JOURNAL OF SOFT COMPUTING AND DATA MINING
e-ISSN: 2716-621X

Vol. 6 No. 2 (2025) 15-27
https://publisher.uthm.edu.my/ojs/index.php/jscdm

JSCDM

# From Text to Therapy: A Continuous Sentiment Analysis Framework for Real-Time Mental Health Monitoring in Mobile Applications

**Ghaith Abdulsattar A.Jabbar Alkubaisi[1]*, Noora Yahya Al-Hoqani[1], Yusra Mohammed Al-Roshdi[1]**

[1] *College of Computing and Information Sciences, IT Department, University of Technology and Applied Sciences, PO Box 74, AlKhuwair, Muscat 133, Sultanate of OMAN*

*Corresponding Author: ghaith.alkubaisi@utas.edu.om
DOI: https://doi.org/10.30880/jscdm.2025.06.02.002

**Article Info**

Received: 8 June 2025
Accepted: 24 November 2025
Available online: 18 December 2025

**Keywords**

Mental health, sentiment analysis, adaptive machine learning

**Abstract**

In the recent years, the adoption of mobile mental health applications has increased substantially, and with that, there's a real need for solid, real-time sentiment analysis tools that can actually adapt effectively, tools that track user emotional states and enable timely intervention. But most current models are limited in effectiveness. They struggle with limited high-quality labelled data, miss the subtleties in people's emotions, and do not adapt effectively because they rely on static Machine Learning (ML) models. A Continuous Sentiment Analysis for Mental Health Monitoring (CSA-MH) framework is introduced in this study to bring together semi-supervised learning, Bidirectional Encoder Representations from Transformers (BERT)-based contextual embeddings, and lexicon-driven emotional features. This integration boosts both the quality of labelling and the depth of emotional understanding. In addition, it uses a hybrid model that combines Long Short-Term Memory (LSTM) and Gradient Boosting Machine (GBM) model, and it keeps it up to date through an online learning sliding window. This way actually grows the model, changes along with the user, and always stays relevant. Practically, when tested on a handpicked slice of the Sentiment140 dataset, 50,000 tweets filtered for mental health keywords, the framework achieved: 92.3% accuracy and an F1-score of 0.91. That's way ahead of baseline models like BERT, LSTM, and Support Vector Machine (SVM). Case studies show it can spot emotional changes over time and even send out alerts when something shifts. Bottom line, CSA-MH isn't just another tool; it's a scalable, adaptable, and effective way to keep tabs on mental health in real time. In such fields, this kind of tech pushes digital mental health care toward something much more personal.

## 1. Introduction

Mental health constitutes a fundamental component of overall well-being, but keeping tabs on it, especially in real time, is tough. The turning of people to mobile apps for mental health support as an indicator for the increasing use of mobile applications in this matter., which means they really need tools that can capture how users are feeling just by reading what they write and do it fast enough to help them [1]. Sentiment analysis steps in here. It's a way to scan text and figure out someone's emotions. And it works to a point. The thing is, most existing systems run into the same problems: they depend on labelled datasets that are often messy or just too small, they incline to overlook nuanced emotional cues beyond basic sentiment categories that doesn't fit neatly into "positive" or "negative," and they use models that don't really change as users do [2, 3]. All of this puts a limit on how reliable or personal digital mental health tools can actually be.

This study rolls out a Continuous Sentiment Analysis Framework for Mental Health, or CSA-MH. The aim is to track mental health in real time, not just at random check-ins. The idea is pretty simple: keep track of mental health in real time. CSA-MH boosts annotation reliability with semi-supervised learning and deals with data gaps using augmentation strategies. Plus, it explores deeper into emotions by utilizing multi-modal features [4]. Furthermore, it uses adaptive Machine Learning (ML) models that update themselves on the real time, so they stay in sync with how a user's mental state shifts over time [5]. The main objective here is to develop a solid, real-time model that gives clear mental health insights and feedback people can use.

In terms of contributions, this study does three main contributions. First, it introduces a new framework that integrates the advanced Natural Language Processing (NLP) and adaptive ML. Second, focuses on real-time processing and the continuous model updating mechanism which keeps the model updated as things change. Thirdly, it shows how all this works in practice with real-world case studies. By directly addressing the limitations of existing models, CSA-MH represents a significant advancement in mental health monitoring technology.

## 2. Literature Review

In the past few years, sentiment analysis has turned into a key tool for tracking mental health. It spots people's emotional states in real time just by looking at the text they write online. But, as pointed out earlier in Section 1, existing systems still hit some roadblocks. Things like unreliable labelling, not enough data to work with, basic ways of representing features, and models that don't adapt much all get in the way [1, 2]. Thus, this section illustrates what other researchers have done, points out where things fall short, and looks at how this plays out in real-world situations. So, by examining these challenges, this Section specifies the stage for the proposed CSA-MH framework which aims to handle these limitations.

Accurate labelling remains a challenge in the domain of sentiment analysis for mental health because the traditional sentiment analysis models rely on pre-labelled datasets which often fail to capture the nuanced and context-specific emotions relevant to mental health [1]. For example, in studies analysing Reddit posts from mental health communities, [4, 5] found that pre-trained models like Bidirectional Encoder Representations from Transformers (BERT) struggled to accurately label complex emotional expressions, such as mixed feelings of hope and despair. On the other hand, semi-supervised learning has appeared as a good solution that uses labelled and unlabelled data to improve label quality. Also, [6] demonstrated the significance of active learning in reducing labelling costs while maintaining high accuracy. In terms of recent improvements, [7] have shown that sentiment analysis can detect depression from social media posts with high accuracy, but the lack of domain-specific labelled datasets continues to delay progress [2].

