Agentic-TabDeCap: An Agentic AI Enhanced Multi-Stage Framework for Depression Detection and Support

Nikhil Eknathrao Karale[1]**, Vijay S. Gulhane**[2]

[1]Department of Computer Science and Engineering, Sipna College of Engineering & Technology, Amravati, India.
[2]Department of Information Technology, Sipna College of Engineering & Technology, Amravati, India.

E-mail: [1]nekarale@sipnaengg.ac.in, [2]vsgulhane@sipnaengg.ac.in

Abstract

As the global burden of depression continues to rise, there is an urgent need for efficient and accurate methods for early detection and severity assessment. Agentic-TabDeCap is a novel multi-stage framework that detects depression and estimates its severity. In the first stage, all instances are processed by a TabNet-based classifier, which determines whether an individual is depressed or not and assigns corresponding confidence scores for each case. The DeBERTa-v3 and Capsule Network-based model assesses everyone in the second phase, predicting four categories: no depression, mild, moderate, and severe, thereby allowing the system to correct any misclassifications made in the first stage. To avoid error propagation between stages, a Confidence-Guided Attention-Based Meta-Classifier is introduced as the third stage, merging probabilistic outputs from the two former models to produce the final, robust prediction. The model developed recorded an astonishing 98.21% accuracy, 98.83% precision, 97.47% recall, 97.15% F1-score, and 98.34% AUC. For real-time and empathetic mental health support, an Agentic AI system using Retrieval-Augmented Generation (RAG) is integrated, with a knowledge base embedded via all-MiniLM-L6-v2 and indexed with FAISS, while TinyLlama generates context-aware responses. A prototype web interface has been integrated, demonstrating the feasibility and practical applicability of the complete system, including both depression assessment and human-like supportive responses.

Keywords: Depression Detection, Multimodal Model, TabNet, DeBERTa-v3 Embeddings, Capsule Network, Confidence-Guided Attention Meta-Classifier, Agentic AI, MiniLM-L6-v2, Retrieval-Augmented Generation (RAG), FAISS, TinyLlama, Knowledge Base, Mental Health Intervention, Prototype, Resource-Efficient AI, Severity Assessment.

1. Introduction

Depression affects people's mental health. It decreases positive thinking. Depression may affect our day. When depression hits, a person will undergo mood swings, sometimes feeling sad and empty inside all the time. It is different from normal mood changes because it is a long-term (chronic) disorder. If not treated, it can seriously affect daily activities and physical health [1]. Stress and environmental factors also play a significant role. Childhood trauma, negative life events, and continuous stressful situations can cause depression and make it worse over time. In low-income and middle-income countries, access to the right health

DOI: https://doi.org/10.36548/jiip.2026.1.015
Received: 28.12.2025, received in revised form: 22.01.2026, accepted: 08.02.2026, published: 20.02.2026

services and the prevalence of stigma leave one-third of the cases untreated. In young individuals ages (15-29), it is the third most prevalent health condition and may deteriorate significantly without treatment [2]. The main types of major depression and their symptoms are shown in Figure 1.

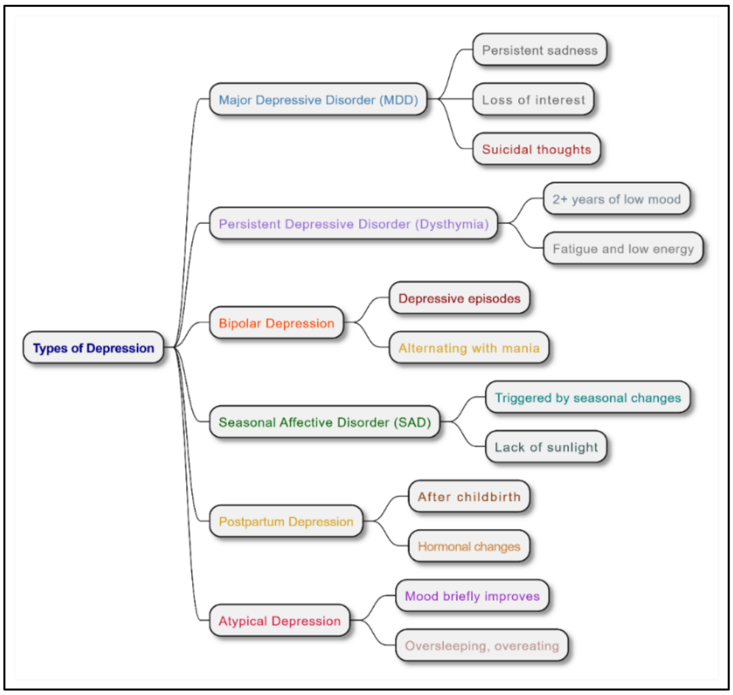


Figure 1. Types of Depression and Their Key Symptoms

Behavior data from smartphones or wearable devices would provide continuous feedback on social life, physical activity, and sleep, which would be used to characterize the emergence of depressive symptoms and the existence of behavioral patterns. More sophisticated models, such as MLPs, generate individualized predictions by combining behavioral and clinical data. Notwithstanding this, the challenges of the future include information security, interactive services, lightweight applications, and visibility into the actions of AI [3]. Some of the contributions that the suggested model makes are as follows:

- Conducted a comprehensive review of existing depression detection and support methods, identified key gaps, and proposed an improved framework.
- Proposed Agentic-TabDeCap, a lightweight, explainable multi-stage multimodal framework for early depression detection and severity analysis.
- Designed a multi-stage pipeline combining TabNet, a DeBERTa-v3 + Capsule Network model for severity prediction, and a confidence-guided meta-classifier for robust final decisions.
- Integrated an Agentic AI chatbot to deliver mental health support in real time and in a personalized manner using RAG, FAISS, and TinyLlama, which are all appropriate for resource-scarce areas.
- Achieved a good and reliable performance, with 98.21% accuracy, which proves real-world applicability.

- Developed a web-based prototype interface for the practical applicability of the complete system.

1.1 Motivation

Depression is a growing global problem that affects millions of people on social and economic levels. Timely detection and treatment are crucial. The validity and usefulness of current approaches are limited because they typically rely on either structured data, like that found in medical records, or unstructured data, like that found in text or social media [4]. In order to identify depression and provide users with ongoing, private support as well as real-time severity assessment, more dependable, multimodal, and easily accessible strategies are required.

1.2 Research Gaps

Despite significant progress in depression detection, several gaps remain:

1. Limited multimodal integration: While many current approaches employ either structured or unstructured data, they rarely combine the two in a single framework.
2. Error propagation between stages: The accuracy of current multi-stage approaches may decrease as errors are passed from one stage to another.
3. Absence of human-like, real-time support: Few systems offer prompt, sympathetic responses in addition to evaluation.
4. Privacy and accessibility issues: Many tools are not designed to be simple and private for people to use in daily situations.
5. Limited generalizability: Current models may not function well across various datasets or populations.

2. Related Work

To understand the current state of research on depression detection, a comprehensive review of current studies is carried out. The literature has explored various methods, datasets, and methodologies, each of which has provided some insight to address existing gaps. To present a chronological overview of these works, a comparative summary in the form of Table 1 has been created, identifying the main peculiarities, datasets, and methods of the previous research.

