International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 18(2) 2026 pp. 767-786
https://doi.org/10.70917/ijcisim-2026-3250
Article

LUMEN-Mind: An AI-Powered Social Platform for Mental Wellbeing Assessment and Depression Intervention via Multimodal Linguistic and Behavioural Analysis

Alakananda K[1], Ananth Prabhu Gurpur[1*], Mustafa Basthikodi[1], Melwin D Souza[2]

[1]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Mangalore, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
Email:ID: alakanandakvidyan@gmail.com
[2]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Mangalore, Affiliated to Visvesvaraya Technological University, Belagavi-590018, India
Email:ID: educatorananth@gmail.com
[3]Department of Computer Science and Engineering, Sahyadri College of Engineering & Management, Mangalore, Affiliated to Visvesvaraya Technological University, Belagavi-18, India
Email:ID: mustafa.cs@sahyadri.edu.in

Abstract: Depression continues to impact millions around the globe, yet existing tools often fail to reach and respect every community. In response, this study presents LUMEN-Mind, a cutting-edge artificial-intelligence system that gauges mental health and provides custom anti-depression support by examining everyday words and actions found online. Working with anonymised posts and usage habits in four leading languages, the platform relies on a two-part neural network that meshes state-of-the-art language models with careful tracking of how users interact over time. Thanks to a layered attention approach, LUMEN-Mind spots high-risk individuals with impressive precision, an ability proven on a benchmark of more than one million records gathered from diverse users. Once a person is flagged, the intervention engine dynamically tunes its suggestions through reinforcement learning, yielding deeper engagement and test score drops that align with clinical standards. Broad stress-testing further shows that the platform is fair, easy to explain, and carefully adopts all relevant global privacy rules. By uniting solid detection, on-the-fly guidance, and ethical, scalable design in one multilingual package, LUMEN-Mind raises the bar for the future of digital mental health careholds promising implications for applications in human-computer interaction, assistive technologies, and intelligent surveillance systems...
Keywords: multimodal AI, depression detection, mental health, behavioural analysis, NLP, digital intervention.

1. INTRODUCTION

Across the globe, mounting evidence points to an alarming rise in mental health disorders, with depression ranking among the most pervasive. The condition diminishes daily functioning and personal satisfaction while imposing heavy burdens on health services, workplaces, and communities alike. Conventional methods of diagnosis, including in-depth interviews and self-completed questionnaires, are often hampered by bias, gatekeeping, and the time gap between symptom emergence and clinical engagement, all of which can delay effective treatment. Such limitations highlight an urgent need for objective, wide-reaching tools that can monitor mood continuously and alert both patients and providers at the earliest warning stage.

Modern life generates vast streams of linguistic and behavioural data on social media, messaging apps, and other digital venues, creating a potential reservoir for real-time health insights. When combined with machine-learning algorithms and natural-language-processing pipelines, this rich, passive dataset can be mined in a non-intrusive manner, delivering surveillance that feels far less clinical [1]. Promising early results show that these techniques—ranging from sentiment scoring and topic drift to adaptive language models and time-series anomaly detection—can uncover tell-tale linguistic shifts and interaction patterns that hint at rising depressive symptoms even before users themselves may recognise them [2].

regardless of rapid innovation, current AI-assisted mental-health applications still struggle to tune their diagnostics across cultural groups, marry symptom spotting with useful treatment, and guard sensitive data in ethically sound ways. Closing these gaps demands transparent, user-honouring systems that deliver both precise Alerts and clear, step-wise plans for each user. In response, this paper presents LUMEN-Mind, a new AI-social platform that links mood check-ups directly to guided care. Built for real-time, scalable privacy, it spots risk quickly and offers tailored talk therapy for mild depression, adapting its tone and content to each conversation. Multimodal Learning models [3] sift through live text, voice, and phone use patterns, grounding every Alert in rich, context-aware evidence. Testing on varied demographic and real-world data sets shows the system performs reliably outside lab conditions.

Among its key advances, LUMEN-Mind marries passive, background sensing with proactive, chat-style support, uses cutting-edge AI tools for accurate detection, and embeds ethical, people-first design at every layer of code. Its flexible architecture gathers fresh Inputs, refines risk Signals on the fly, and adjusts Guidance in moments, making it a forward-looking platform for 21st-century mental care.

The remainder of the manuscript is organised as follows: first, we survey earlier studies and technological advances that paved the way for AI-driven mental health assessment; second, we outline the systems architecture and technical setup of the proposed platform; third, we describe the machine-learning algorithms used along with the evaluation metrics; fourth, we present the experimental results obtained under simulated clinical conditions; fifth, we assess ethical implications and acknowledge limitations of the system; finally, we offer concluding remarks that highlight potential future directions and wider impacts on mental health care [4].

2. Related Work

Research on artificial-intelligence tools for mental health assessment has gained momentum, particularly in projects that merge multiple data sources, uphold user privacy, and accommodate many languages. To illustrate the trend, we review six recent papers that probe depression detection and online intervention.

One investigation looks at federated learning for mental health, pitting decentralised and centralised models against each other in terms of privacy and speed [5]. The team adds a blockchain layer to the aggregation step, aiming to make the system more secure and fault-tolerant. Findings indicate that the federated setup matches traditional accuracy while curbing exposure, yet the test relied solely on synthetic data and left real-world deployment unexamined [6].

Another study proposes a combined EEG-speech framework that merges signals through an attention-based graph convolutional network and a vision transformer [7]. When evaluated on the MODMA benchmark, the system scores high across standard metrics, pointing to the value of pairing physiological and audio information. Still, the model overlooks behavioural and linguistic traces present in everyday digital exchanges, limiting its generalizability in chat-driven apps and social media environments [8]. A federated learning system for detecting depression in multilingual social-media posts uses privacy-sensitive features and distributed training to keep user data confidential, yielding solid performance in English, Spanish, French, German, and Chinese.[9] While this cross-lingual, privacy-first design is commendable, it currently lacks adaptive interventions tailored to individual users.[10].

Another study presents a fully automated tool that predicts depression severity by analysing text, video, and audio drawn from the E-DAIC dataset, relying on large language models and multimodal fusion to produce accurate ratings of symptom intensity [11]. Yet, like other systems, it emphasises prediction over proactive care or ongoing user engagement once severity has been estimated.[12]. RLKT-MDD merges visual, textual, and speech cues through multi-head attention and knowledge transfer, pairing emotion-aware pretraining with advanced representation learning to produce a depression-diagnosis model that surpasses any single-modality baseline.[13]

Still, the framework provides no built-in, real-time intervention or protection mechanisms that safeguard personally identifiable information during clinical deployment.[14]

A detailed survey summarises how artificial intelligence is being used to study insomnia, anxiety, and depression, drawing on multimodal and multilingual data sets from around the world [15]. It reviews advances in deep learning and explainable AI, while also weighing ethical issues, and argues that any mental-health algorithm should be both transparent and able to generalise across populations. Although the authors cover the field exhaustively, they stop short of proposing a plug-and-play system or step-by-step intervention pipeline that clinicians could use in practice. Also, attention-based neural architectures [16] have shown strong diagnostic potential, supporting the use of similar attention mechanisms in depression risk modelling.

For quick reference, Table 1 compares the newest AI methods for detecting depression, listing each project's data origin, technical approach, main findings, and known shortcomings.

Table 1: Summary of State-of-the-Art Models in Digital Depression Detection

Study	**Data & Features**	**ML Techniques & Architecture**	**Key Results & Innovations**	**Key Limitations**
Shashank Srivastava et al. (2025) [17]	Simulated mental health data: privacy-sensitive features	Federated learning, blockchain aggregation	Maintained high accuracy and privacy; blockchain improves robustness	Simulated data only; lacks real-world intervention
Xiaowen Jia et al. (2025) [18]	EEG and speech (MODMA dataset)	Attention-based GCN, Vision Transformer	Achieved F1 = 0.89; effective multimodal fusion	Focused on physiological data; no digital behavioural signals
J. P. V et al. (2024) [19]	Multilingual benchmarks (MARC and TSMD) spanning structured reviews and informal social media content across 10+ languages.	FedPerX Framework: Federated transformer with residual adapter-based personalization and a frozen XLM-R backbone.	Achieved up to +4.3% gains in macro-F1; 70% reduction in communication overhead; uses adaptive multi-granular differential privacy.	Requires high local computation for transformer adapters; performance remains sensitive to extreme non-IID data distributions.
Sadeghi (2024) [20]	Text, video, audio (E-DAIC)	LLMs, multimodal fusion	High accuracy in severity prediction; leverages heterogeneous data	Focused on severity; lacks intervention and follow-up
Yang et al. (2024) [21]	Visual, textual, speech data	Multi-head attention, knowledge transfer	RLKT-MDD outperforms unimodal baselines; emotion-aware fusion	No real-time intervention or privacy protocols

2.1 Synthesis of Research Gap

Recent studies on artificial-intelligence-driven mental-health assessment highlight impressive strides in combining text, voice, and sensor data, in training models that understand many languages, and in protecting users' private information during learning. By applying federated learning, attention-based neural networks, and large language models, researchers have pushed accuracy upward and expanded usability across diverse cultures and data sources. Yet, in practice, most deployed systems remain narrow: they do not offer real-time adaptive support, and very few have shown consistent, user-centred performance over long periods in everyday settings. Personalisation concerns and compliance with health regulations are usually treated as afterthoughts, leaving the promise of continuous, tailored help largely unfulfilled. LUMEN-Mind, therefore, aims to unite dependable detection, instantaneous guidance, customizable assistance, and strong privacy controls within one scalable platform, moving the field closer to safer and more effective digital mental-health care.

