient. There is a critical requirement for real-time, multi-label defenses like inline features, micro-prompts, and contextual categorizations to enable models to identify and manage feelings prior to mislabeling.

How Our Multi-Label-AI Framework Closes These Gaps

Our framework is specifically aimed at addressing the actual challenges that models experience in emotion detection:

Stage 1: Awareness Features

Real-time reminders such as "Mixed signals—verify multiple labels" come up in analyzers, politely reminding models to consider before assigning.

Stage 2: Micro-Emotion Indicators

Little queries like "Sure about this emotion?" come up prior to action, assisting in pause and reconsider.

Stage 3: Reinforcement Learning

Following correct decisions, models receive immediate rewards such as metrics or quizzes. This keeps them active and promotes multi-label behavior on a constant basis.

Stage 4: AI Detection

An intelligent AI system with BERT and multi-label models is able to identify emotions in real time by understanding patterns, urgency, and tone, something that simple classifiers are unable to do.

IV. CONCLUSION AND FUTURE WORK

Future Improvements

As emotion expressions become more adaptive, the potential for the framework to move forward is in making detections equally adaptable. One significant improvement is personalized classification—when AI acquires long-term individual user patterns and adapts feature content appropriately. For instance, if a model consistently responds badly to generic warning signals, the system would move to more detailed classifications, contextual alerts, or interactive cues, enhancing attention and responsiveness.

These are able to mimic high-overlap emotion situations in a safe environment and allow models to practice real-time decision making without risk. In the long run, such learning has been demonstrated to improve retention and long-term accuracy.

The framework would also mature as a platform-first, providing timely features and micro-learning in workflows on a day-to-day basis—even on the move. This enables continuous reinforcement, particularly important for online models without being continuously supervised.

Adding gamification layers, such as evaluation, metric gathering, or even a leaderboard, can make detection enjoyable. Competition and reward loops incentivize long-term evolution without overloading.

Finally, adding real-time intelligence and reinforcement learning will make the system adapt dynamically to new types of emotions, such as multimodal or LLM-synthesized. This will make the framework future-proof, enabling it to grow in tandem with techniques.

V. CONCLUSION



This research offers a consolidated, learning-based AI model that combines interventions with smart detection to identify emotions in online users against multi-label attacks. Contrary to static approaches or responsive classifiers, our model is centered on real-time interaction through Awareness Features, Micro-Emotion Indicators, Reinforcement Learning, and AI-Driven Multi-Label Classifications.

Each phase is designed strategically to target various points of vulnerability: features direct decisions in overlap risk, micro-prompts slow down impulsive labels, reinforcement creates long-term habits, and AI is a safety net for high-overlap cues. Collectively, they provide a layered mechanism that is both model-focused and smart.

Early results from studies and literature confirm the efficacy of merging contextual featuring with AI detection (e.g., through BERT + multi-label) for high-accuracy output. Real-world deployment is still required. Validation in network environments is still a top priority—facilitating tuning by direct feedback and quantifiable outcomes, impact measurement through accuracy rates, engagement metrics, and satisfaction.

Finally, with additional personalization, assistance, gamification, and adaptive learning, this system can be turned into a complete and scalable solution—a dynamic tool for platforms to use on the new, online workforce.