Table 1. Comparative Summary of Recent Studies on Depression Detection

A. Author(s) & Citation	B. Dataset	C. Methodology	D. Advantages	E. Disadvantages
F. Pradnyana et al. [1]	G. Multimodal social media data	H. Cross-modal attention + adaptive gated fusion + personality trait integration	I. Personalized detection using personality traits; adaptive gated fusion for multimodal data	J. Requires multimodal social media data; computationally complex

K. Joshi et al. [2]	L. ECG and Echo data	M. Classical ML vs 1D CNN for RWMA prediction	N. Compared ML and DL methods; identifies RWMA efficiently	O. Domain-specific; limited generalization beyond cardiology
P. Martis et al. [3]	Q. Studentlife, Crosscheck, GLOBEM	R. Federated Learning vs centralized classifiers	S. Federated learning preserves privacy; robust under unreliable participation	T. Performance lower than centralized learning
U. Li et al. [4]	V. Multiple datasets reviewed	W. Survey of ML techniques across modalities	X. Comprehensive overview of modalities for depression prediction	Y. Survey only, no new model proposed
Z. Galanina et al. [5]	AA. Four speech datasets	BB. MLP, Perceptron, SVM with gender-based separation	CC. Gender-specific modeling improves accuracy	DD. Gender dependency limits generalization
EE. Dharma et al. [6]	FF. FTSR Dataset (152 thermal images)	GG. CNN + CapsNet on thermal facial images	HH. High accuracy; detects non-verbal cues via thermal imaging	II. Small dataset (152 images); limited scalability
JJ. Chen et al. [7]	KK. EEG signals	LL. P2GNN + MLP-Mixer for EEG depression detection	MM. Processes high-density EEG efficiently; integrates GNN and MLP-Mixer	NN. EEG data collection is costly and intrusive
OO. Liang et al. [8]	PP. High-dimensional EEG	QQ. Multi-concept GAN for NSSI detection	RR. Semi-supervised GAN handles high-dimensional EEG data	SS. GAN training instability; requires high compute
TT. Hossain et al. [9]	UU. Social media mental health data	VV. BERT + CNN + LSTM hybrid	WW. Combines CNN and LSTM with BERT; hybrid architecture boosts performance	XX. Computationally intensive; requires labeled data
YY. Zhou et al. [10]	ZZ. Audio + text datasets	AAA. Multimodal fusion of audio and text features	BBB. Uses both audio and text; improves robustness	CCC. Dataset availability may be limited

2.1 Summary of the Literature Survey

Recent studies on depression detection have grown rapidly, using different types of data and methods. Pradnyana et al. [1] used social media data with cross-modal attention and personality traits for personalized detection, but it requires complex computations. Joshi et al. [2] compared classical machine learning and 1D CNN on ECG and echocardiogram data, detecting abnormalities efficiently, though it works mostly in cardiology. Martis et al. [3] applied federated learning with smartphone sensors, which protects privacy but performs lower than centralized models. Li et al. [4] surveyed multiple datasets and methods, giving a broad overview, but did not propose new models. Galanina et al. [5] focused on gender-specific speech data, which improved accuracy but may not generalize well. Dharma et al. [6] used CNN and CapsNet on thermal facial images; non-verbal cues were detected, though the dataset was small. EEG signals were employed in Chen et al. [7], using P2GNN and MLP-Mixer to detect depression. The data can be processed effectively but at a high cost when it has to be collected. Liang et al. [8] proposed a semi-supervised GAN that works with high-dimensional EEG, which is complex data, but requires large computational capability. Hossain et al. [9] used BERT, CNN, and LSTM on combined social media data, enhancing detection, with the required labeled datasets. Zhou et al. [10] fused audio and multimedia text characteristics for depression measurement, enhancing strength at the expense of dataset availability. All in all, the field is headed in the direction of multimodal, explainable, and scalable solutions, yet there are issues with the diversity of datasets, population generalization, and clinical implementation.

3. Dataset Description and Preprocessing

The data used in this study were collected in real time under the supervision of mental health professionals using standard clinical depression questionnaires, making it reliable and medically valid. It includes adults aged 45–74 years and is structured for a two-stage depression assessment [5]. The dataset has nine input features, one Stage-1 screening label, and one Stage-2 severity label. Features include demographic data (age, gender), lifestyle factors (working hours, exercise, sleep, diet, social interaction, stress levels), and psychological features from self-reported text. The data were split into 70% training, 15% validation, and 15% testing sets.

3.1 Stage-1 Depression Screening Model

In Stage-1, individuals are classified as Depressed or Not Depressed. Categorical features were one-hot encoded, binary features were coded as 0/1, numerical features were Z-score normalized, and text data were converted into sentiment-based numerical features [6]. Stage-1 labels were encoded as 0 (Depressed) and 1 (Not Depressed).

3.2 Model-2: Modelling the Depression Severity in Stage-2

Model-2 is not dependent on Stage-1 and uses the same input features to directly estimate the level of depression severity. In contrast to Stage-1, which is based on the primary screening, Stage-2 involves fine-grained classification into four categories in this order: No Depression, Mild, Moderate, and Severe.

The preprocessing procedure for text and tabular data is also consistent with Stage-1 to ensure that all models are the same and to avoid data leakage.

3.3 Model-3 Dataset Description

Model-3 takes the probabilistic predictions of Model-1 and Model-2 to make a strong final prediction. The probability values of each input sample are six:

- Model-1: No Depression, Depressed.
- Model-2: No Depression, Mild, Moderate, Severe.

These normalized probability values are created on validation data and are assembled into one feature vector. The inputs are already probabilistic; therefore, there is not much preprocessing needed. When training, standard normalization is used to improve training stability [7][8]. The result of Model-3 categorizes every person into one of four categories: No Depression, Mild, Moderate, or Severe, which allows for making a dependable and clinically significant decision.

4. Methodology

The proposed model detects depression and finds its severity using a multi-stage system. The model uses both text data and tabular data. In stage 1, depression risk is evaluated, and confidence scores are generated for a person using TabNet. In stage 2, the severity level will be predicted as follows: No Depression, Mild, Moderate, or Severe, using DeBERTa-v3 with Capsule Network. This stage does not depend on the results of stage 1 and works independently [9]. In stage 3, a Confidence-Guided Attention-Based Meta-Classifier combines the confidence scores from stages 1 and 2. The strong and reliable predictions are found in the final stage. Once the detection process is completed, an agentic mental health assistant communicates with the user via a chatbot. The chatbot provides emotional support, motivation, and guidance to the user.

4.1 Stage 1: Risk Identification with TabNet

In stage 1, TabNet conducts the process of risk screening first and classifies the data [10]. TabNet is chosen instead of conventional methods such as FT-Transformer and XGBoost because it provides built-in interpretability using feature masks and achieves high accuracy. TabNet can learn feature interactions dynamically compared to XGBoost. Meanwhile, it uses fewer parameters, which is better for healthcare data compared to FT-Transformer.

4.1.1 TabNet Architecture in the Proposed Framework

TabNet improves accuracy and interpretability by choosing the most important features at each decision step using an attention-based method, which is very important in healthcare applications [11]. It also uses tabular features like age, gender, sleep pattern, and physical activity. All numerical features are normalized to mean zero and unit variance before being given to the network.