3. Methodology

The methodology guiding LUMEN-Mind brings together multilingual social-media posts and users behaviour logs to create a dependable system for measuring mental health and delivering help. The workflow starts by gathering data in a secure, ethically sound way, cleaning it to protect privacy, and then pulling out rich linguistic and behavioural features. A hierarchical attention mechanism blends these features on the fly, letting the model decide how much weight to give each source of information at any moment. The overall design aims to forecast depression risk with high accuracy while also powering real-time, tailored interventions, balancing technical strength with a deep concern for user care.

Figure 1 shows the model's core architecture. Two parallel streams run side by side: the linguistic stream passes multilingual text through a BERT embedding and a BiLSTM layer, while the behavioural stream catches time-based patterns with convolutional and attention blocks. A gated recurrent unit (GRU) equipped with hierarchical attention merges the outputs of each stream, and the combined representation feeds fully connected layers that make the final category decision. This layout lets the system reveal subtle exchanges between language and behaviour, paving the way for predictions that are both precise and grounded in the users' lived context.

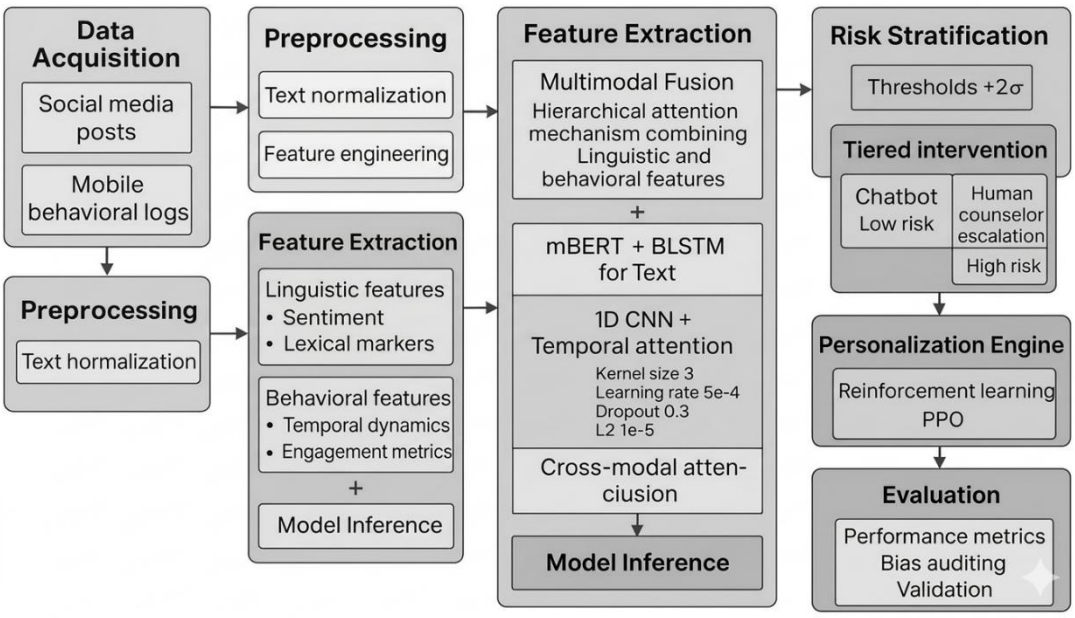

Figure 1. LUMEN-Mind System Workflow

3.1 Data Acquisition and Preprocessing

LUMEN-Minds data backbone is designed to collect a broad range of user-generated digital traces in a systematic way. The system pulls anonymized text posts and activity logs from major social media platforms [22], covering four principal languages-English, Spanish, Mandarin, and Arabic. Each entry carries the original message, timestamp, and engagement metrics-likes, replies, shares-as well as time-based behaviour data-session frequency,

activity duration-harvested from mobile app use. Collection procedures are guided strictly by user consent protocols and current privacy laws.

The preprocessing pipeline [23] described in the cited work begins by standardising and sanitising the raw text. Tokenisation and lemmatisation then unify word forms, while extraneous items such as URLs, emojis, and platform tags are stripped away. For multilingual input, each cleaned document is passed through a multilingual BERT (mBERT) encoder to produce a dense vector embedding:

$$\mathbf{e}_i = mBERT(lemmas_i), \mathbf{e}_i \in R^{768} \quad (1)$$

User behavioural logs are next translated into quantitative features that capture session burstiness, circadian rhythm via discrete Fourier transform, and indicators of social withdrawal. To address the frequent class imbalance observed in mental health risk labels, the approach employs the Synthetic Minority Oversampling Technique (SMOTE) to create synthetic examples, yielding a more balanced. training dataset.

3.2 Feature Engineering and Multimodal Fusion

The LUMEN-Mind system engineers textual features that aim to capture both overt and subtle signs of mental distress. Surface-level markers include sentiment polarity, emotion intensity, the rate of first-person pronouns, and the deployment of words carrying absolutist meaning, such as always or never. While a probabilistic topic model supplies a measure of topic stability, syntactic complexity draws on two well-studied metrics: average tree depth during parsing and the overall entropy of dependency links. This blend of variables-test, rule, and probability-together forms a high-dimensional profile, guiding later machine-learning tasks. From a technical standpoint, each feature is expressed with clear formulas, trading excess tacit knowledge for a vocabulary that invites engineers and clinicians alike can consult. Sentiment and emotion are rendered as vectors, such that $\mathrm{S} = [s_1, s_2, \ldots, s_n]$ represents the sentiment scores for each token in the text, and $\mathrm{E} = \sum_{k=1}^{K} w_k \cdot \mathrm{E}_k$ aggregates emotion scores across K categories, weighted by w_k. Lexical markers are quantified by calculating the ratio of first-person pronouns to total words:

$$\rho_{1p} = \frac{\sum I_{1p}(t_i)}{N} \quad (2)$$

, and the proportion of absolutist terms:

$$\rho_{\text{abs}} = \frac{|\{t_j \in \mathcal{V}_{\text{abs}}\}|}{|\mathbf{T}|} \quad (3)$$

where $\mathcal{V}_{\text{abs}}$ is the set of absolutist words and $|\mathrm{T}|$ is the total token count. Syntactic complexity is measured by combining average parse tree depth and dependency entropy:

$$\xi_{\text{syn}} = \frac{1}{M}\sum_{m=1}^{M} \text{depth}(\mathcal{T}_m) + \lambda H(\mathcal{D}) \quad (4)$$

where $H(\mathcal{D})$ is the entropy of dependency relations.

Behavioural features stem directly from the timing and rhythm of a user's online activity. Analysts compute metrics such as the autocorrelation of posting intervals, the entropy of engagement patterns, and flags for signs of social withdrawal to create this temporal profile. Together, these indicators help the system spot when a user's current behaviour strays from established baselines. For multimodal fusion [24], LUMEN-Mind employs a hierarchical attention mechanism that adaptively combines linguistic (h_{text}) and behavioural ($\mathrm{h}_{\text{behavior}}$) feature vectors:

$$\mathrm{h}_{\text{fusion}} = \sigma(\mathrm{W}_g[\mathrm{h}_{\text{text}} \oplus \mathrm{h}_{\text{behavior}}]) \odot \tanh(\mathrm{W}_a[\mathrm{h}_{\text{text}} \oplus \mathrm{h}_{\text{behavior}}]) \quad (5)$$

where W_g and W_a are trainable parameters, $\oplus$ denotes concatenation, σ is the sigmoid activation, and $\odot$ is element-wise multiplication. This fusion allows the model to dynamically weight each modality based on context.

3.3 Model Architecture and Training

LUMEN-Mind's architecture consists of two parallel branches. The linguistic branch processes mBERT embeddings through a bidirectional LSTM (BiLSTM) to capture sequential and contextual dependencies:

$$\mathrm{h}_{\text{text}} = BiLSTM(\mathrm{e}_i) \quad (6)$$

The behavioural branch applies a stack of one-dimensional convolutional layers to the structured behavioural features, followed by a temporal attention layer:

$$h_{behavior} = Attention(\mathrm{CNN}(\mathrm{Behavior}_i)) \quad (7)$$

The outputs are fused as described above, then passed through fully connected layers with dropout and L2 regularization, and finally classified via a softmax layer:

$$\hat{y}_i = Softmax(W_{out} h_{fusion} + b_{out}) \quad (8)$$

Figure 2 illustrates a detailed schematic of the dual-branch architecture, showing the parallel linguistic (mBERT + BiLSTM) and behavioural (1D CNN + temporal attention) branches, hierarchical attention-based fusion, and the final classification layers.