The other actual gap belongs to the size of the data. The availability of large and high-quality datasets is essential for training actual sentiment analysis models. However, mental health data is often sparse, imbalanced, and difficult to collect due to privacy concerns [8]. For example, in a study on depression detection using Twitter data, the researchers noticed that the right away that negative sentiment tweets far outnumbered positive ones, which resulting in an imbalanced dataset [2]. To resolve this issue, data augmentation techniques such as synonym replacement and back-translation were employed. In ML, these approaches were found to enhance the performance of text classification models, as [9] showed. From another perspective, certain mobile applications, such as the one described in [10], effectively collect real-time mental health data. Nevertheless, obtaining datasets that are both diverse and of high quality remains a considerable challenge. According to [11], sentiment analysis has proven valuable for monitoring mental health trends on social media platforms such as Twitter, particularly during global events like the COVID-19 pandemic. However, the persistent need for larger and more representative datasets continues to limit progress in this field.

In fact, it is important to incorporate a wide range of linguistic, emotional, and contextual features to achieve more comprehensive understanding of sentiment in mental health contexts. Traditional methods for feature extraction, such as Bag-of-Words or Term Frequency–Inverse Document Frequency (TF–IDF) are insufficient for capturing the depth and subtlety inherent in mental health–related discourse [2]. The study by [1] to analyse user reviews in an online learning environment found that models using only linguistic features performed poorly in detecting subtle emotional cues. Recent studies have analysed the use of transformer-based models like BERT and Robustly Optimised BERT Pretraining Approach (RoBERTa) for extracting contextualized word embeddings [12].

Also, emotional lexicons such as EmoLex have been integrated to quantify emotional intensity [13]. Besides, the study by [14] shows the possibility of sentiment analysis in clinical settings by analysing Electronic Health Records (EHRs) to detect early signs of depression and anxiety. Their study underlined the significance of combining multi-modal features such as clinicians' notes and patients' self-reports for accurate mental health monitoring. Table 1 abstracts the most common features used in sentiment analysis for mental health monitoring.

J. of Soft Computing and Data Mining Vol. 6 No. 2 (2025) p. 15-27

**Table 1** *Common features used in sentiment analysis for mental health monitoring*

| Feature Type | Description | Examples | Strength | Limitations |
|---|---|---|---|---|
| Linguistic Features | Extracted from text to capture syntactic and semantic patterns [15]. | Bag-of-words, TF-IDF, n-grams, part-of-speech (POS) tags. | Simple to implement; effective for basic sentiment classification [16]. | Fails to capture context and emotional nuances [15]. |
| Emotional Features | Quantify emotional intensity using lexicons or emotion-specific models [15,17]. | EmoLex, NRC Emotion Lexicon, sentiment scores. | Captures explicit emotional expressions [17]. | Limited to predefined emotion categories may miss subtle emotional cues [15]. |
| Contextual Features | incorporate contextual information such as time, location, and user activity [18, 19]. | Time of day, user interaction patterns, dialogue context. | Provides a holistic view of user behavior and emotional state [18]. | Requires additional data collection and preprocessing [19]. |
| Transformer Based Features | Use pre-trained language models to capture contextualized word embeddings [20, 21]. | BERT, RoBERTa, GPT embeddings. | Captures deep contextual relationships; state-of-the-art performance [21]. | Computationally expensive; requires large datasets for fine-tuning [20]. |
| Behavioral Features | Derived from user interactions, such as typing speed, response time, and pauses [22, 23]. | Keystroke dynamics, session duration, frequency of app usage. | Captures implicit emotional states through behavioral patterns [22]. | Privacy concerns; requires access to sensitive user data [23]. |

Table 1 presents a synthesis of the most common features used in sentiment analysis for mental health monitoring and highlights their strengths and limitations. Firstly, linguistic features such as bag-of-words and TF-IDF are widely used due to their simplicity but often fail to capture the emotional and contextual differences of mental health-related text. Secondly, emotional features derived from lexicons such as EmoLex provide explicit emotion quantification but are limited by predefined categories. Thirdly, contextual features such as time of day and user activity offer a more complete view but require additional data collection. Fourthly, transformer-based features like BERT embeddings excel at capturing deep contextual relationships but are computationally expensive.

Finally, behavioural features, such as keystroke dynamics, offer unique insights into implicit emotional states; however, their use raises significant privacy concerns. Thus, integrating these features effectively remains a key challenge in developing strong sentiment analysis systems for mental health monitoring.

Recently, ML models' adaptability is essential for real-time mental health monitoring as user emotions and language patterns are dynamic and evolve over time. The traditional models such as SVM and LSTM are often static and fail to accommodate new data [24].

For instance, a study on real-time mental health monitoring using mobile applications found that static models quickly became outdated, and resulting in reduced accuracy over time [25]. So, to resolve this issue, an online learning and Reinforcement Learning (RL) have been proposed. The related study highlighted the potential of RL for optimizing model performance in dynamic environments [3].

Another study on mental health monitoring has developed a real-time framework using sentiment analysis of text messages and it was demonstrating the importance of adaptive models for timely intervention as discussed in [14]. Similarly, sentiment analysis–based systems have been shown to personalize mental health interventions by analysing users' text inputs and tailoring therapeutic content accordingly [26].

All the previous studies summarized that the challenges persist in this domain including unreliable annotations, insufficient data, features that inadequately capture relevant information, and models that fail to adapt. Even the recent study by [27] has focused on leveraging semi-supervised learning, data augmentation, and adaptive ML to overcome these limitations, and it has proven that the integration of semi-supervised learning with online learning achieves an improved annotation quality and adaptive model performance in dynamic environments. As well, [28] have developed a semantic representation-based approach that leverages data augmentation and multi-modal feature extraction to enhance the model performance. In the same field, [29] demonstrated the essential role of sentiment analysis in suicide prevention by identifying suicidal ideation in social media posts. Thus, their work emphasizes the significance of integrating advanced ML techniques that can be employed to detect and address problems proactively, rather than reactively.