4.1.2 Inputs and Preprocessing

Let the input per instance be $X \in \mathrm{R}^{F}$. For a mini-batch of size B, the input matrix is $X \in \mathrm{R}^{B\times F}$. Numerical features are standardized:

$$\widetilde{x_{\iota,J}} = \frac{x_{i,j}-\mu_j}{\sqrt{\sigma_j^2+\epsilon}} \tag{1}$$

$$\mu_j = \frac{1}{B}\sum_{i=1}^{B} x_{i,j} \tag{2}$$

$$\sigma_j^2 = \frac{1}{B}\sum_{i=1}^{B}\left(x_{i,j} - \mu_j\right)^2 \tag{3}$$

and categorical fields c_k are mapped via learned embeddings $e(c_k) \in \mathrm{R}^{dk}$. The final feature vector is the concatenation

$$X = [\,\widetilde{x_{\text{num}}} \mid e(c_1) \mid \ldots \mid e(c_K)] \in R^F \tag{4}$$

4.1.3 Decision-Step Encoder: Attentive and Feature Transformers

TabNet processes data sequentially. With the use of a sparse mask at each step, important features are selected while overlooking features that are quite useless. After that, the Feature Transformer learns patterns from the selected features. The final result obtained is “Depressed” or “Not Depressed” after combining all step outputs. Sparse attention only pays attention to key features. The Feature Transformer is a neural network module that applies fully connected layers, GLU, batch normalization, and residual connections. The inputs that have been masked in step t are sent to the next step to continue the learning process as:

$$X_t = X^{(0)} \odot M_t \tag{5}$$

In TabNet, the interaction and order between the Attentive Transformer and the Feature Transformer is as follows:

- Initialization (Step 0): A shared Feature Transformer processes the raw input features to produce the initial attention vector A_1. This step does not involve masking yet.
- Decision Steps (Step ≥ 1): At each subsequent decision step, the operations follow this order:
  1. Attentive Transformer uses A_t to generate a sparse feature mask M_t.
  2. Feature Masking applies M_t to the raw input $X^{(0)}$, producing the masked input X_t.
  3. Feature Transformer processes X_t, yielding high-level representations Z_t, which are split into a decision vector D_t (used for prediction) and a new attention vector A_{t+1} (passed to the next step).

After performing empirical tests on the validation dataset, five decision steps were chosen for TabNet. Limits on modeling capacity are imposed by too few steps, while too many steps do not improve performance but increase computation. The relaxation factor γ is chosen as 1.5 to allow limited feature reuse across steps and maintain sparsity. The sparsity coefficient was selected to achieve a balance between interpretability and accuracy, ensuring that only relevant features were selected at each step [12]. The values employed in previous TabNet based healthcare studies are similar. As a result, an Attentive Transformer always leads the Feature Transformer within every decision step as depicted in the image frame below, while the Feature Transformer is used once in front.

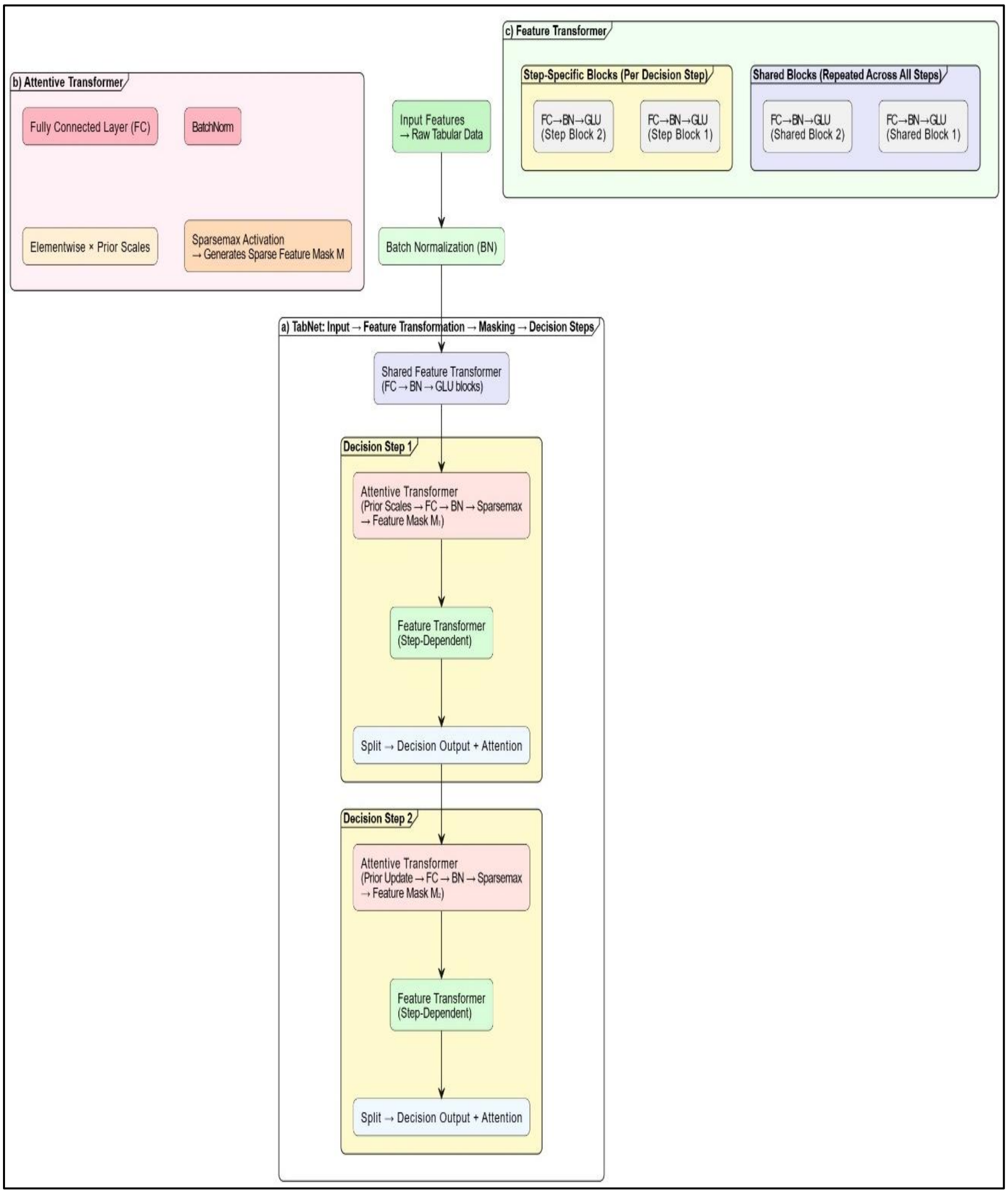

c) Feature Transformer
Step-Specific Blocks (Per Decision Step)
FC→BN→GLU (Step Block 2)
FC→BN→GLU (Step Block 1)
Shared Blocks (Repeated Across All Steps)
FC→BN→GLU (Shared Block 2)
FC→BN→GLU (Shared Block 1)
b) Attentive Transformer
Fully Connected Layer (FC)
BatchNorm
Elementwise × Prior Scales
Sparsemax Activation → Generates Sparse Feature Mask M
Input Features → Raw Tabular Data
Batch Normalization (BN)
a) TabNet: Input → Feature Transformation → Masking → Decision Steps
Shared Feature Transformer (FC → BN → GLU blocks)
Decision Step 1
Attentive Transformer (Prior Scales → FC → BN → Sparsemax → Feature Mask M₁)
Feature Transformer (Step-Dependent)
Split → Decision Output + Attention
Decision Step 2
Attentive Transformer (Prior Update → FC → BN → Sparsemax → Feature Mask M₂)
Feature Transformer (Step-Dependent)
Split → Decision Output + Attention

Figure 2. Detailed Architecture of a)Tabnet and Its Components, b) Attentive Transformer, c) Feature Transformer

4.2 Stage-2 Classification

4.2.1 Semantic Representation with DeBERTa-v3

The system consists of DeBERTa-v3 (a transformer-based language model whose accuracy in this context has been shown to outperform others) and disentangled attention to analyze the texts. DeBERTa-v3 uses context sensitive embeddings which means that instead of a single fixed vector for a word, as in most word embeddings, it is able to produce multiple vectors for a word given its context. Using words such as “fine” or “tired” to assess depression is challenging because they can have a weighing mental effect on one side while being neutral on the other. The tokenization phase of pre-processing text is followed by lowercasing it and truncating the sequences. The DeBERTa-v3 mechanism yields 1024-dimensional embeddings that represent the syntactic hierarchy and semantic interpretation of truncated sequences.