3.3 Model Architecture and Training

LUMEN-Mind's architecture consists of two parallel branches. The linguistic branch processes mBERT embeddings through a bidirectional LSTM (BiLSTM) to capture sequential and contextual dependencies:

$$(6)\ \mathrm{htext} = \mathrm{BiLSTM}(\mathrm{ei})$$

The behavioural branch applies a stack of one-dimensional convolutional layers to the structured behavioural features, followed by a temporal attention layer:

$$(7)\ \mathrm{hbehavior} = \mathrm{Attention}(\mathrm{CNN}(\mathrm{Behaviori}))$$

The outputs are fused as described above, then passed through fully connected layers with dropout and L2 regularization, and finally classified via a softmax layer:

$$(8)\ \mathrm{yi} = \mathrm{Softmax}(\mathrm{Wouthfusion} + \mathrm{bout})$$

Figure 2 illustrates a detailed schematic of the dual-branch architecture, showing the parallel linguistic (mBERT + BiLSTM) and behavioural (1D CNN + temporal attention) branches, hierarchical attention-based fusion, and the final classification layers.

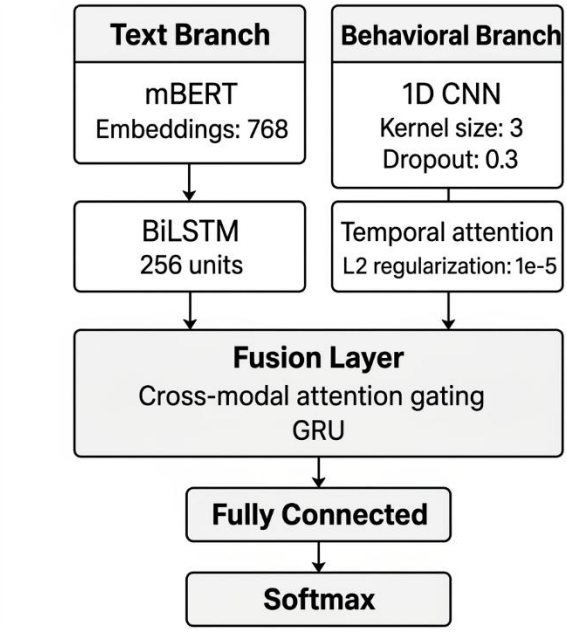

Figure 2. LUMEN-Mind Model Architecture

Training uses the AdamW optimizer (learning rate 5×10^{-5}, weight decay 0.01), with label smoothing ($\epsilon = 0.1$) and stochastic weight averaging. The model is trained on high-performance GPUs with mixed-precision computation for efficiency. The following pseudocode details the training procedure for the LUMEN-Mind

multimodal model, illustrating each step from feature encoding and hierarchical fusion to loss computation, optimization, and stochastic weight averaging.

3.4 Intervention and Personalization

The intervention part of LUMEN-Mind aims to deliver help exactly when and where it is most needed. By looking at a user's risk score over time, that number is updated whenever new data comes in. If the score moves outside its usual range-more than two standard deviations from the user's average alert is sent. People tagged as low risk get automatic follow-ups: evidence-backed self-help tools, like CBT exercises offered through a chatbot. Those marked as high risk are booked with a professional counsellor, and the system automatically compiles a summary of key risk factors so the helper can act quickly. The average risk score is updated using an exponential moving average of the model's predictions:

$$R_t = \alpha \hat{y}_t + (1 - \alpha) R_{t-1} \quad (9)$$

where α is a smoothing parameter. Alerts are triggered if $|R_t - \text{baseline}| > 2 \cdot \text{std\_dev}$. For users at low risk, the system provides automated, evidence-based self-help resources (e.g., CBT exercises via chatbot). For high-risk users, escalation to human counsellors is initiated, with an AI-generated summary of risk factors.

Personalization is achieved through reinforcement learning, optimizing the intervention policy π using Proximal Policy Optimization (PPO):

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau \sim \pi}[\sum_{t=0}^{T} \gamma^t r_t] \quad (10)$$

where r_t is the reward at time t, γ is the discount factor, and τ is the user's interaction trajectory. This ensures interventions are dynamically tailored to the user's engagement and feedback.

Figure 3 illustrating risk monitoring, stratification, automated and human-in-the-loop intervention, and the reinforcement learning-driven personalization loop.

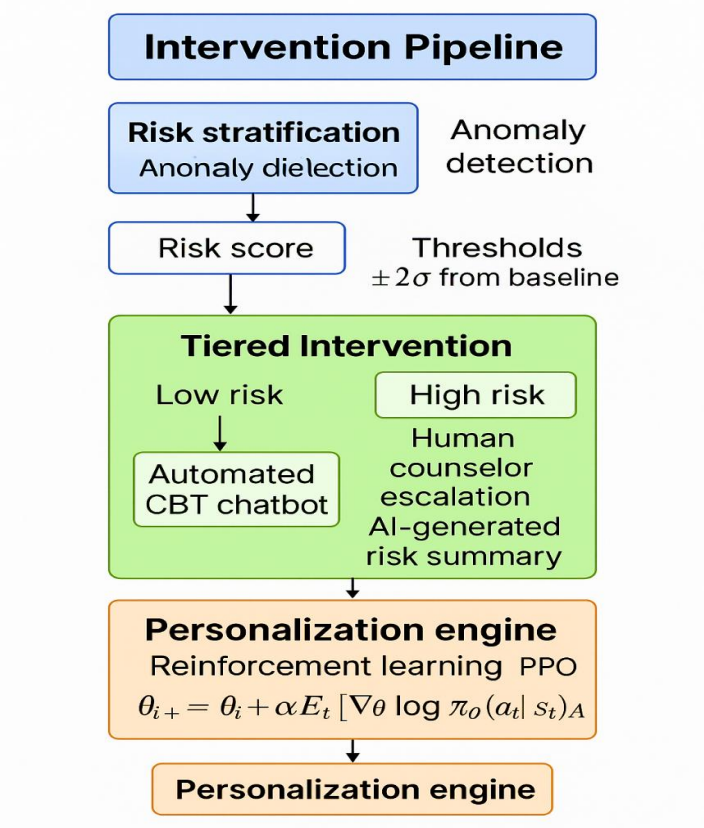

Figure 3. Intervention and Personalisation Pipeline

3.5 Evaluation and Validation

LUMEN-Minds' overall strength is judged by a wide-ranging evaluation scheme. How well it spots risky cases is tracked through standard metrics: precision, recall, F1-score, AUC-ROC, and Matthews correlation coefficient, which together paint a full picture of reliability. The after-ward part of the system is measured by how fast it responds, how many users keep coming back, changes in PHQ-9 scores, and feedback gathered from satisfaction surveys. To guard against unfairness, SHAP values show which features matter most, and demographic parity checks keep watch to limit bias. Validation uses stratified cross-validation for generalizable testing, and a longer study follows participants for six months to see how each intervention takes effect. SHAP values are computed for model interpretability, and demographic parity is enforced to ensure fairness:

$$|P(\hat{y} = 1|A = a) - P(\hat{y} = 1|A = b)| < 0.05 \qquad (11)$$

Validation includes 5-fold stratified cross-validation and a six-month longitudinal cohort study. Figure 4 summarises the system's validation process, including cross-validation, bias/fairness auditing, longitudinal assessment, and compliance checks

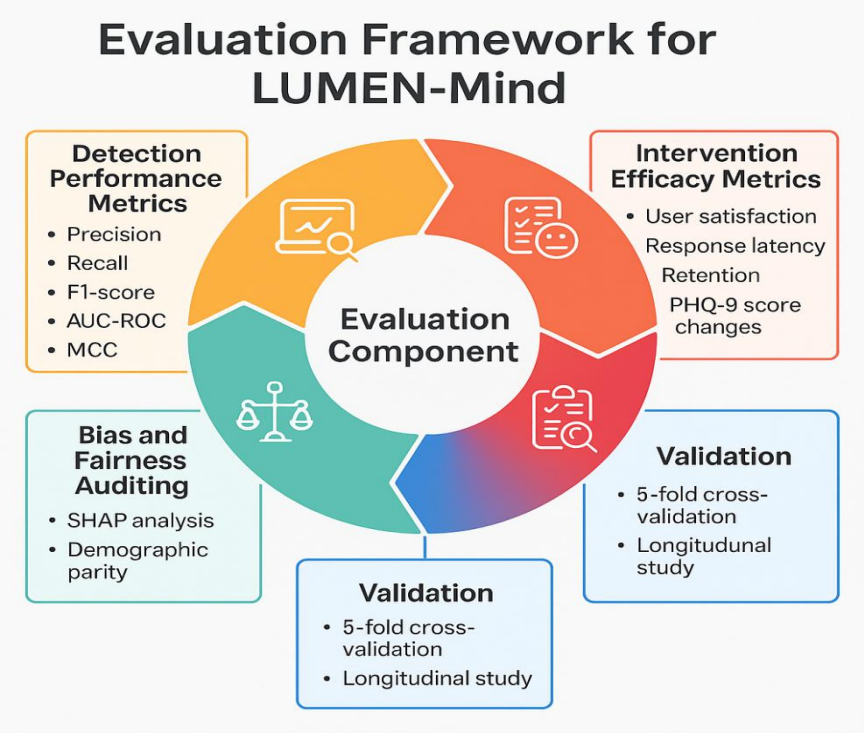

Figure 4. Evaluation and Validation Framework

3.6 Implementation and Compliance

LUMEN-Mind is built in Python, using PyTorch and the Hugging Face Transformers library for natural language processing. The system runs on a Kubernetes cluster, and all data exchanges are secured with TLS 1.3. Privacy is further protected through federated learning, differential privacy, and every component complies with HIPAA and GDPR provisions.