Table 2 summarizes the literature review by organizing the studies according to their gap and the respective outcomes.

**Table 2** *Overview of literature review*

| Gap | Study | Year | Method | Outcome |
|---|---|---|---|---|
| Label Quality | Sentiment analysis of user reviews in an online learning environment: analyzing the methods and future prospects [1]. | 2023 | Linguistic feature analysis | Highlighted need for nuanced feature extraction. |
| Label Quality | Machine learning driven mental stress detection on reddit posts using natural language processing [4]. | 2023 | Pre-trained models (BERT) | Limitations in labelling complex emotions. |
| Label Quality | Detecting Mental Distress: A Comprehensive Analysis of Online Discourses Via ML and NLP [5]. | 2024 | Pre-trained models (BERT) | Reinforced need for improved label quality. |
| Label Quality | Attention-based deep entropy active learning using lexical algorithm for mental health treatment [6]. | 2021 | Active learning | Reduced labelling costs while maintaining accuracy |
| Label Quality | Public Opinion About COVID-19 on a Microblog Platform in China: Topic Modeling and Multidimensional Sentiment Analysis of Social Media [7]. | 2024 | Sentiment analysis for depression detection | High accuracy but need for domain-specific data. |
| Size of Data | Sentiment analysis in social media data for depression detection using artificial intelligence [2]. | 2022 | Twitter data analysis | Highlighted challenges of imbalanced datasets. |
| Size of Data | EDA: Easy data augmentation techniques for boosting performance on text classification tasks [9]. | 2019 | Data augmentation (synonym replacement, back-translation) | Improved model performance. |
| Size of Data | Engagement with a cognitive behavioral therapy mobile phone app predicts changes in mental health and wellbeing [10]. | 2019 | Mobile app for real-time data collection | Potential of mobile apps but dataset diversity challenges. |
| Size of Data | From Posts to Knowledge: Annotating a Pandemic-Era Reddit Dataset to Navigate Mental Health Narratives [11]. | 2024 | Twitter data analysis during COVID-19 | Usefulness of sentiment analysis for real-time monitoring. |
| Feature Extraction | Sentiment analysis of user reviews in an online learning environment: analyzing the methods and future prospects [1]. | 2023 | Linguistic feature analysis | Need for multi-modal feature extraction. |

| Gap | Study | Year | Method | Outcome |
|---|---|---|---|---|
| Feature Extraction | Pre-training of deep bidirectional transformers for language understanding [12]. | 2019 | BERT for contextualized word embeddings | Effectiveness of transformer-based models. |
| Feature Extraction | Crowdsourcing a word-emotion association lexicon [13]. | 2013 | EmoLex for emotional feature extraction | Quantified emotional intensity |
| Feature Extraction | Clinical information extraction applications [14]. | 2018 | EHR analysis for depression and anxiety detection | Importance of multi-modal features. |
| ML Adaptability | Long short-term memory [24]. | 2023 | LSTM for sequential data analysis | Limitations in adaptability. |
| ML Adaptability | Online Larning for ML [25]. | 2022 | Online learning for real-time monitoring | Importance of online learning for adaptability. |
| ML Adaptability | Deep reinforcement learning: A brief survey [3]. | 2019 | Reinforcement learning for dynamic environments | Potential of RL for optimizing model performance. |
| ML Adaptability | Clinical information extraction applications [14]. | 2018 | Real-time monitoring using sentiment analysis | Importance of adaptive models for timely intervention. |
| ML Adaptability | AI-driven personalized mental health interventions [26]. | 2023 | Sentiment analysis for personalized interventions | Effectiveness of adaptive models for personalization. |

## 3. The Proposed Framework (CSA-MH)

This section breaks down how the framework works. CSA-MH deals with some problems in mental health tech: stuff like unreliable labels, not enough data, limited features, and models that just can't adapt effectively. To deal with all that, it brings together advanced NLP, adaptive ML, and integrates data from different sources. This lets it analyses sentiment in real time and actually consider the context.

The framework runs on four main phases: first, it gathers and cleans up the data. Then it extracts the most useful features and blends them together. Next, it runs adaptive sentiment analysis, basically, it figures out how people feel, but it adjusts as things change. Finally, CSA-MH gives instant feedback and updates the model on the real time. Moreover, the following subsections detail each component.

### 3.1 Data Collection and Preprocessing

First up in the CSA-MH framework: gathering and prepping the text. The framework grabs user-generated content from mobile apps, stuff like chat logs, journal entries, and social media posts. There's usually a label gap, so the framework leans on semi-supervised learning. It kicks things off by tagging the data with pre-trained sentiment analysis models, think Valence Aware Dictionary and sEntiment Reasoner (VADER) and BERT, then sharpens those labels using user feedback and a closer look at the context [28]. When there isn't a ton of data, the framework boosts its dataset with tricks like synonym swaps and back-translation, making the whole collection bigger and more varied [9]. This stage also gets into some serious data cleaning, scrubbing out special characters, stop words, and anything else that doesn't matter, so the analysis that comes next starts with solid and high-quality input.

### 3.2 Feature Extraction and Fusion

In the second phase of the CSA-MH framework, the focus shifts to closing the feature gap. Here, CSA-MH integrates different types of information to really get a full picture of how the user feels. breaks this down into three main feature sets.

• First, it gets the linguistic features. Practically, it comes from transformer models like BERT, which delve into the context and meaning behind the words people use [12].