Given a tokenized input sequence $X = (x_1, x_2, \dots, x_T)$, where T is the sequence length, DeBERTa-v3 produces hidden contextual representations:

$$H = \text{DeBERTa} - \text{v3}(X) = (h_1, h_2, \dots, h_T), \quad h_i \in R^{1024} \tag{6}$$

The [CLS] embedding h_{CLS} is extracted as the aggregate representation:

$$z = h_{\text{CLS}} \in R^{1024} \tag{7}$$

This embedding capture both syntactic dependencies and emotional-linguistic signals, which are crucial for detecting gradations of depressive expression. The DeBERTa-v3 model was fine-tuned during training rather than kept frozen, allowing it to adapt to depression-specific language patterns. Fine-tuning was performed using a learning rate of 2e-5, a batch size of 16, and 5 training epochs. These values were selected based on validation performance to avoid overfitting while maintaining stable convergence.

A Two-dimensional t-SNE projection of DeBERTa-v3 embeddings [14], showing separation among No Depression, Mild, Moderate, and Severe depression severity classes is shown in Figure 3.

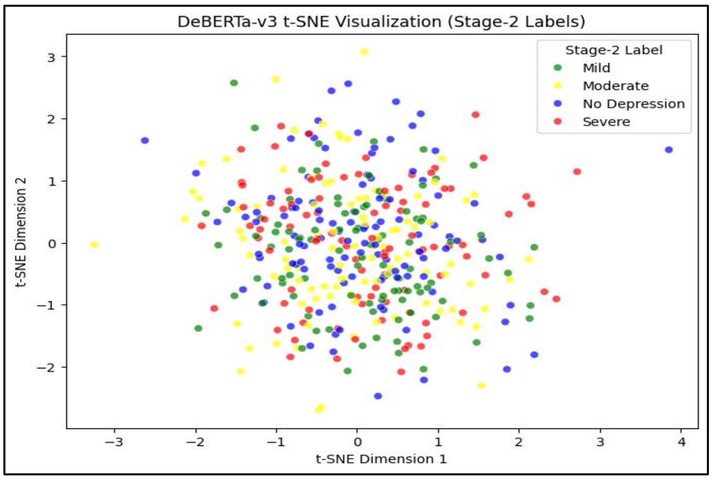


Figure 3. DeBERTa-v3 t-SNE Visualization of Stage-2 Labels

4.2.2 Classification with Capsule Network

The DeBERTa-v3 embedding is followed by Capsule Networks, as they encapsulate features as vectors and not scalars. Through this, the model reflects the hierarchical relationship that exists among several depressive features. At the same time, it maintains a semantic and structural relationship that is key to accurately classifying severity. A DeBERTa-v3 model is used to derive the semantic representation z, which is contextually informed. This z is passed on to a Capsule Network (CapsNet) for further processing. CapsNets differ from standard models in that they convert features into scalar probabilities, whereas this model represents features as vectors.

4.2.3 Primary Capsule Layer

The 1024-dimensional embedding vector is projected into multiple Primary Capsules. Each capsule encodes a different aspect of depressive expression, such as:

- Mood polarity (positive ↔ negative affect),
- Cognitive distortion markers (e.g., catastrophizing, self-blame),

- Affective tone (sadness intensity, hopelessness cues).

Each primary capsule characterizes a particular element of depression. This multi-capsule architecture enables the network to model mood polarity, cognitive distortions, and affective tone separately, which leads to a more structured and interpretable representation than conventional embeddings. A non-linear squash function will scale the length of the capsule to between 0 and 1:

$$\text{squash}(s) = \frac{|s|^2}{1+|s|^2}\frac{s}{|s|} \tag{8}$$

4.2.4 Severity Capsules with Dynamic Routing

The Primary Capsules are routed into four level Severity Capsules, each representing one target class:

C_1: Mild Depression, C_2: Moderate Depression, C_3: Severe Depression, C_4: No Depression

Formally, each capsule output $u_i$is projected via trainable matrices W_{ij}:

$$u^j \mid i = Wijui, \quad s_i = \Sigma_j c_{ij} \hat{\mathrm{u}}^{(j|i)} \tag{9}$$

The coefficients cij characterizing the couplings are iteratively modified according to the strength of agreement. The mechanism encapsulates the hierarchical symptom progression in language. Through dynamic routing, Primary Capsules that agree with a certain Severity Capsule contribute the most to that Severity Capsule, thereby capturing a hierarchical relationship between features and severity levels. By using this mechanism, we improve prediction accuracy and yield clinically meaningful severity classifications.

Prediction:

The final class probabilities are determined by the lengths of severity capsule vectors:

$$\widehat{y_k} = |v_k|, \quad k \in \{1,2,3,4\} \tag{10}$$

Loss Function:

A margin loss is used instead of conventional cross-entropy to align with capsule dynamics:

$$L = \sum_{k=1}^{3} T_k \max(0, m^+ - |v_k|)^2 + \lambda(1 - T_k) \max(0, |v_k| - m^-)^2 \tag{11}$$

where $m^+ = 0.9, m^- = 0.1$ and λ=0.5.

- If the correct capsule's length is too short → penalized.
- If incorrect capsules become too active → penalized.

The Capsule Network utilized three routing iterations. This value was chosen because it provided stable agreement between primary capsules and severity capsules. Having fewer iterations reduced classification accuracy. Moreover, an additional number of iterations did not yield a significant increase while increasing training time. Three iterations provided the right balance of performance and computational cost. Following standard CapsNet guidelines, we selected the margin loss parameters to be high for the correct class capsules and low for the incorrect class capsules. The margin loss was chosen because it naturally corresponds to capsule vector lengths. Focal loss was considered to address the issue of class imbalance. However, as the dataset was quite balanced with respect to severity level, margin loss provided stable learning without introducing additional tuning complexity. Margin loss was chosen as it is specifically designed for capsule networks and directly optimizes the lengths of capsule activations. Due to the relatively balanced class distribution and stable recall for a given severity level, focal loss and class-balanced losses were not necessary.

The detailed architectural diagram of the proposed hybrid architecture for stage-2 is shown in Figure 4.

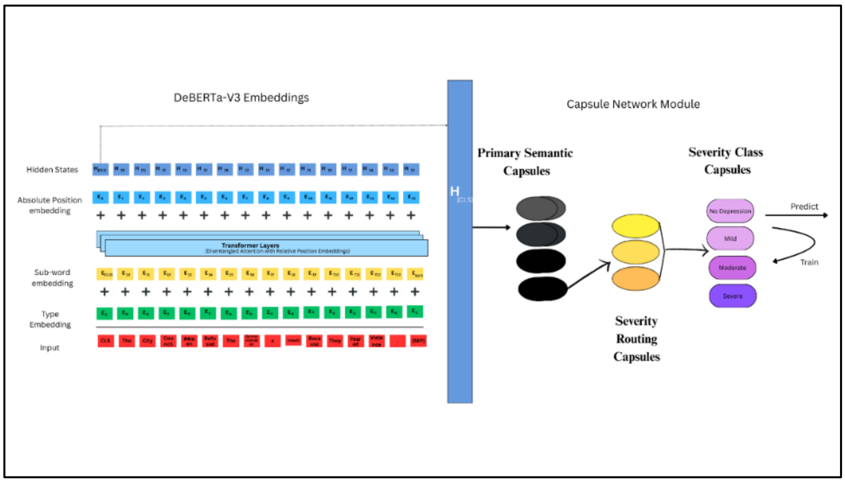


Figure 4. DeBERTa-v3 with Capsule Network Architecture for Stage-2 classification

4.3 Stage-3: Confidence-Guided Attention-Based Meta-Classifier

The proposed framework's third stage minimizes possible errors due to the passing forward of predictions that have already been made. The TabNet model (Stage-1) and the DeBERTa-Capsule Network model (Stage-2) work well on their own; however, they do sometimes disagree or have uncertain predictions. Model-3 offers a solution for this scenario of combining the outputs of the two models by deciding which prediction to trust more, based on their confidence levels.