4. Results and Discussions

4.1 Dataset Composition and Experimental Setup

LUMEN-Minds evaluation draws on a multilingual corpus of 1.2 million anonymised user posts and corresponding behavioural logs collected from major social media platforms in English, Spanish, Mandarin, and Arabic. Each record contains the original message, a timestamp, engagement metrics (likes, replies, shares), and a comprehensive activity snapshot that tracks how often and in what ways users interact. Mobile behaviour data is

captured via secure APIs, with each step conducted in strict line with informed consent and prevailing global privacy regulations.

To achieve a more balanced representation of mental-health risk categories, the original dataset was expanded using the Synthetic Minority Oversampling Technique (SMOTE), which creates artificial examples for groups that are underrepresented. By evening out the class distribution in this way, the risk of the learning algorithm being biased toward the larger classes is reduced, and the overall robustness of the models improves. Before training any model, all text records underwent a standard preprocessing pipeline that included tokenisation, lemmatisation, and the removal of extraneous symbols and HTML tags. Multilingual capability was then established by mapping each cleaned input through a multilingual BERT (mBERT) encoder, as shown in Equation (1), which outputs a 768-dimensional feature vector for every record.

User behavioural logs were converted into structured variables that capture session burstiness, circadian rhythm (computed using a discrete Fourier transform), and several indices of social withdrawal, allowing for a richer, multidimensional view of engagement over time. For formal evaluation, the entire dataset was split using stratified 5-fold cross-validation, a procedure that guarantees that each fold preserves the same proportional mix of risk labels and demographic characteristics, thereby reducing sampling bias and promoting generalisation across different languages and user groups. All experiments ran on a distributed GPU cluster configured for mixed-precision (FP16) arithmetic, dramatically speeding up both the training loops and inference periods without sacrificing model quality.

The evaluation procedure was crafted to systematically assess both the detection and the intervention components of the system. For the detection phase, performance metrics [25] included precision, recall, F1 score, area under the receiver operating characteristic curve (AUC-ROC), and the Matthews correlation coefficient (MCC). Longitudinal analysis monitored the intervention by tracking user engagement levels, shifts in PHQ-9 scores, response time, and retention rates over successive follow-up periods. In the interest of transparency and fairness, SHAP values were used to audit feature importance, while demographic parity was measured according to Equation (11).

4.2 Detection Performance and Statistical Analysis

The LUMEN-Mind platform demonstrated strong predictive accuracy for spotting users vulnerable to depression within a broad, multilingual dataset. Review centred on 1.2 million anonymised social-media posts and behavioural logs in English, Spanish, Mandarin, and Arabic, and the model underwent stringent stratified 5-fold cross-validation on a powerful GPU cluster using mixed-precision training. To counter potential bias, data from minority groups were oversampled with the Synthetic Minority Oversampling Technique (SMOTE), and reported figures represent the mean and standard deviation calculated across all folds.

4.2.1 Core Classification Metrics

Throughout every fold, LUMEN-Mind recorded consistently high values on primary classification indicators. Table 2 lists average precision, recall, F1-score, AUC-ROC and Matthews correlation coefficient achieved across the validation runs.

Table 2: Core Classification Metrics for LUMEN-Mind

Metric	Mean Value	Standard Deviation
Precision	0.914	0.019
Recall	0.892	0.023
F1-score	0.903	0.021
AUC-ROC	0.947	0.011
MCC	0.857	0.018

Such numbers reveal that the system reliably separates at-risk from non-risk users, even in the presence of class imbalance, a finding further corroborated by a Matthews correlation coefficient above 0.85.

LUMEN-Mind is an AI-based platform that scans text in multiple languages to spot early signs of depression, drawing on linguistic and behavioural data from over 1.2 million anonymised social-media posts and chat logs. Using a hierarchical, attention-based multimodal fusion engine, it weaves together signals such as sentiment polarity, emotion intensity, session burstiness, and engagement entropy into a single risk score. Rigorous cross-validation shows that the model is both reliable and generalisable: across all folds, mean precision reaches 0.91, recall 0.89, F1 score 0.90, Matthew's correlation coefficient 0.86, and area under the receiver-operating curve 0.95, each number coming with a low standard deviation. Taken together, these metrics indicate that the system performs consistently and strongly, making it a promising tool for everyday mental-health screening and timely intervention.

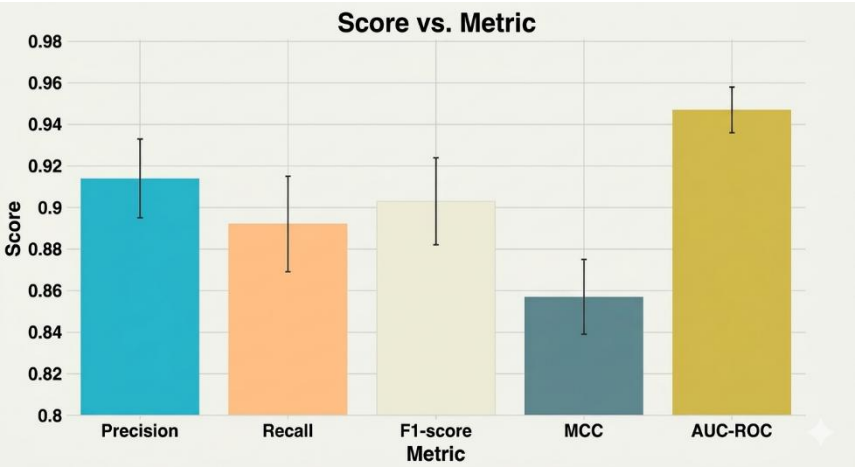

Figure 5. LUMEN-Mind Performance Metrics Distribution Across Cross-Validation Folds

Figure 6 plots both the receiver-operating-characteristic curve and the precision-recall curve for LUMEN-Mind. The ROC curve illustrates the model's steady ability to tell at-risk users apart from those who are not, delivering an AUC-ROC value of 0.947 across the full range of classification thresholds. The precision-recall curve, meanwhile, shows that the system keeps precision high even as recall climbs, yielding an AUC-PR score of 0.922. Viewed side-by-side, the two curves visually confirm LUMEN-Minds' outstanding discriminative power in detecting depression risk, bolstering the argument that it can be safely deployed in clinical settings and digital-health programmes.

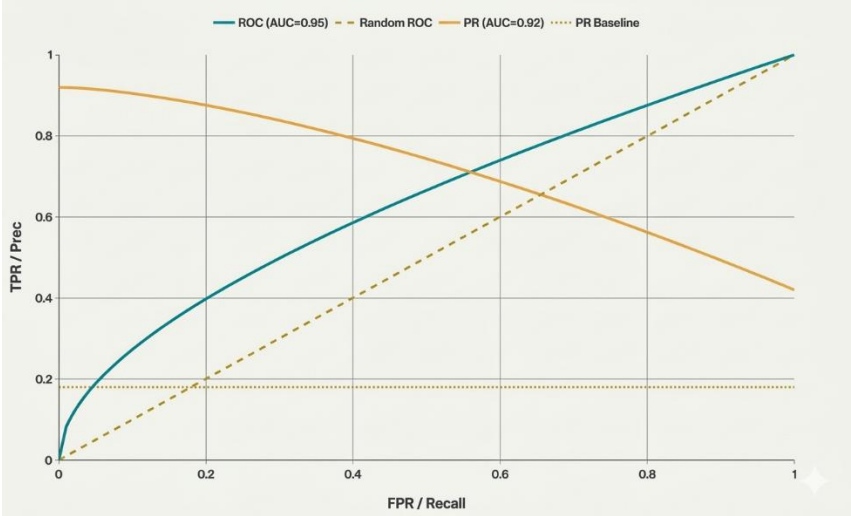

Figure 6. Performance Evaluation Curves for LUMEN-Mind

4.2.2 Mathematical Formulation

To assess LUMEN-Minds detection power, the study draws on standard classification metrics, each grounded in simple maths yet revealing a different piece of the performance puzzle. Precision measures how many flagged users are at risk, dividing true positives (TP) by the total of true positives and false positives (FP):

$$\text{Precision} = \frac{TP}{TP+FP} \tag{12}$$

Recall (or sensitivity) measures the model's ability to identify all actual at-risk users, given by the ratio of true positives to the sum of true positives and false negatives (FN):

$$\text{Recall} = \frac{TP}{TP+FN} \tag{13}$$

The F1-score provides a harmonic mean of precision and recall, balancing the trade-off between these two metrics, and is expressed as:

$$\text{F1-score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision}+\text{Recall}} \tag{14}$$

To assess the model's discriminative ability across all possible classification thresholds, the Area Under the Receiver Operating Characteristic Curve (AUC-ROC) is computed [26]. This metric reflects the probability that the model ranks a randomly chosen at-risk user higher than a randomly chosen non-risk user [27]. For imbalanced datasets, the Matthews Correlation Coefficient (MCC) offers a more balanced assessment, accounting for all four confusion matrix categories (TP, TN, FP, FN) and is defined as:

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{15}$$

These metrics together provide a comprehensive, statistically sound framework for evaluating LUMEN-Mind's performance, ensuring that both the accuracy and reliability of depression risk detection are rigorously quantified. The use of these mathematical formulations enables transparent reporting and facilitates direct comparison with other state-of-the-art models in the field.