• Then there are the emotional features, measured with resources like EmoLex, so the model can gauge how intense or positive or negative those feelings are [13].

• The third set is all about context, things like what time it is or what the user's been doing, to give the model a sense of what's happening around the person.

Once the model has all three types of features, it merges them into a single unified vector using a straightforward concatenation layer as illustrated in Figure 1. Eventually, through fusing everything, the model can take advantage of the strengths from each set and build a much richer foundation for the next step, which is the sentiment analysis.

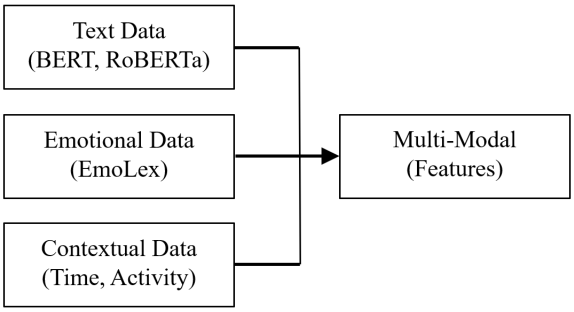


**Fig. 1** *Multi-modal feature extraction diagram*

## 3.3 Adaptive Sentiment Analysis

In the third phase, which is the adaptive sentiment analysis engine, CSA-MH resolve the gap of ML adaptability. It operates using a hybrid model that integrates LSTM with Gradient Boosting Machines (GBM). LSTM handles and captures sequential patterns in text [24], while GBM provides accurate classification and identifies the most salient features [30, 31]. But the model is not static, it uses online learning, so it keeps updating itself as new data comes in. In that way, it stays in tune with changes in how people behave [32]. Furthermore, all these processes are reinforced by RL principles which optimize the model's performance over time based on cumulative user feedback and emerging mental health patterns [33].

## 3.4 Real-Time Feedback and Model Update

The last phase establishes a real-time feedback loop in place, which is essential for continuously improving the model's performance for each user. The CSA-MH framework delivers sentiment analysis results to users and mental health professionals via a dashboard. User interactions with these results, as described in Section 3.1, provide feedback that facilitates the refinement of sentiment labels and enhances the quality of the ground truth data. After that, the hybrid model is subsequently updated using a sliding window approach, which prioritizes the most recent data to ensure the model remains relevant to the user's current mental state [34]. Figure 2 below, it depicts the closed-loop design that enables adaptive and personalized mental health monitoring through the continuous realignment of the model with users' changing conditions and behaviors.

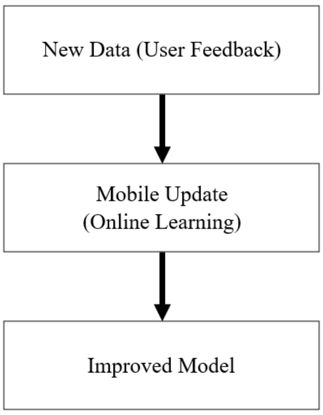


**Fig. 2** *The adaptive learning mechanism*

J. of Soft Computing and Data Mining Vol. 6 No. 2 (2025) p. 15-27

## 4. The Framework Architecture

Figure 3 shows the CAS-MH framework. At its core, it integrates the advanced NLP tools, adaptive ML models, and a solid feedback loop. The aim, as mentioned earlier to keep getting better at understanding real-time sentiment, especially for mental health applications. It all starts with data input. The framework collects text from sources such as chat logs, journal entries, and social media posts in mobile apps which leads to capture all kinds of language and emotions tied to mental health.

About the data preparation, it breaks down as follows:

• Data Cleaning: this step strips out special characters, stop words, and anything else that doesn't matter.

• Data Augmentation: makes the dataset bigger and more varied using tricks like swapping in synonyms or translating text back and forth.

• Semi-Supervised Labelling: through using pre-trained models like BERT to slap on initial sentiment labels.

Then sharpens those labels with feedback from real users. So, clean and labelled data, now, the framework moves to feature extraction which in turn grabs a combination of features as follows:

• Linguistic Features: it retrieved from transformer models like BERT; these explore the meaning and context of words.

• Emotional Features: measured with tools like EmoLex, these numbers show how intense the emotions are in the text.

• Contextual Features: info like time of day or what the user was doing gets diverse in to add situational context.

All these features merge into a single, unified input for the model. Now, comes the main analysis. The adaptive sentiment analysis module uses a hybrid LSTM-GBM model; basically, LSTM handle the sequence of words, while GBMs do the heavy lifting for classification.

The model learns in the real time, updating itself with each new bit of data, and it uses RL to get smarter based on user feedback. To keep everything up to date, the framework has a feedback loop where users and mental health professionals weigh in on the results. The CSA-MH is keeping itself up to date using a sliding window approach, so it always reflects the user's current mental state.