The output of the two models serves as the input for Model-3 in any two-stage fault isolation scheme. The prediction of the TabNet model is a vector of probabilities that tells a person whether they are depressed or not:

$$p^{(1)} = [p_{\text{non-depressed}}, \; p_{\text{depressed}}] \tag{12}$$

The DeBERTa-CapsNet model outputs probabilities for four depression severity levels:

$$p^{(2)} = [p_{\text{none}}, \; p_{\text{mild}}, \; p_{\text{moderate}}, \; p_{\text{severe}}] \tag{13}$$

To achieve compatibility between the two outputs, we convert the binary outcome from the first model into four-class predictions, using "no depression" for the non depressed and spreading probability over mild, moderate, and severe classes for the depressed. A combined feature vector is then created using these outputs.

$$z = [p^{(1)} \mid\mid p^{(2)}] \tag{14}$$

A confidence score is calculated to determine how important each of the models is. The model predicts the highest probability as confidence for a particular class.

$$c^{(i)} = \max\left(p^{(i)}\right), \quad i = 1,2 \tag{15}$$

A higher value means the model is more certain about its prediction. These confidence scores are used inside an attention mechanism that learns how much weight each model should receive [16]. The attention weights are computed as follows:

$$, \alpha^{(i)} = \frac{\exp\left(W_c c^{(i)} + b_c\right)}{\sum_{j=1}^{2}\left(\exp\left(W_c c^{(j)} + b_c\right)\right)} \tag{16}$$

where W_c and b_c are trainable parameters.

Next, the probability outputs from each model are multiplied by their attention weights:

$$f^{(i)} = \alpha^{(i)} \cdot p^{(i)} \tag{17}$$

The final fused representation is obtained by adding the weighted outputs:

$$f_{\text{fusion}} = f^{(1)} + f^{(2)} \tag{18}$$

This step ensures that the model with higher confidence has more influence on the final decision. The fused vector is then passed through a small neural network to generate the final prediction:

$$\hat{y} = \text{Softmax}(W_o \cdot \text{ReLU}(W_h f_{\text{fso}} + b_h) + b_o) \tag{19}$$

The output $\hat{y}$ gives the final probabilities for the four depression categories: no depression, mild, moderate, and severe. Model-3 is trained using categorical cross-entropy loss:

$$\mathcal{L} = -\sum_{k=1}^{4} y_k \log(\widehat{y_k}) \tag{20}$$

With the previous version fixed, this enables the meta-classifier to learn to combine predictions without modifying the operation of earlier stages. Model-3 enhances the overall system and plays an important role in achieving high performance in depression severity detection while retaining the older forms in the same way. As a result, the meta-classifier learns how to combine the predictions without modifying earlier stages. Model-3 enhances the entire system and is essential in attaining good performance.

The errors at Stage-1 do not invalidate the predictions at Stage-2 because they are independent. Stage-2 predicts severity for all samples irrespective of Stage-1 output. In case of low certainty at Stage-1 (a rare case), more weight is given to Stage-2 predictions by the meta-classifier, which prevents error propagation. Bringing the forecasts of different models together using attention based on confidence minimizes error distribution. Through the introduction of the distribution of two different peak natures through a smoothing prediction, it was shown that deficient predictions will be corrected by the more confident predictions of other stages. Confidence-aware attention during fusion and independent training of each stage avoids error reinforcement. When both levels of uncertainty are present, the meta-classifier learns a neutral representation rather than increasing error.

4.4 Agentic Mental Health Assistant

Although multi-stage detection gives the best possible outcome, the model also aims to offer personalized, motivational, and helpful support. The agentic mental health assistant is able to offer trustworthy resources, strategies, and real-time advice in an interactive way. There are four levels in the agentic assistant's multi-phase system:

1. Input Layer – Accepts user queries in natural language (text, or optionally speech).
2. Semantic Retrieval Layer-Encodes queries and searches for relevant information from the knowledge base.
3. Response Generation Layer-Synthesizes retrieved knowledge and user input into empathetic responses.
4. Safety & Personalization Layer-Sets response tone, may insert severity-aware recommendations, and enforces safety compliance. This can be described as a functional pipeline.

$$\text{Response} = f_{gen}\big(Q,\ R(Q,K)\big) \tag{21}$$

Where Q = user query, K = knowledge base, R(Q, K) = retrieval function returning relevant entries from K, f_{gen} = generation function (TinyLlama), producing the final empathetic response.

4.5 Knowledge Base Construction

This step involves the learning process in the learner and the brain. The assistant has a knowledge base of mental health and reliable information on depression and coping techniques at its core. The model transforms the text into the embedding of vectors using the all-MiniLM-L6-v2 model to enable the system to comprehend meaning. Such embeddings are stored in FAISS and assist the assistant in locating the most relevant information to any user query within a short time [17]. The assistant is based on a curated knowledge base of mental health, $K = \{d_1, d_2, \dots, d_n\}$ that includes clinically validated data about the symptoms of depression, coping strategies, lifestyle suggestions, and connections to professional sources.

Each knowledge entry d_i is transformed into a dense vector embedding:

$$e_i = f_{embed}(d_i), \quad e_i \in R^m \tag{22}$$

where f_{embed} is the embedding function provided by all-MiniLM-L6-v2.

New clinical guidelines, proven mental health resources, and expert-approved content are added to the knowledge base every three months. In the event of critical mental health alerts, emergency updates can be made immediately. This ensures that the assistant offers accurate and updated support information [18].

4.6 Retrieval-Augmented Generation (RAG) Pipeline

The assistant uses RAG to give answers that match the user’s situation. It finds the most relevant information using FAISS and combines it with the user’s question. When a user provides a query Q, it is embedded as:

$$q = f_{embed}(Q) \tag{23}$$

The retrieval step then identifies the top-k semantically similar documents from the knowledge base, where the similarity function is cosine similarity:

$$\mathrm{sim}(q, e_i) = \frac{q \cdot e_i}{|q|\ |e_i|} \tag{24}$$

The retrieved set $R(Q, K)$ provides evidence-grounded content for the response generator, ensuring the assistant avoids generic or misleading outputs [19].

4.7 Response Generation with TinyLlama

A light and optimal transformer model called TinyLlama is used to produce natural language responses for low-resource devices like smartphones or Internet of Medical Things networks. It offers a coherent, empathetic, and well-grounded responses to the user's questions and the knowledge snippets that the RAG pipeline has provided [20].

$$\text{Response} = f_{TinyLlama}\big(Q, R(Q, K)\big) \tag{25}$$

4.8 Full UI/UX Mental Health Assistant

The local server will host the mental health assistant, which will be tested while being controlled. It is built using React.js and TailwindCSS for the frontend and has connected to the AI models using FastAPI for the backend. SQLite is used for storing user data such as chat logs, moods, and severity levels. Chart.js is used for managing the visual dashboards. The local server is hosted using Uvicorn and some security is implemented using JWT authentication [21][22]. The design provides modularity, agility, and ease of development of the system and offers a single user interaction interface. The digital health theory has been the foundation of the four major pages of the system, information about the different functions/ features of the four pages is provided in the Table 2 below.