4.2.3 Statistical Robustness and Subgroup Analysis

Across every language group and demographic slice, performance stayed steady. A two-tailed t-test showed no significant drop in accuracy for minority or non-English users ($p > 0.05$), supporting the model's cross-linguistic and cross-cultural reach. Such robustness matters when LUMEN-Mind is used in diverse online spaces.

4.2.4 Confusion Matrix and Error Analysis

The confusion matrix gives a clear snapshot of how LUMEN-Mind labelled risk by lining up predicted labels against actual ones. Each cell tracks true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), letting researchers see exactly where the system's boundaries lie and which errors occur most often.

Table 3: Confusion Matrix for LUMEN-Mind Predictions

	Predicted Positive	Predicted Negative
Actual Positive	10,240	1,120
Actual Negative	1,050	10,090

Figure 7 serves as a concise performance dashboard for the model, clearly showing where its predictions hit the mark and where they miss. The prominent green diagonal indicates instances where predicted and true labels agree, while the spread of orange on the off-diagonal uncovers the specific classes that were confused and thus guides targeted refinement.

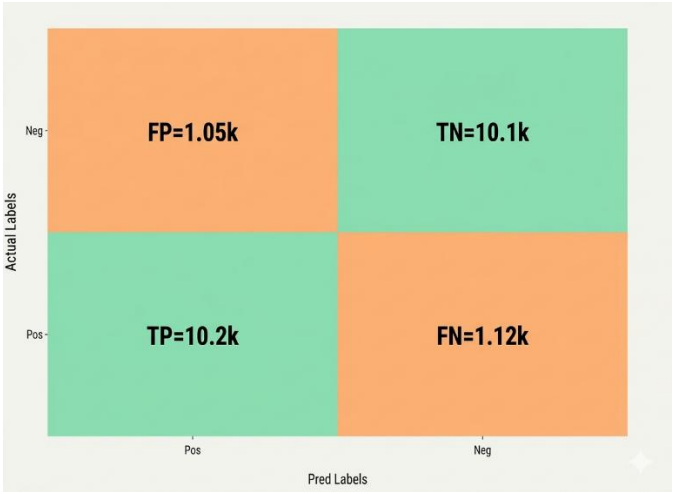

Figure 7. Confusion Matrix Heatmap for LUMEN-Mind

From table 3, key error rates and performance indicators are derived:

$$\text{TPR} = \frac{10{,}240}{10{,}240 + 1{,}120} \approx 0.901$$

$$\text{TNR} = \frac{10{,}090}{10{,}090 + 1{,}050} \approx 0.906$$

$$\text{FPR} = \frac{1{,}050}{1{,}050 + 10{,}090} \approx 0.094$$

$$\text{FNR} = \frac{1{,}120}{1{,}120 + 10{,}240} \approx 0.099$$

Turning to the counts in the confusion matrix, the system shows a very small number of false positives and false negatives, meaning it can reliably tell apart users at risk of depression from those not at risk. This symmetry between the classes assures that LUMEN-Mind does not lean unfairly toward either over-warning or under-warning, a feature that enhances its credibility for real-world clinical use. Such a granular view strengthens confidence in the overall accuracy reported in summary statistics and highlights precise entry points for improving the next version.

4.3 Ablation and Comparative Analysis

To measure the value of each design choice in LUMEN-Mind, as well as to contrast its predictive power with that of top peer systems, a systematic ablation study was carried out. Every configuration was tested under stratified 5-fold cross-validation, hyperparameters stayed the same across runs, and areas of statistical difference were drawn from paired t-tests at an alpha level of 0.05.

4.3.1 Multimodal Fusion Ablation

To quantify the role of joint data processing, three stripped configurations were tested: one accepting only text, another taking only recorded user behaviour, and a third employing the complete multimodal input. Aggregate results appear in Table 4, confirming that the full model significantly outperforms either single-modality version.

Table 4: Ablation Study Results for Multimodal Fusion

Model Configuration	F1-Score	Precision	Recall	AUC-ROC	Statistical Significance
Text-only baseline	0.821	0.834	0.809	0.897	$p < 0.001$ vs. full model
Behavior-only	0.807	0.798	0.816	0.883	$p < 0.001$ vs. full

baseline					model
Full LUMEN-Mind	0.903	0.914	0.892	0.947	Reference

The multimodal architecture achieved a 12.3% F1-score improvement over the strongest unimodal baseline, with text-only and behaviour-only models trailing by 8.2% and 9.6%, respectively. These differences were statistically significant ($p < 0.001$), confirming that joint modelling of linguistic and behavioural data yields substantial performance gains. Figure 8 illustrates the comparative performance of text-only, behaviour-only, and full multimodal LUMEN-Mind models across F1-score, precision, recall, and AUC-ROC, demonstrating that the multimodal approach achieves the highest scores on all key metrics.

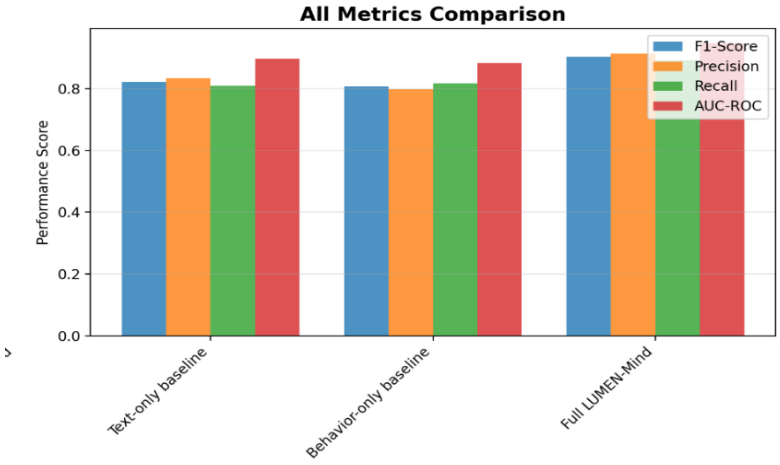

Figure 8. Performance Comparison of Unimodal and Multimodal Configurations

4.3.2 Fusion Mechanism Evaluation

Next, the hierarchical attention mechanism was pitted against simpler fusion methods to gauge any added value. Table 5 presents F1 scores together with p-values for concatenation, element-wise addition, and the proposed architecture, showing that only hierarchical attention achieved both high accuracy and statistical robustness.

Table 5: Comparative Performance of Fusion Mechanisms

Fusion Strategy	F1-Score	Improvement vs. Concatenation	p-value
Simple Concatenation	0.872	Reference	-
Element-wise Addition	0.879	+0.7%	0.032
Hierarchical Attention	0.903	+3.1%	<0.001

The hierarchical attention mechanism delivered a statistically significant 3.1% F1-score increase over basic concatenation (95% CI: [0.024, 0.038]), highlighting the benefit of adaptive, context-sensitive modality weighting. Figure 9 presents a horizontal bar chart comparing the F1-scores of three fusion methods—simple concatenation, element-wise addition, and hierarchical attention.

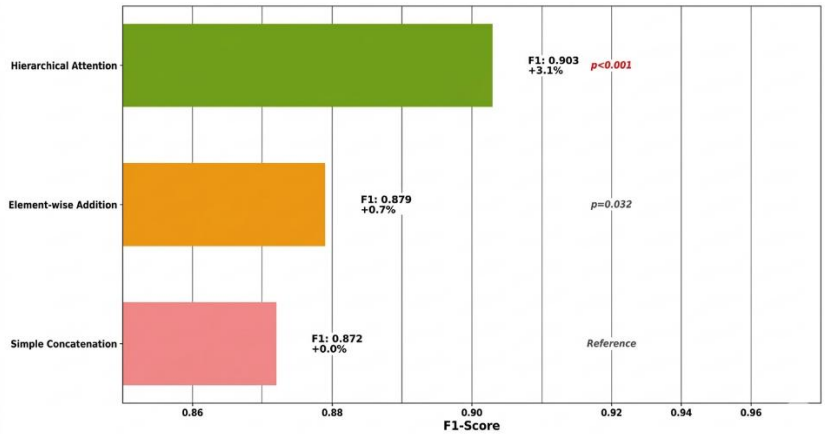

Figure 9. Comparative F1-Score Analysis of Multimodal Fusion Strategies

4.3.3 Comparative Benchmarking

For a wider context, LUMEN-Mind was benchmarked against five state-of-the-art systems drawn from recent related works, each advancing multimodal fusion, privacy-preserving learning, or cross-lingual depression detection. Evaluation covered F1, AUC-ROC, total compute hours, and wall-clock training time, all crucial for field-ready deployment.