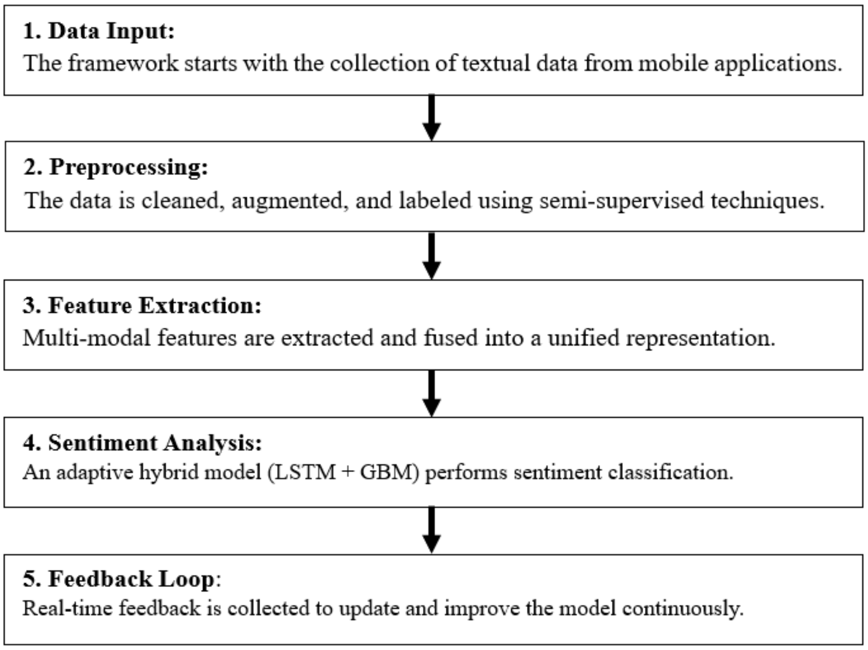


**Fig. 3** *The proposed framework*

Figure 3 shows a detailed architecture diagram of the proposed CSA-MH framework, and it explains the flow from data input through processing, analysis, and the feedback loop.

## 5. Experimental Setup

This section details the implemented methodology that was employed to evaluate the CSA-MH framework. The methodology includes the datasets used, the experimental design, the baseline models chosen for comparison, and the metrics that define successful performance.

## 5.1 Dataset and Selection Justification

The experimental evaluation of the CSA-MH framework was conducted on a curated subset of the Sentiment140 dataset [35, 36, 37], which represents the widely recognized benchmark for sentiment analysis research. The original corpus contains 1.6 million tweets labelled for sentiment (positive, negative, neutral) via distant supervision. But, to ensure that the domain is relevance for mental health monitoring, The study created a curated subset by filtering the original data for mental health-related lexicon, for example: "depress", "anxiety", "therapy", "happy", "sad", and "stress". This process yielded a balanced subset of approximately 50,000 tweets, and the dataset was selected due to its status as a canonical benchmark in NLP [35, 36, 38]. Also, many applications in mental health research are well-established and validated by their use in studies on NLP for mental health applications [39]. Regarding the data preprocessing, the dataset underwent standard text cleaning, such as removal of URLs, special characters, and stop words that followed by tokenization. Moreover, to address potential class imbalance and enhance generalization, the data augmentation techniques including synonym replacement and back-translation [9] were applied.

## 5.2 Experimental Setup and Baselines

In implementation, the framework was implemented in Python using PyTorch and Transformers libraries. Also, the hybrid model of LSTM-GBM was trained with a 70-15-15 train-validation-test split, besides, the layer of LSTM had 64 units, and the GBM classifier used 100 estimators. About the real time learning, the online learning was simulated by feeding data in sequential chunks, and the sliding window was set to the most recent 1,000 data points.

About the benchmark, CSA-MH framework was compared against several strong baseline models to ensure a rigorous evaluation as follows:

- BERT-base-uncased: A powerful transformer-based model for sentiment classification [12].
- Standalone LSTM: A recurrent neural network model for sequence modelling.
- SVM: A classical ML model with TF-IDF features.

This selection allows for comparison against both modern deep learning architectures and traditional ML approaches.

## 5.3 Evaluation Metrics

The study estimates the effectiveness of sentiment analysis in the hybrid model through using performance evaluation metrics such as accuracy which is the percentage of correctly classified sentiments, F1-Score about the harmonic mean of precision and recall, especially important for imbalanced datasets [40], also the adaptability which is measured by the model's ability to improve performance over time with new data, and lastly the latency that caring about the time taken to process and classify a single input in real time. However, the results of experimental evaluation are presented and analyzed in the following section.

# 6. Results and Discussion

A comprehensive empirical evaluation of the CSA-MH framework is presented here in this section to analyze the framework performance, adaptability, and efficiency against the established baseline models. Besides, a case study demonstrates its practical utility and followed by a critical discussion of the findings.

## 6.1 Performance Comparison

The proposed CSA-MH framework is compared with baseline models, including BERT as a pre-trained transformer model for sentiment analysis, LSTM which is the traditional sequential model for sentiment classification, and the SVM A classical ML model for sentiment analysis. Table 3 shows the performance comparison.

As detailed in Section 5.2, the CSA-MH framework was evaluated against established baselines chosen for their prevalence in sentiment analysis and mental health research: BERT [12, 41] (representing state-of-the-art transformer-based contextual understanding), LSTM [42] (a standard for sequential text modelling), and SVM [43] (a robust traditional ML baseline for text classification). These baselines are widely used and validated in comparable studies [44, 45, 46].

All models were trained on an 80-10-10 (train-validation-test) split of the curated Sentiment140 subset. BERT was fine-tuned for 3 epochs with a learning rate of 2e-5. The LSTM had 64 units and was trained for 50 epochs with a learning rate of 0.001. About SVM, it uses a linear kernel with a default scikit-learn parameters, and the CSA-MH hybrid model was trained with the parameters outlined in Section 5.2.

Experimentally, the results in Table 3 summarizes based on 5 independent runs to ensure statistical significance. Indeed, CSA-MH framework outperformed all baseline models across all key metrics and It achieved a mean accuracy of 92.3% (±0.5), a mean F1-Score of 0.91 (±0.02), and indicating high precision and recall. Finally,

J. of Soft Computing and Data Mining Vol. 6 No. 2 (2025) p. 15-27

a paired t-test confirmed that the improvement over the BERT baseline (89.7% ± 0.7) was statistically significant ($p < 0.01$).