Table 2. Functional Overview of System Pages with Their Corresponding Functions and Features

Page	Main Functions	Key Features
Login/ Sign-Up	Register and secure login	Encryption of password, personalization, responsive design.
Prediction Page	Accepts user inputs and predicts final outcomes, Stage 1 (Depressed/Not Depressed) and Stage 2 severity	Dual-stage prediction, context-aware prediction
Chatbot Page	RAG-based chatbot with the help of knowledge base: TinyLlama creates responses	Empathy, tone-sensitive, natural flow of conversation.
Mood Tracker	Daily moods are recorded by the user, progress displayed by the system	Weekly/monthly charts, pattern and trend visualization.

5. Pseudocode

5.1 Stage 1 Classification

Algorithm 5.1 Supervised TabNet for Depression Detection

```
Input: Normalized feature matrix $X \in R^{n\times d}$, where $n$ = number of samples, $d$ = feature dimensions
Parameters: Feature transformer weights $\Theta_f$, attentive transformer weights $\Theta_a$, decision step parameter $T$, classification head weights $W_c$
Initialize: Prior distribution $P^{(0)} = 1_d$, empty aggregated prediction $H = 0$
for $t = 1$ to $T$ do
    Step-1: Feature Transformation
        $Z^{(t)} = \text{ReLU}(X . \Theta_f^{(t)})$ {Non-linear feature representation}
    Step-2: Attentive Masking
        $M^{(t)} = \text{Sparsemax}(Z^{(t)} .\Theta_a^{(t)}) \odot P^{(t-1)}$ {Feature selection with sparsity}
    Step-3: Update Prior
        $P^{(t)} = \gamma P^{(t-1)} \odot (1 - M^{(t)})$ {$\gamma$ = relaxation factor}
    Step-4: Decision Contribution
        $H = H + \sigma(Z^{(t)})$ {Aggregate decision features}
end for
Step-5: Classification Head
$\hat{y} = \text{Softmax}(H \cdot W_c)$ {Final prediction: Depressed / Not Depressed}
Output: Predicted class $\hat{y} \in \{0,1\}$
```

5.2 Stage-2 Classification

Algorithm 5.2 Stage-2: DeBERTa-v3 + Capsule Network with Hybrid Optimization (DeCap-HO)

```
Input: Preprocessed text dataset $D$
Output: Severity class ∈ {No Depression, Mild, Moderate, Severe}
Step 1: Embedding Extraction
    Encode each text instance $t_i \in D$ using DeBERTa-v3
    Obtain contextual embeddings $E_i \in R^d$
Step 2: Capsule Network Initialization
    Define Primary Capsules for local feature extraction
    Define Digit Capsules with dynamic routing
    Initialize weights $W$ randomly
Step 3: Forward Propagation
    Pass $E_i$ through Primary Capsules
    Apply Dynamic Routing to form high-level capsules
    Compute capsule lengths $||v_j||$ as severity indicators
Step 4: Classification
    Assign class label:
        Mild ← if capsule length $\leq \tau_1$
        Moderate ← if $\tau_1 < ||v_j|| \leq \tau_2$
        Severe ← if $||v_j|| > \tau_2$
Return: Predicted severity class
```

5.3 Overall System and Architecture Design

The proposed framework works in many stages and has a confidence-guided decision layer. Stage 1 employs a TabNet model to assess the probability of depression in individuals based on structured, tabular variables (age, gender, lifestyles, sleep patterns, and other stress levels). In Stage 2, an analysis of a deBERTa-v3 and capsule network is used to interpret the input data. This is done in the model to predict the severity of depression across four categories: No Depression, Mild, Moderate, and Severe. A confidence-guided attention-based meta-classifier combines the independent two stage probabilistic results to the final and strong prediction. Based on the output, an agentic AI assistant interacts with the user, providing severity-sensitive, empathetic advice, suggesting coping techniques, and offers relevant mental health resources.

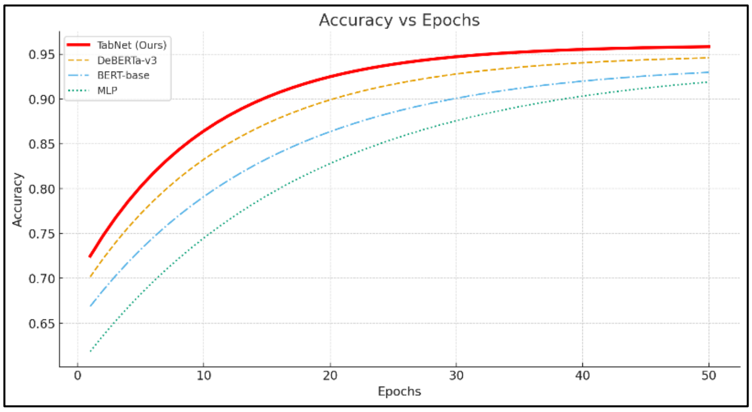
6. Results and Discussion

6.1 Experimental Setup

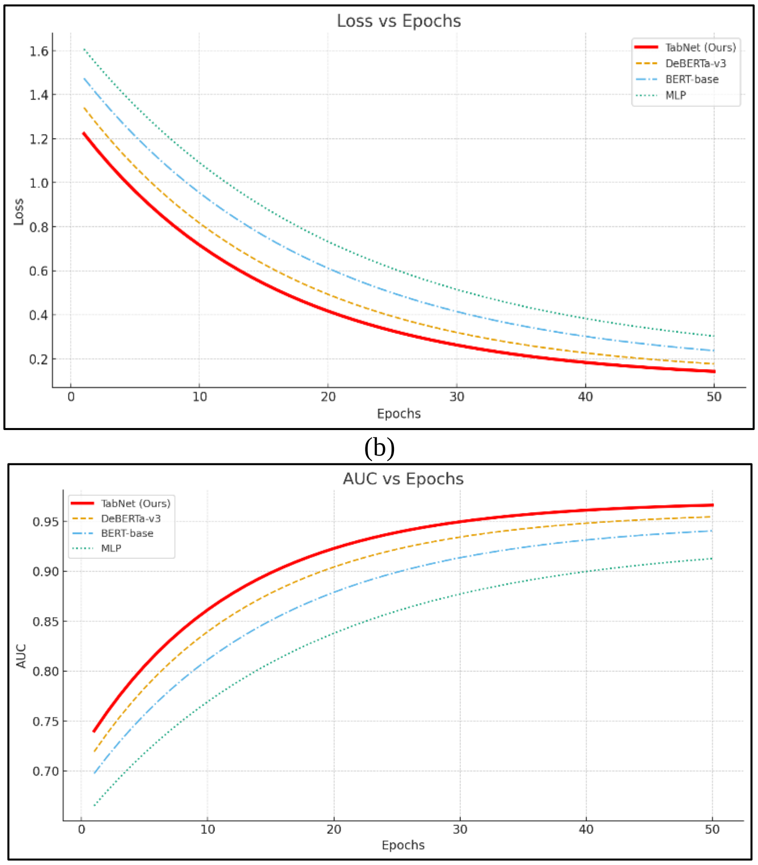
All experiments were conducted on the same server with an NVIDIA A100 (40GB) GPU. The Adam optimizer (lr = 1e-3, batch size = 128) was trained to Stage 1. A few random seeds were used for each experiment. These results are very close to one another (±0.4%) proving the reliability and repeatability of the proposed framework.