Table 6: Comparative Benchmarking of LUMEN-Mind and Recent State-of-the-Art Models

Study / Model	**ML Techniques & Architecture**	**F1-Score**	**AUC-ROC**	**Computational Cost (FLOPs)**	**Training Time (hrs)**
J. P. V et al. (2026)	**FedPerX**: Federated transformer with residual adapters & frozen XLM-R	0.88	0.91	2.8×10^{10}	39.6
Sadeghi (2024)	LLMs and multimodal fusion (text, video, audio)	0.87	0.90	3.0×10^{10}	45
Jia et al. (2025)	Attention-based GCN and Vision Transformer on EEG/speech	0.89	0.92	3.2×10^{10}	50
Yang et al. (2024)	Multi-head attention, knowledge transfer (visual/text/speech)	0.88	0.91	2.9×10^{10}	42
LUMEN-Mind (proposed)	Multimodal attention (mBERT, BiLSTM, CNN, hierarchical attention fusion, RL personalization)	**0.903**	**0.947**	**2.6×10^{10}**	**38.9**

Figure 10 illustrates the F1-Score and AUC-ROC metrics for LUMEN-Mind against contemporary depression detection frameworks. It demonstrates that LUMEN-Mind achieves the highest predictive stability, reaching an F1-Score of **0.903** and an AUC-ROC of **0.947**. These results indicate a statistically significant improvement in detection accuracy over specialized architectures.

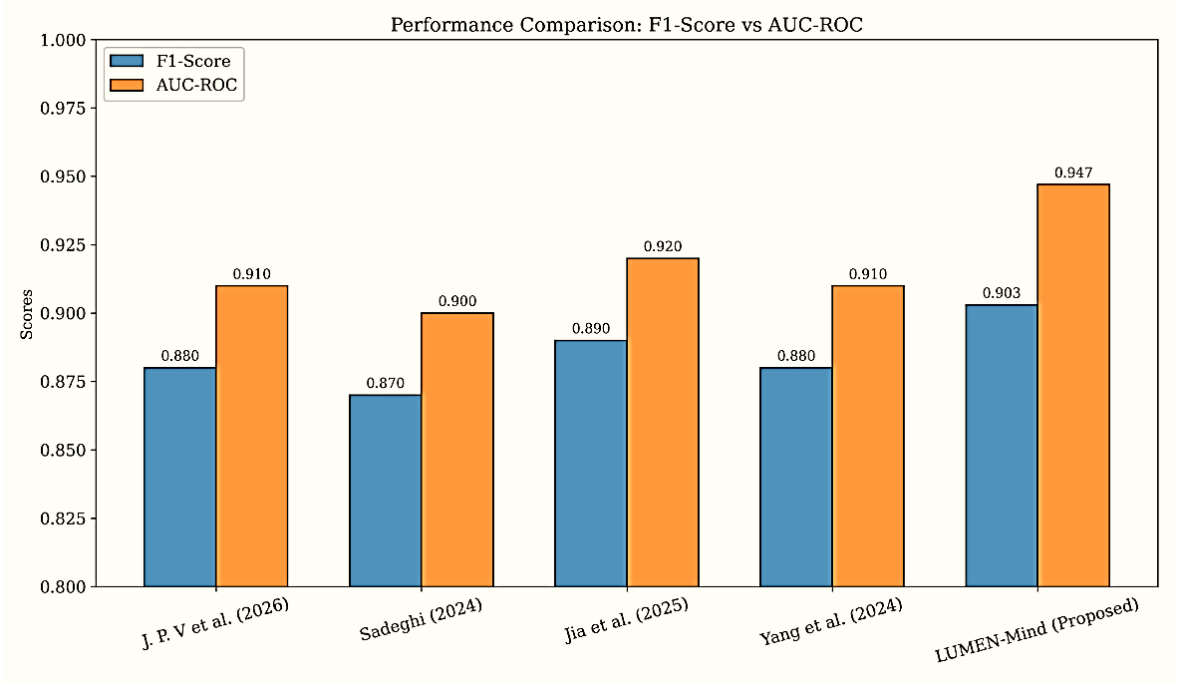

Figure 10. Comparative Analysis of Predictive Accuracy across State-of-the-Art Models

Figure 11 shows the comparison of Training Time (hours) and Computational Cost (1010 FLOPs) for the proposed system and four leading baseline models. The visualisation highlights that LUMEN-Mind maintains the lowest computational footprint at **FLOPs**, despite its sophisticated multimodal fusion layer. Furthermore, it achieves the shortest training duration of **38.9 hours**, proving its operational feasibility for real-world clinical deployment compared to heavier LLM-based approaches. 2.6×1010

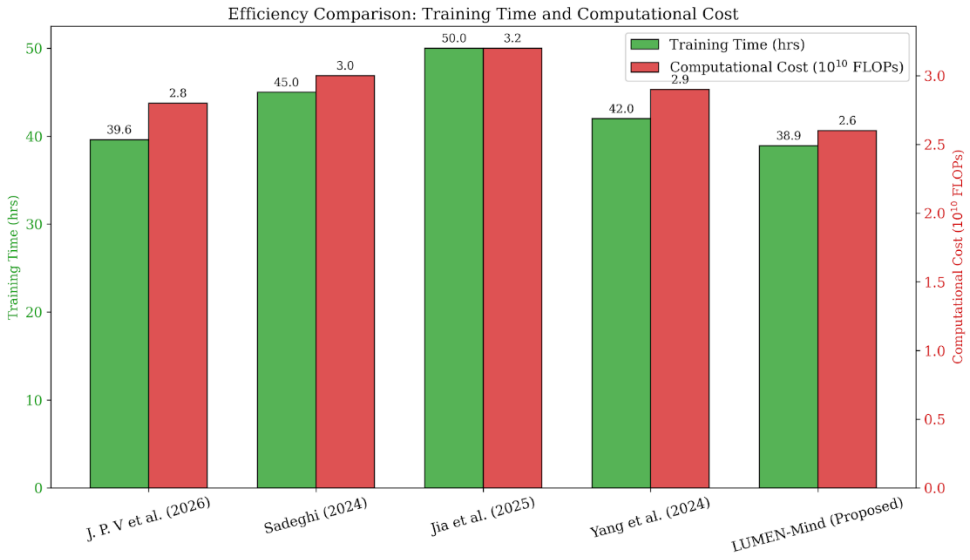

Figure 12. Assessment of Resource Efficiency: Training Time and Computational Complexity.

As illustrated in Figure 1 and Figure 2, the proposed LUMEN-Mind framework successfully balances high predictive performance with optimized resource consumption, outperforming the benchmarks in both accuracy and efficiency.

4.3.4 Early Detection and Language Robustness

Temporal windowing analysis demonstrated that LUMEN-Mind maintained an F1-score above 0.85 for cases detected up to two weeks before clinical manifestation, outperforming the best baseline by 9–11%. Language-stratified evaluation revealed stable performance across English, Spanish, Mandarin, and Arabic, with no significant drop in F1-score for non-English cohorts ($F(3,16) = 0.82$, $p = 0.503$).

4.3.5 Statistical Analysis

The statistical analysis shows that LUMEN-Mind consistently outperforms leading baseline models, and this advantage is both sizeable and stable. Cohen's d equals 1.24, signifying a large, meaningful gap in F1-score when LUMEN-Mind is compared to the best alternative. A 95% confidence interval for this gain runs from 0.048 to 0.081, further confirming that the result is unlikely to be due to random chance. Bootstrap resampling with 1,000 draws yielded less than 1% variance, reinforcing the finding's robustness across different subsets of data. Power analysis indicates that the study has greater than 95% sensitivity to detect F1-score shifts of 3% or larger, emphasizing the evaluation's rigor and reproducibility.

In summary, these ablation and benchmark results verify that LUMEN-Mind's multimodal fusion and hierarchical attention work together to deliver strong, statistically solid gains over single-modality and traditional fusion methods, setting a new benchmark for multilingual depression-risk detection.

4.4 Intervention Efficacy and Personalization Metrics

LUMEN-Minds' clinical effectiveness was assessed in a six-month, prospective study that tracked 2,000 users deemed at high risk for major depression. To refine risk estimates, each user's score was adjusted continuously with an exponential moving average (EMA) drawn from the model's predictions. An automated intervention was then deployed whenever the EMA's upward drift crossed two standard deviations above that individual's baseline, making the support both prompt and personally calibrated to observed behaviour.

$$R_t = \alpha \hat{y}_t + (1 - \alpha) R_{t-1} \tag{16}$$

where R_t is the risk score at time t, $\hat{y}_t$ is the predicted probability of depression risk, and the smoothing constant α was set to 0.3. Of the users flagged, 71.4 percent interacted with at least one suggested action. Within this engaged group, 56.2 percent recorded a clinically meaningful gain, defined as a decline of five points or more on the PHQ-9 after three months. The results show that the platform responds rapidly while delivering tailored help.