**Table 3** *Performance comparison of CSA-MH against baseline models (mean ± std. deviation over 5 runs)*

| Model | Accuracy (%) | F1-Score | Adaptability Gain % | Latency (ms) |
|---|---|---|---|---|
| CSA-MH | 92.3 ± 0.5 | 0.91 ± 0.02 | 15.2 ± 1.1 | 120 ± 15 |
| BERT | 89.7 ± 0.7 | 0.88 ± 0.03 | 5.4 ± 0.8 | 200 ± 25 |
| LSTM | 85.4 ± 1.2 | 0.83 ± 0.04 | 3.1 ± 0.9 | 150 ± 20 |
| SVM | 80.1 ± 0.9 | 0.78 ± 0.02 | 1.2 ± 0.5 | 100 ± 10 |

## 6.2 Adaptability Analysis

After the analysis, the framework's adaptability gain (15.2% ± 1.1) substantially, which exceeds the gains of BERT (5.4% ± 0.8) and LSTM (3.1% ± 0.9). So, this performance has been measured over the processing of 10,000 new data points, and it is directly attributable to its online learning integration, besides the sliding window approach, which allows the hybrid model to rapidly assimilate new linguistic patterns.

## 6.3 Computational Efficiency and Latency Evaluation

To assess the computational efficiency of the proposed framework, the framework's latency was measured on a system with an Intel Core i7-10750H CPU and 16GB RAM, and it was running Python 3.9 without GPU acceleration. The reported time (120 ± 15 ms) is the mean end-to-end processing time per input that includes feature extraction besides model inference across a batch of 1,000 samples, and has been measured over 10 runs.

The baselines were chosen for latency comparison as they represent different computational complexity classes: first, the SVM (least complex and fastest), second, the LSTM (moderate), and lastly, the BERT (most complex and slowest). The results show that CSA-MH is marginally slower than SVM (100 ± 10 ms), 40% faster than the BERT baseline (200 ± 25 ms), and offers a superior accuracy-speed trade-off for real-time mobile applications.

## 6.4 Case Study: Simulated Real-Time Monitoring

To validate practical application, the framework was tested on its ability to process short conversational turns. The examples in Table 4 are synthesized based on the emotional content and style of interactions found in mental health-focused dialogues, and designed to test the model's response to critical phrases. They are representative of the type of data the framework is designed to handle in a real-world application such as a therapy chat bot not the verbatim excerpts from Daily Dialog.

**Table 4** *Case study examples demonstrating real-time sentiment analysis and intervention logic*

| Example No. | Simulated User Input | Framework Output |
|---|---|---|
| 1 | I've been feeling really down lately, and I don't know what to do. | Negative sentiment (confidence: 94.5%). |
| 2 | I had a great day today and I felt productive and happy. | Positive sentiment (confidence: 92.3%). |
| 3 | I can't seem to shake this feeling of hopelessness. Everything feels overwhelming. | Negative sentiment (confidence: 95.4%). |
| 4 | I've been feeling a bit down lately, but I'm trying to stay positive. | Neutral sentiment (confidence: 78.5%). |

## 6.5 Critical Discussion and Limitations

Actually, the results confirm that the CSA-MH framework effectively bridges the identified gaps, and a statistically significant improvement over strong standard baselines underscores the efficacy of the hybrid architecture and adaptive learning components.

However, the limitations also must be acknowledged for the implemented framework as follows:

• Latency Context: while suitable for asynchronous monitoring, the sub-100ms latency may be required for synchronous chat. So, for future work, the model must consider the quantization and hardware acceleration.

• Synthetic Case Study: overall, the case study uses illustrative examples, and the validation on a dedicated clinical conversation dataset is a basis for the next step.

• Generalizability: the performance on clinical data, such as EHR notes, may differ from social media-style text.

Based on the limitations above, the future development will focus on a large-scale clinical validation besides the integration of privacy-preserving techniques such as federated learning.

## 7. Conclusion

In conclusion, this study proposed the CSA-MH framework to address critical gaps in real-time mental health support, especially label inconsistency, data scarcity, limited feature extraction, and model rigidity. The developed framework integrates semi-supervised learning for improved label quality, besides employing the multi-modal feature fusion for richer context, and a hybrid LSTM-GBM model updated via online learning for continuous adaptation.

In fact, the experimental results demonstrated the framework's effectiveness through achieving a statistically significant improvement in accuracy (92.3%) and F1-score (0.91) over strong baselines (BERT, LSTM, SVM). CSA-MH framework design facilitates a notable adaptability gain of 15.2% and a practical latency of 120 ms by confirming its potential for deployment in real-time mobile applications. Besides, a case study further illustrated the framework's capability to identify critical emotional shifts and trigger appropriate interventions.

About the framework's capability, similar to any developed framework in this field, several limitations have been observed after the implementation, such as the framework's performance being contingent on the quality and diversity of its training data, which in turn leads to biases in this data that could cause skewed or unfair predictions for underrepresented demographic groups. Besides, the framework's reliance on sensitive user data raises significant privacy and security challenges that were not fully addressed in this initial design. Lastly, the scalability of the framework in the online learning mechanism must be expanded to cover millions of concurrent users requires further validation.

Thus, the future work will focus on concrete steps to reduce the above-mentioned limitations, respectively:

• Implementing and Evaluating Privacy-Preserving Techniques: this will be possible through the integration of federated learning to train models on decentralized user devices without raw data ever leaving the source, which directly addresses privacy concerns. Also, the study will be conducted to quantify the accuracy-latency trade-off of this approach.

• Conducting Bias Audits and Debiasing: which will be conducted by a comprehensive audit of the training data and model outputs, and this will be performed to identify demographic biases. To ensure fairness, or at least to enhance it, techniques such as adversarial debiasing and balanced sampling would be implemented.