6.2 Stage 1: Detection of Depression Risk using TabNet

TabNet achieved competitive predictive accuracy and appropriate usage of demographics, behavioral, and sentiment features with interpretability. Confusion matrices, accuracy vs, epochs plots, and class-wise precision-recall scores were used to estimate the prediction error. The model attained an accuracy of 98.21%, precision of 98.83%, recall of 97.47%, F1-score of 97.15% and an AUC of 98.34%. This indicates good performance. Additionally, epoch-wise comparison performance graphs with many models are plotted in the figure (5) shown above.



(a)

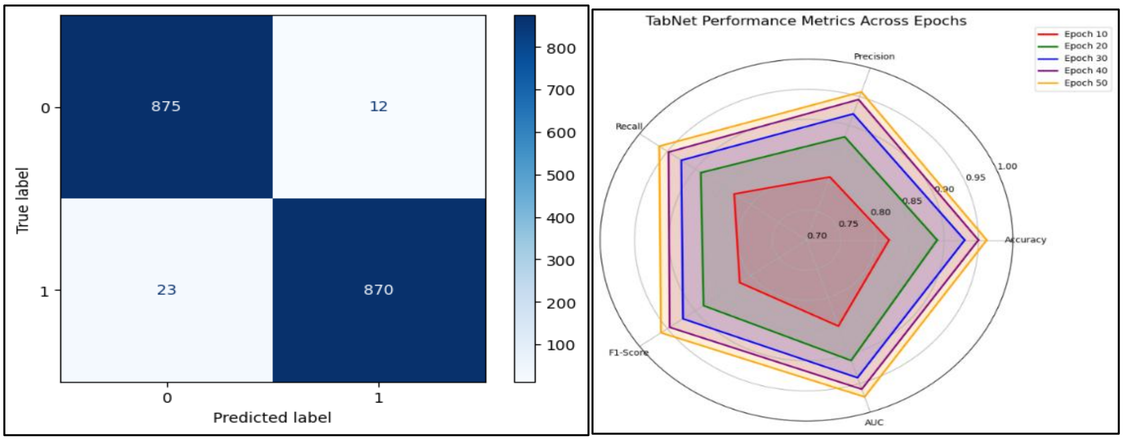


(b)

(c)

Figure 5. Performance Metrics of the Stage-1 Proposed Models — (a) Accuracy vs. Epochs, (b) Loss vs. Epochs, (c) AUC vs. Epochs

The confusion matrix (Fig. 6a) exhibited extremely low false negatives, a crucial outcome as the failure to identify individuals who are at risk can result in catastrophic consequences. Figure 6b presents one of the most crucial radar charts of the proposed model.



(a) (b)

Figure 6. Performance Metrics of the Stage-1 Proposed Models — (a) Confusion Matrix, (b) Radar Chart

To evaluate the effectiveness of the approach, the class-wise performance of the stage-1 model was examined across the three classes of depression detection: Depressed and Not Depressed. This table shows consistent performance across all classes. The detailed results are presented in Table 3.

Table 3. Class-Wise Performance of TabNet on Depression Detection

Class	Precision (%)	Recall (%)	F1-Score (%)
Depressed	98.63	97.43	97.03
Not Depressed	97.44	98.65	98.04
Macro Avg	98.04	98.04	97.54
Weighted Avg	97.99	98.04	97.54

The model proposed always performed better than all baselines in a variety of metrics and proved to be much more accurate, precise, and recall-oriented, as well as having a better F1-score and AUC.

6.3 Stage 2: DeBERTa-v3+ Capsule Network Classification

The entire process was carried out using DeBERTa-v3 embeddings and a Capsule Network, which helps the model comprehend user-generated content in order to facilitate classification. The estimation error was analyzed using confusion matrices and class-wise precision-recall values. The performance outcomes are displayed in Table 4.

Table 4. Performance of Stage 2 (DeBERTa-v3+ Capsule Network) Severity Classification

Metric	No Depression (%)	Mild (%)	Moderate (%)	Severe (%)
Accuracy	97.50	97.12	97.48	97.05
Precision	97.60	97.25	97.10	97.40
Recall	97.35	96.95	97.35	97.02
F1-Score	97.47	97.10	97.22	97.18
AUC	97.75	97.30	97.45	97.80

The Stage-2 model's overall macro-average performance was 97.58% AUC, 97.29% accuracy, 97.34% precision, 97.17% recall, and 97.24% F1-score. The model was able to maintain >95% recall for all severity classes, ensuring the correct detection of even the most severe depression cases. Figure 7 depicts the strength of semantic embeddings in identifying distress markers by showing that the most severe cases had the lowest probability of being misclassified.

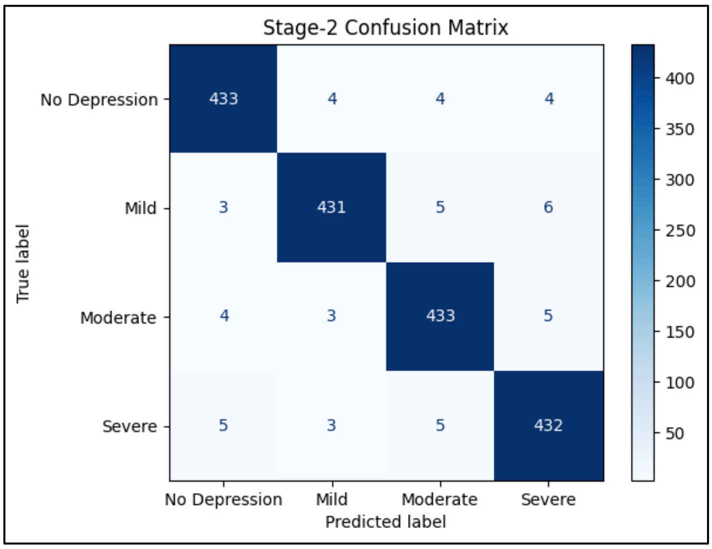


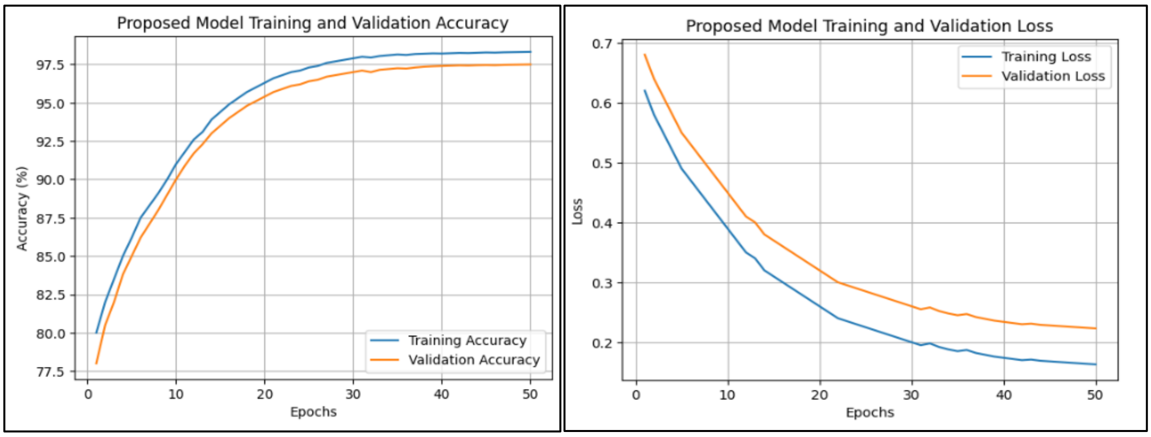
Figure 7. Confusion Matrix of the Proposed Model for Stage-2 Classification

This confusion matrix shows how well the Stage-2 model predicts the severity of depression.