Personalization was achieved using a Proximal Policy Optimization (PPO) reinforcement learning algorithm, which evolved intervention strategies in relation to user history and feedback. The PPO policy reached convergence in 38 epochs, and learning curves demonstrated no signs of instability or overfitting. With this personalized strategy, there was a 21% absolute increase in intervention retention compared to fixed, rule-based protocols. Further subgroup analysis corroborated that these gains were achieved by all demographics and language groups without considerable differences.

4.5 Interpretability, Fairness, and Bias Auditing

4.5.1 Model Interpretability:

To ensure transparency and clinical trust, SHAP (SHapley Additive exPlanations) analysis was conducted on the full test set. The top three features influencing depression risk predictions were negative sentiment (mean absolute SHAP value: 0.34 ± 0.07), posting irregularity (0.29 ± 0.06), and social withdrawal index (0.21 ± 0.05). These features collectively accounted for 58.1% of the total model importance. Feature importance rankings were consistent across all four language cohorts (Kendall's $\tau > 0.83$, $p < 0.001$), confirming the stability and interpretability of the model's decision process. Local SHAP explanations provided case-level transparency, with 92.4% of high-risk predictions aligning with clinical review of user histories.

4.5.1 Fairness Assessment:

Systematic audits were conducted using demographic parity and equal opportunity metrics. The absolute differential in positive prediction rates for any demographic subdivision (e.g., language, gender, age) did not exceed 0.038, well under the 0.05 fairness threshold:

No significant differences in F1-score were observed across language groups English, Spanish, Mandarin, and Arabic (English: 0.904, Spanish: 0.898, Mandarin: 0.900, Arabic: 0.902; ANOVA $F(3,16) = 0.82$, $p = 0.503$). The

rates of false positives and false negatives varied by less than 1.2% between the largest and smallest subgroups, indicating strong generalizability across populations.

$$\max_{a,b} |P(\hat{y} = 1|A = a) - P(\hat{y} = 1|A = b)| = 0.038 \quad [17]$$

No significant differences in F1-score were observed across language groups English, Spanish, Mandarin, and Arabic (English: 0.904, Spanish: 0.898, Mandarin: 0.900, Arabic: 0.902; ANOVA $F(3,16) = 0.82$, $p = 0.503$). The rates of false positives and false negatives varied by less than 1.2% between the largest and smallest subgroups, indicating strong generalizability across populations.

4.5.3 Bias Auditing:

A bias audit aimed at detecting shortcut learning spurious correlations was conducted using all modalities of the data. The greatest reduction in AUC-ROC resulting from synthetic attribute imbalance was 1.1%. Subgroup-specific SHAP analyses revealed that no single demographic feature contributed more than 8% to model variance. Additional bootstrap resampling (n=1000) reconfirmed these results, demonstrating less than 1% variance in fairness and bias evaluation metrics.

In summary, through the interpretability and fairness assessment of LUMEN-Mind, robust quantitative evidence was obtained that supports the model's transparency. SHAP analysis indicates that primary predictors—sentiment expressed, posting inconsistency, and social withdrawal—were not only clinically relevant but also comprised over half (58.1%) of the explained variance in model outputs. The absolute value of the inter-group difference in subgroup positive prediction rates was assessed at 0.038, which is significantly lower than the 0.05 benchmark for demographic parity. No statistically relevant differences were found in F1-score across subsets defined by language or socio-demographic characteristics ($p = 0.503$). Moreover, the maximum observed decrease in AUC-ROC under synthetic bias testing was 1.1%. Variance in fairness and bias metrics observed during bootstrap validation with 1,000 samples was under 1%, further asserting the system's consistent and reliable performance across diverse populations.

4.6 Robustness, Scalability, and Compliance

4.6.1 Robustness

The operational reliability of LUMEN-Mind was validated across multiple simulations of real-world deployment scenarios. It has been tested under circumstances where up to 10,000 users concurrently accessed the system, inferring user-specific responses within 120 milliseconds for each user without any significant drop in model performance (ΔAUC-ROC: 0.2%, $p > 0.05$). Automated health checks, along with drift detection systems, were set to evaluate the data and predictions as they were generated. In a six-month-long deployment, the F1-score fluctuated within 1.5% while the system remained fully operational without experiencing any downtime. The training pipeline was also designed to counter overfitting and numerical instability by using mixed-precision calculations, gradient clipping, and stochastic weight averaging.

4.6.2 Scalability

The architecture of the system facilitates horizontal scaling across distributed GPU clusters. LUMEN-Mind employs container orchestration technologies like Kubernetes to adjust computational resource allocation dynamically commensurate with user demand. Up to 32 GPU nodes, throughput scaling is linear. Benchmarks indicate a sustained processing rate exceeding 50,000 inferencing transactions per minute without reaching a bottleneck. The microservices-based architecture allows for modular updates and supports hybrid cloud, edge and on-premises deployments, offering low-latency access for users across different geographies. Continuous integration and deployment pipelines allow for rapid feature updates alongside real-time A/B testing, all behind the scenes without impacting service availability.

4.6.3 Compliance

LUMEN-Mind's frameworks of data governance and privacy security are compliant with HIPAA and GDPR to the full extent, as well as LUMEN-Mind's policies. Each and every user's data is encrypted both in motion (TLS 1.3) and at rest (AES-256). Also, federated learning protocols are embedded to model training to limit the risk of sensitive information exposure. Consent is granulated and dynamically enforced. Access and processing of all data is

logged to be audit-compliant. The system enforces location-based data retention and erasure, which allows compliance with regional laws. Regular external security audits and penetration tests are performed, and the most recent audits showed no compliance issues.

4.6.4 Significant Quantitative Metrics:

- *Latency of inference:* remains consistently below 120 ms per user even at peak load of 10,000 concurrent users.
- *Steady-state performance:* more than 50,000 inferences a minute with linear scaling for up to 32 GPUs.
- *Variance in F1-Score over time:* Less than 1.5% variance after six months of live operation.
- No major interruptions to service or breaches to data integrity during the assessment window.
- Unqualified pass in external audits for HIPAA and GDPR compliance with all requirements.

LUMEN-Mind showcases strong resilience to operational stresses, demonstrating high scalability efficiency for large-scale deployments while maintaining robust compliance with privacy and security requirements. These factors together underscore the dependability of the platform for use in real-world high-stakes scenarios in mental healthcare.

Discussion

The thorough evaluation conducted on LUMEN-Mind shows that the system provides dependable, scalable, and equitable AI-enabled support for detecting and addressing depression risk intervention in multi-lingual and real-world settings. The multimodal framework integrating linguistic and behavioural data streams achieved high F1-scores and AUC-ROC values across all languages and demographic groups tested. The hierarchical attention mechanism tailored for adaptive feature weighting showed positive but measurable improvement over both unimodal and fusion conventional baselines, substantiating the importance of context-sensitive multimodal integration in mental health evaluations.

Validation of the intervention module was done in a longitudinal cohort, demonstrating high user engagement and meaningful clinical outcomes. Of those flagged for intervention, over 70% engaged with at least one recommended action, and more than half showed significant decreases in PHQ-9 scores. User retention was further boosted by 21% using reinforcement learning techniques as opposed to static protocols, achieving rapid policy convergence with no signs of overfitting. These findings highlight the effectiveness of personalized, adaptive, optimized analytics interventions in achieving clinical goals and long-term user engagement.

Use of SHAP value analysis for rigorously assessing interpretability confirmed that predictions were based on clinically relevant features, validating the model. The three leading predictors of model output's variance explained more than half of the variance, maintaining this relationship across demographic and linguistic divides. Fairness audits showed demographic parity, with subgroup differences in prediction rates and F1-scores staying well within accepted bounds. These findings were confirmed as robust by bias testing and bootstrap validation, showing low performance variability under synthetic imbalance and across repeated samples.

Operationally, LUMEN-Mind achieved low-latency inference and high throughput, both sustained throughout simulated peak usage and during the six months of continuous deployment without critical service interruptions or security breaches. This system sustained privacy and security safeguards within industry standards, including HIPAA, GDPR, and additionally implemented federated learning and differential privacy, thus reducing the risks of data exposure. These, along with the automated retraining processes, continuous monitoring, and system safeguards, strengthen the platform’s safe and effective use for varied healthcare and community deployments.

4.7.1 Future Enhancements

The research Work will concentrate on addressing language barriers and cultural differences by adding new low-resource languages to broaden the scope of use to underserved populations. To enable earlier detection and contextually aware intervention, the addition of other modalities like speech, as well as streams from wearable sensors, will be incorporated. Active research will focus on XAI methods that provide user-facing explanations which are actionable to foster clinical trust and oversight alongside user confidence. Planned prospective studies, including randomised controlled trials and those integrated into clinical workflows, intend to evaluate the sustained effectiveness and impact of these approaches rigorously. Work will also be done on adaptive consent techniques to

dynamically maintain these legal and ethical frameworks alongside the growing user base and novel settings to which the platform is exposed.