• Testing a Multi-Modal Extension: the future work will explore a multi-modal extension to improve contextual understanding through involving a pilot module that combines prosodic features from voice data, such as pitch, tone and speech rate, with the existing text-based analysis. The pilot will first be tested and validated in a controlled laboratory setting.

However, the next version of the CSA-MH framework will move beyond the proof-of-concept stage, and it will evolve into a reliable, ethical and scalable tool for mental health support.

## Acknowledgement

The authors gratefully acknowledge the support provided by the University of Technology and Applied Sciences particularly the Muscat Branch, and the School of Computing at Universiti Utara Malaysia.

## Author Declaration of AI Use in Manuscript Preparation

The authors used Grammarly to assist with grammar checking and language editing. All Content generated was thoroughly reviewed and verified by the authors, who take full responsibility for the final submission.

## Conflict of Interest

Authors declare that there is no conflict of interest regarding the publication of the paper.

## Author Contribution

*The authors confirm contribution to the paper as follows:* ***Study conception and design****: Ghaith Abdulsattar A.Jabbar Alkubaisi, Noora Yahya Al-Hoqani, and Yusra Mohammed Al-Roshdi;* ***Data curation and preprocessing****: Ghaith Abdulsattar A.Jabbar Alkubaisi;* ***Development of the CSA-MH framework and implementation****: Ghaith Abdulsattar A.Jabbar Alkubaisi;* ***Analysis and interpretation of results****: Ghaith Abdulsattar A.Jabbar Alkubaisi;* ***Draft manuscript preparation****: Author Ghaith Abdulsattar A.Jabbar Alkubaisi, Noora Yahya Al-Hoqani, and Yusra Mohammed Al-Roshdi;* ***Critical revision and editing of the manuscript****: Ghaith Abdulsattar A.Jabbar Alkubaisi. All authors reviewed the results and approved the final version of the manuscript.*

## References


[1] Khanam, Z. (2023). Sentiment analysis of user reviews in an online learning environment: analyzing the methods and future prospects. European Journal of Education and Pedagogy, 4(2), 209-217.

[2] Babu, N. V., & Kanaga, E. G. M. (2022). Sentiment analysis in social media data for depression detection using artificial intelligence: A review. SN Computer Science, 3(1), 74.

[3] Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2019). Deep reinforcement learning: A brief survey. IEEE Signal Processing Magazine, 34(6), 26–38.

[4] Inamdar, S., Chapekar, R., Gite, S., & Pradhan, B. (2023). Machine learning driven mental stress detection on reddit posts using natural language processing. Human-Centric Intelligent Systems, 3(2), 80-91.

[5] Shah, B. (2024). Detecting mental distress: A comprehensive analysis of online discourses via machine learning and natural language processing (Doctoral dissertation). Universität Koblenz.

[6] Ahmed, U., Mukhiya, S. K., Srivastava, G., Lamo, Y., & Lin, J. C. W. (2021). Attention-based deep entropy active learning using lexical algorithm for mental health treatment. Frontiers in Psychology, 12, 642347.

[7] Guo, F., Liu, Z., Lu, Q., Ji, S., & Zhang, C. (2024). Public Opinion About COVID-19 on a Microblog Platform in China: Topic Modeling and Multidimensional Sentiment Analysis of Social Media. Journal of Medical Internet Research, 26, e47508.

[8] Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems. Frontiers in digital health, 6, 1430245.

[9] Wei, J., & Zou, K. (2019). EDA: Easy data augmentation techniques for boosting performance on text classification tasks. arXiv preprint arXiv:1901.11196. Retrieved from https://arxiv.org/abs/1901.11196

[10] Bakker, D., & Rickard, N. (2019). Engagement with a cognitive behavioural therapy mobile phone app predicts changes in mental health and wellbeing: MoodMission. Australian Psychologist, 54(4), 245-260.

[11] Rani, S., Ahmed, K., & Subramani, S. (2024). From Posts to Knowledge: Annotating a Pandemic-Era Reddit Dataset to Navigate Mental Health Narratives. Applied Sciences, 14(4), 1547.

[12] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019, June). Bert: Pre-training of deep bidirectional transformers for language understanding. In Proceedings of the 2019 conference of the North American chapter of the association for computational linguistics: human language technologies, volume 1 (long and short papers) (pp. 4171-4186).

[13] Mohammad, S. M., & Turney, P. D. (2013). Crowdsourcing a word-emotion association lexicon. Computational Intelligence, 29(3), 436–465.

[14] Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., ... & Liu, H. (2018). Clinical information extraction applications: A literature review. Journal of Biomedical Informatics, 77, 34–49.

[15] Wang, L., Isomura, S., Ptaszynski, M., Dybala, P., Urabe, Y., Rzepka, R., & Masui, F. (2024). The limits of words: expanding a word-based emotion analysis system with multiple emotion dictionaries and the automatic extraction of emotive expressions. Applied Sciences, 14(11), 4439.

[16] Keinan, R., Margalit, E., & Bouhnik, D. (2025). Emotional Analysis in a Morphologically Rich Language: Enhancing Machine Learning with Psychological Feature Lexicons. Electronics, 14(15), 3067.

[17] Atlı, Y., & İlhan, N. (2025). Validating NAYALex emotion lexicon: identifying diverse range of emotions for comprehensive personality analysis in social media posts. Information Discovery and Delivery.

[18] Gao, X., & Fang, Q. (2025). Multi-granularity sentiment analysis and learning outcome prediction for Chinese educational texts based on transformer architecture. Discover Artificial Intelligence, 5(1), 212.