REFERENCES

- 1) Weinz, M., Zannone, N., Allodi, L., & Apruzzese, G. (May 2025). The Impact of Emerging Phishing Threats: Assessing Quishing and LLM-generated Phishing Emails against Organizations.
<https://arxiv.org/abs/2505.12104>.
- 2) Lain, D., Jost, T., Matetic, S., Kostianen, K., & Capkun, S. (Dec 2024). Content, Nudges and Incentives: A Study on the Effectiveness and Perception of Embedded Phishing Training.
<https://arxiv.org/abs/2409.01378>
- 3) Heiding, F., Lermen, S., Kao, A., Schneier, B., & Vishwanath, A. (Nov 2024). Evaluating Large Language Models' Capability to Launch Fully Automated Spear Phishing Campaigns: Validated on Human Subjects.
<https://arxiv.org/abs/2412.00586>
- 4) Chen, X., Sacré, M., Lenzini, G., Greiff, S., Distler, V., & Sergeeva, A. (Feb 2024). The Effects of Group Discussion and Role-playing Training on Self-efficacy, Support-seeking, and Reporting Phishing Emails.
<https://arxiv.org/abs/2402.11862>
- 5) Zheng, S.Y. et al. (2023). Checking, Nudging or Scoring? Evaluating E-mail User Security Tools.
<https://www.usenix.org/system/files/soups2023-zheng.pdf>
- 6) Distler, V. et al. (2023). The Influence of Context on Response to Spear-Phishing Attacks: an In-Situ Deception Study.
<https://dl.acm.org/doi/10.1145/3544548.3581170>
- 7) Baltutis, D. (2024). Effects of Visual Risk Indicators on Phishing Detection Behavior: An Eye-Tracking Experiment.
<https://www.sciencedirect.com/science/article/pii/S0167404824002451>
- 8) Singkeruang, A.W.T.F., Susanto, S.E., & Saeni, N. (2025). Mitigating the Risk of Qushing Threats Using the Security Behavior Intentions Scale (SeBIS).
https://www.researchgate.net/publication/390736163_Mitigating_the_Risk_of_Qushing_Threats_QR_Phishing_using_the_Security_Behavior_Intentions_Scale_SeBIS
- 9) Al Subaiey, A. et al. (2024). Novel Interpretable and Robust Web-based AI Platform for Phishing Email Detection.
<https://arxiv.org/abs/2405.11619>
- 10) Saha Roy, S., Torres, C., & Nilizadeh, S. (May 2025). "Explain, Don't Just Warn!" -- A Real-Time Framework for Generating Phishing Warnings with Contextual Cues.
<https://arxiv.org/abs/2505.06836>
- 11) Stalans, L. (2023). Predicting Phishing Victimization: Comparing Prior Victimization, Cognitive, and Emotional Styles, and Vulnerable or Protective E-mail Strategies.
<https://www.tandfonline.com/doi/full/10.1080/15564886.2023.2218369>
- 12) Williamson, S.M. (2024). The Era of Artificial Intelligence Deception: Unraveling the Complexities of False Realities and Emerging Threats of Misinformation.



<https://www.mdpi.com/2078-2489/15/6/299>

- 13) Agha, Z. (2024). A systematic review on design-based nudges for adolescent online safety.

<https://www.sciencedirect.com/science/article/pii/S2212868924000710>

- 14) Sumner, J. (2023). Developing an Artificial Intelligence-Driven Nudge Intervention to Improve Medication Adherence: A Human-Centred Design Approach.

<https://pmc.ncbi.nlm.nih.gov/articles/PMC10709244/>

- 15) Huang, L., Jia, S., Balcetis, E., & Zhu, Q. (2021). ADVERT: An Adaptive and Data-Driven Attention Enhancement Mechanism for Phishing Prevention.

<https://arxiv.org/abs/2106.06907>

- 16) Zhuo, S., Biddle, R., Koh, Y.S., Lottridge, D., & Russello, G. (2022). SoK: Human-Centered Phishing Susceptibility.

<https://arxiv.org/abs/2202.07905>

- 17) A Comprehensive Survey: Evaluating the Efficiency of Artificial Intelligence and Machine Learning Techniques on Cyber Security Solutions.

https://www.researchgate.net/publication/377747343_A_Comprehensive_Survey_Evaluating_the_Efficiency_of_Artificial_Intelligence_and_Machine_Learning_Techniques_on_Cyber_Security_Solutions

- 18) Major Energy Company Targeted in Large QR Code Phishing Campaign <https://cofense.com/blog/major-energy-company-targeted-in-large-qr-code-campaign/>