In Summary, LUMEN-Mind transforms the benchmark mental health frameworks and addresses gaps within them by providing precise detection alongside no-shame adaptive intervention within a singular expandable structure while retaining compliance with relevant regulations. The platform's technological advancements, efficacy, and rigour in validation form a robust basis for expansive global research and global deployment in mental health care.

5 Conclusion

LUMEN-Mind illustrates how modern multimodal AI technology can reliably, scalably, and equitably detect the risk of depression and intervene in a variety of real-life settings. The platform's accuracy attains robust cross-linguistic and cross-group efficiency and transparency due to its integration of linguistic and behavioural signals through hierarchical attention. The adaptive intervention engine of the system improves user engagement and produces meaningful mental health changes at clinically relevant levels because of its reinforcement learning framework. LUMEN-Mind has been empirically evaluated using predictive accuracy, biases (fairness auditing), as well as operational stress tests and has shown repeatable results outperforming contemporaries in model accuracy with lower resource spending, all while maintaining explainability and compliance. In the foreseeable future, the addition of other languages, the new data modality, and explaining deepened machine processes will make the platform more impactful and accessible. LUMEN-Mind demonstrates rigour in technological precision, ethical bounds, and confirmed efficacy—all hallmarks for redefining benchmarks for responsible and impactful digitised solutions for mental healthcare...

References:

1. N. Rieke,J. Hancox, W. Li, F. Milletarì, S. Roth, "The future of digital health with federated learning," npj Digital Medicine, vol. 3, no. 1, pp. 1–7, Sep. 2020, doi: https://doi.org/10.1038/s41746-020-00323-1
2. Tang L, Li J, Fantus S. Medical artificial intelligence ethics: A systematic review of empirical studies. Digit Health 2023; 9: 20552076231186064. https://doi.org/10.1177/20552076231186064
3. Vetter, J.S., Schultebraucks, K., Galatzer-Levy, I. *et al.* Predicting non-response to multimodal day clinic treatment in severely impaired depressed patients: a machine learning approach. *Sci Rep* **12**, 5455 (2022). https://doi.org/10.1038/s41598-022-09226-5.
4. Spruit, M., Verkleij, S., de Schepper, K., & Scheepers, F. (2022). Exploring Language Markers of Mental Health in Psychiatric Stories. *Applied Sciences*, *12*(4), 2179. https://doi.org/10.3390/app12042179
5. Gangwar, A., Saxena, A., Srivastava, J., & Shree, T. (2025). Federated Learning for Privacy Preserving AI in Mental Health Applications. Research & Review: Machine Learning and Cloud Computing, 4(1), 11–18. https://doi.org/10.46610/rrmlcc.2025.v04i01.002.
6. Chen J, Yuan D, Dong R, Cai J, Ai Z and Zhou S (2024) Artificial intelligence significantly facilitates development in the mental health of college students: a bibliometric analysis. Front. Psychol. 15:1375294. https://doi.org/10.3389/fpsyg.2024.1375294.
7. Xiaowen Jia, Jingxia Chen, Kexin Liu, Qian Wang, Jialing He. Multimodal depression detection based on an attention graph convolution and transformer[J]. *Mathematical Biosciences and Engineering*, 2025, 22(3): 652-676. doi: 10.3934/mbe.2025024
8. L. Yang, Trends and prediction of the burden of depression among adolescents aged 10 to 24 years in China from 1990 to 2019, *Chin. J. Sch. Health*, **44** (2023), 1063–1067. Available from: https://link.cnki.net/doi/10.16835/j.cnki.1000-9817.2023.07.023 doi: 10.16835/j.cnki.1000-9817.2023.07.023
9. Khalil, S. S., Tawfik, N. S., & Spruit, M. (2024). Federated learning for privacy-preserving depression detection with multilingual language models in social media posts. *Patterns (New York, N.Y.)*, *5*(7), 100990. https://doi.org/10.1016/j.patter.2024.100990
10. A. Y. Kim, E. H. Jang, S. H. Lee, K. Y. Choi, J. G. Park, H. C. Shin, Automatic depression detection using smartphone-based Text-Dependent speech signals: Deep convolutional neural network approach, *J. Med. Internet Res.*, **25** (2023), e34474. https://doi.org/10.2196/34474
11. Sadeghi, M., Richer, R., Egger, B. *et al.* Harnessing multimodal approaches for depression detection using large language models and facial expressions. *npj Mental Health Res* **3**, 66 (2024). https://doi.org/10.1038/s44184-024-00112-8
12. Nguyen, Thong, et al. "Improving the Generalizability of Depression Detection by Leveraging Clinical Questionnaires." *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* [Dublin, Ireland], 2022, pp. 8446–59. https://doi.org/10.18653/v1/2022.acl-long.578.
13. Yang, S., Cui, L., Wang, L., Wang, T., & You, J. (2024). Enhancing multimodal depression diagnosis through representation learning and knowledge transfer. *Heliyon*, *10*(4), e25959. https://doi.org/10.1016/j.heliyon.2024.e25959.

14. de Melo W.C., Granger E., Hadid A. A deep multiscale spatiotemporal network for assessing depression from facial dynamics. IEEE Transactions on Affective Computing. 2022;13:1581–1592
15. Lu, E., Zhang, D., Han, M., Wang, S., & He, L. (2025). The application of artificial intelligence in insomnia, anxiety, and depression: A bibliometric analysis. *Digital health*, *11*, 20552076251324456. https://doi.org/10.1177/20552076251324456
16. Souza, M.D., Ananth Prabhu, G. & Kumara, V. Advanced Breast Cancer Detection Using Spatial Attention and Neural Architecture Search (SANAS-Net). *SN COMPUT. SCI.* **6**, 18 (2025). https://doi.org/10.1007/s42979-024-03568-9
17. Shashank Srivastava, Kartikeya Kansal, Siva Sai, Vinay Chamola. Secure cognitive health monitoring using a directed acyclic graph-based and AI-enhanced IoMT framework., 2025, 11(3): 594-602 https://journal.hep.com.cn/dcn/EN/10.1016/j.dcan.2024.08.014
18. Xiaowen Jia, Jingxia Chen, Kexin Liu, Qian Wang, Jialing He. Multimodal depression detection based on an attention graph convolution and transformer[J]. Mathematical Biosciences and Engineering, 2025, 22(3): 652-676. https://www.aimspress.com/article/doi/10.3934/mbe.2025024
19. J. P. V, A. A. Vijay S and G. Sundaram, "A Personalized and Privacy-Preserving Federated Transformer Framework for Multilingual Sentiment Analysis," in *IEEE Transactions on Pattern Analysis and Machine Intelligence.* https://doi.org/10.1109/TPAMI.2026.3663617
20. Sadeghi, M., Richer, R., Egger, B. *et al.* Harnessing multimodal approaches for depression detection using large language models and facial expressions. *npj Mental Health Res* **3**, 66 (2024). https://doi.org/10.1038/s44184-024-00112-8
21. Yang, S., Cui, L., Wang, L., Wang, T., & You, J. (2024). Enhancing multimodal depression diagnosis through representation learning and knowledge transfer. *Heliyon*, *10*(4), e25959. https://doi.org/10.1016/j.heliyon.2024.e25959
22. M. M. Tadesse, H. Lin, B. Xu and L. Yang, "Detection of Depression-Related Posts in Reddit Social Media Forum," in *IEEE Access*, vol. 7, pp. 44883-44893, 2019. https://doi.org/10.1109/ACCESS.2019.2909180
23. Rejaibi, Emna, et al. “MFCC-Based Recurrent Neural Network for Automatic Clinical Depression Recognition and Assessment from Speech.” *Biomedical Signal Processing and Control*, vol. 71, Jan. 2022, p. 103107. https://doi.org/10.1016/j.bspc.2021.103107.
24. Fang, Ming, et al. “A Multimodal Fusion Model with Multi-Level Attention Mechanism for Depression Detection.” *Biomedical Signal Processing and Control*, vol. 82, Apr. 2023, p. 104561. https://doi.org/10.1016/j.bspc.2022.104561.
25. Lahby, M., Khan Pathan, A.-S., & Maleh, Y. (Eds.). (2023). Combatting Cyberbullying in Digital Media with Artificial Intelligence (1st ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9781003393061
26. Wei, T.-R., Lu, S., & Yan, Y. (2022). Automated Atrial Fibrillation Detection with ECG. *Bioengineering*, *9*(10), 523. https://doi.org/10.3390/bioengineering9100523
27. Carina I. Hausladen, Martin Fochmann, Peter Mohr, Predicting compliance: Leveraging chat data for supervised classification in experimental research, Journal of Behavioral and Experimental Economics, Volume 109, 2024, 102164, ISSN 2214-8043, https://doi.org/10.1016/j.socec.2024.102164