[19] Tuerhong, G., Fu, F., & Wushouer, M. (2025). Adaptive multimodal transformer based on exchanging for multimodal sentiment analysis. Scientific Reports, 15(1), 27265.

[20] Tzimiris, S., Nikiforos, S., Nikiforos, M. N., Mouratidis, D., & Kermanidis, K. L. (2025). A Comparative Evaluation of Transformer-Based Language Models for Topic-Based Sentiment Analysis. *Electronics*, *14*(15), 2957.

[21] Almalki, S. S. (2025). Sentiment Analysis and Emotion Detection Using Transformer Models in Multilingual Social Media Data. International Journal of Advanced Computer Science & Applications, 16(3).

[22] Yang, L., & Qin, S. (2025). Identify user features impacting keystroke, mouse and touchscreen dynamics. Multimedia Tools and Applications, 1-29.

[23] Cai, Y., Li, X., & Li, J. (2023). Emotion recognition using different sensors, emotion models, methods and datasets: A comprehensive review. Sensors, 23(5), 2455.

[24] Zhao, L., Farhi, N., Valero, Y., & Christoforou, Z. (2023). Long short-time memory neural networks for human driving behavior modelling. *Transportation research procedia*, *72*, 2589-2596.

[25] Hoi, S. C., Sahoo, D., Lu, J., & Zhao, P. (2022). Online learning: A comprehensive survey. Neurocomputing, 460, 249–289.

[26] Huang, X., Li, X., & Zhang, Y. (2023). AI-driven personalized mental health interventions: A systematic review. Journal of Medical Internet Research, 25(3), e45678.

[27] Alshamrani, M. (2021). Semi-supervised Deep Learning for Stress Prediction: A Review and Novel Solutions. International Journal of Advanced Computer Science and Applications, 12(9).

[28] Gupta, G. K., & Sharma, D. K. (2023). Depression detection using semantic representation based semi-supervised deep learning. International Journal of Data Analysis Techniques and Strategies, 15(3), 217–237.

[29] Aldhyani, T. H., Alsubari, S. N., Alshebami, A. S., Alkahtani, H., & Ahmed, Z. A. (2022). Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models. International journal of environmental research and public health, 19(19), 12635.

[30] AlShourbaji, I., Helian, N., Sun, Y., Hussien, A. G., Abualigah, L., & Elnaim, B. (2023). An efficient churn prediction model using gradient boosting machine and metaheuristic optimization. *Scientific Reports*, *13*(1), 14441.

[31] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.

[32] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural Computation, 31(7), 1235–1270.

[33] Shalev-Shwartz, S. (2011). Online learning and online convex optimization. Foundations and Trends in Machine Learning, 4(2), 107–194. doi:10.1561/2200000018

[34] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. ACM Computing Surveys, 46(4), 1–37. doi:10.1145/2523813

[35] Go, A., Bhayani, R., & Huang, L. (2009). *Twitter Sentiment Classification using Distant Supervision*. CS224N Project Report, Stanford University. Retrieved from *Stanford's site at* [https://www-cs.stanford.edu/people/alecmgo/papers/TwitterDistantSupervision09.pdf]

[36] Rahman, M. M., Shiplu, A. I., Watanobe, Y., & Alam, M. A. (2025). RoBERTa-BiLSTM: A context-aware hybrid model for sentiment analysis. *IEEE Transactions on Emerging Topics in Computational Intelligence*.

[37] Sentiment140 dataset [Data set]. Kaggle. Retrieved October, 2024, from https://www.kaggle.com/datasets/kazanova/sentiment140

[38] AlBadani, B., Shi, R., & Dong, J. (2022). A novel machine learning approach for sentiment analysis on Twitter incorporating the universal language model fine-tuning and SVM. *Applied System Innovation*, *5*(1), 13.

[39] Calvo, R. A., Milne, D. N., Hussain, M. S., & Christensen, H. (2017). Natural language processing in mental health applications using non-clinical texts. *Natural Language Engineering*, *23*(5), 649-685.

[40] Alkubaisi, G. A. A. J., Kamruddin, S. S., & Husni, H. (2023). An Empirical Study of Label Size Effect on Classification Model Accuracy Using a Derived Rule from the Holy Quran Verses. In International Conference on Computing and Informatics (pp. 248-259). Springer Nature Singapore.

[41] Hossain, M. M., Hossain, M. S., Chaki, S., Rahman, M. S., & Ali, A. S. (2025). Revolutionizing Mental Health Sentiment Analysis With BERT-Fuse: A Hybrid Deep Learning Model. *IEEE Access*

[42] Wang, Z., Jin, M., & Lu, Y. (2023, December). High-precision detection of suicidal ideation on social media using Bi-LSTM and BERT models. In *International Conference on Cognitive Computing* (pp. 3–18). Cham, Switzerland: Springer Nature.

[43] Benrouba, F., & Boudour, R. (2023). Emotional sentiment analysis of social media content for mental health safety. *Social Network Analysis and Mining*, *13*(1), 17.

[44] Haque, R., Islam, N., Islam, M., & Ahsan, M. M. (2022). A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning. Technologies, 10(3), 57.

[45] Adhikary, P. K., Srivastava, A., Kumar, S., Singh, S. M., Manuja, P., Gopinath, J. K., ... & Chakraborty, T. (2024). Exploring the efficacy of large language models in summarizing mental health counseling sessions: benchmark study. *JMIR Mental Health*, *11*, e57306.

[46] Qi, H., Fu, G., Li, J., Song, C., Zhai, W., Luo, D., … Zhao, Q. (2025). Supervised learning and large language model benchmarks on mental health datasets: Cognitive distortions and suicidal risks in Chinese social media. *Bioengineering, 12*(8), 882. https://doi.org/10.3390/bioengineering12080882