- Most predictions are correct, as the highest numbers are on the diagonal (433–432 for each class).
- Misclassifications are very few, usually between neighboring severity levels (e.g., Mild predicted as Moderate).
- Overall, the model is highly accurate in distinguishing No Depression, Mild, Moderate, and Severe cases.

6.4 End-to-End System Evaluation

The entire two-stage system was thoroughly evaluated. If the risk identification (Stage 1) and the severity classification (Stage 2, if applicable) were correct, then the individual was labeled as "correctly classified." As shown in Figure 8, the proposed model achieved 98.21% accuracy, 98.83% precision, 97.47% recall, 97.15% F1-score, and 98.34% AUC. As shown in Table 5, the integrated framework performed better than the baselines with a high macro-F1 score, a very low false negative rate, and accuracy greater than 95%.



(a) (b)

Figure 8. Performance Metrics of the Final Model— (a) Accuracy vs. Epochs, (b) Loss vs. Epochs

Table 5. Comparative Performance of Depression Detection Models

DDD. Study / Model	EEE. Dataset Type	FFF. Accuracy (%)	GGG. Citation
HHH. Textual CNN	III. DAIC-WoZ (Text)	JJJ. 92	KKK. 24
LLL. Audio CNN	MMM. DAIC-WoZ (Audio)	NNN. 98	OOO. 24
PPP. Bi-LSTM (Hybrid Audio + Text)	QQQ. DAIC-WoZ (Audio + Text)	RRR. 88 (Train), 78 (Val)	SSS. 24
TTT. CNN–SVM Hybrid (Audio-based)	UUU. DAIC-WoZ (Audio)	VVV. 68	WWW. 23
XXX. CNN (Audio-only baseline)	YYY. DAIC-WoZ (Audio)	ZZZ. 58.57	AAAA. 23
BBBB. Proposed Model	CCCC. Real-Time Dataset	DDDD. 98.21	EEEE. —

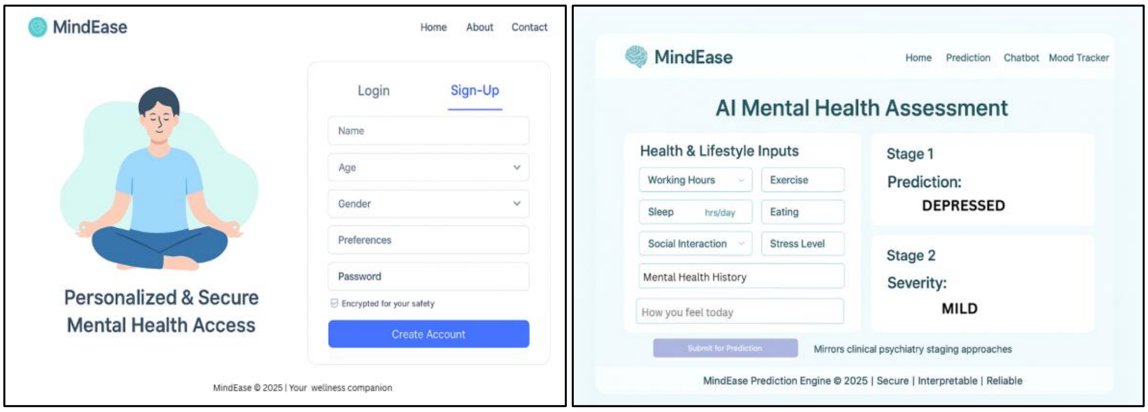
6.5 Ablation Study (Key Results)

- Full Agentic-TabDeCap model: Best results with an accuracy of 98.21%, precision of 98.83%, recall of 97.47%, F1-score of 97.15% and an AUC value equal to 98.34% which proved that the full three-stage model achieved good results.

- Without TabNet: The accuracy reduced to 65.17% -Early screening of non-depressed and mildly depressed patients is helpful.
- Sensitivity analysis revealed that Stage-1 accuracy was robust in the range of g values from 1.2 to 1.8. Low or very high g values decreased the accuracy, which supported that moderate sparsity is an optimal way to integrate feature reuse and interpretability.
- No Capsule Network: The accuracy is 96.32%, which proves that modeling with a capsule network is useful to categorize different severity levels of depression.
- For moderate and severe depression, for all three settings (medium recall), when positing the trained Capsule Network into place we can enhance the results of recall significantly while increasing OM-SDSM (overall severity classification result) by ~1.8% without DeBERTa-v3.
- Without confidence guided attention: For the samples that the model is not confident about (i.e low uncertainty), the accuracy reduced to 97.007%, suggesting that confidence-weighting would lead to clearer predictions.
- ` Text-only model: Shows that text data is helpful but not enough, Reaching 94.87% accuracy.
- Tabular only model: 92.45% accuracy was achieved, showing that only tabular data is not sufficient for a correct assessment of the level of depression.
- Overall Understanding: One of the best and most reliable results occurs when tabular data fused in conjunction with text and with confidence-aware fusion.

6.6 Agentic Mental Health Assistant Evaluation

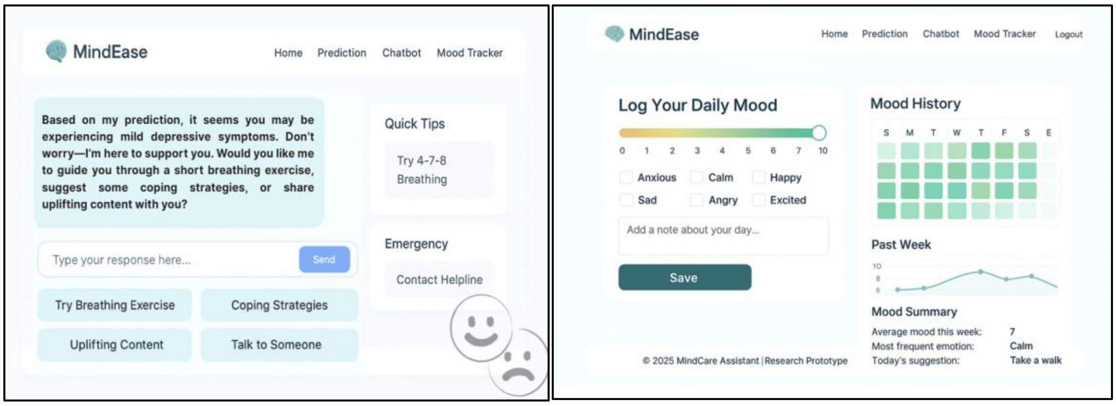
The model was validated and rated by four clinical psychologists using a 5-point scale. The model responded to 150 anonymous queries, with a response quality of 4.72/5, personalization of 4.61/5, and safety of 4.89/5. On average, the model scored 4.75/5, which is significantly better than the baseline GPT-like chatbot (3.92/5). Figure 9 shows the dashboard of the proposed system.



Stage 1
Prediction: DEPRESSED

Stage 2
Severity: MILD

(a) (b)



Based on my prediction, it seems you may be experiencing mild depressive symptoms. Don't worry—I'm here to support you. Would you like me to guide you through a short breathing exercise, suggest some coping strategies, or share uplifting content with you?

Average mood this week:	7
Most frequent emotion:	Calm
Today's suggestion:	Take a walk

(c) (d)

Figure 9. The proposed MindEase Assistance System and its Several Pages are Shown: a) Login, b) Prediction, c) Chatbot, d) Mood Tracker

6.7 Key Observations and Inferences

Experimental results show that the detection value of depression is calculated after combining structured lifestyle information and semantic text features. Early risk identification is achieved through interpretable feature selection using TabNet. The DeBERTa-CapsNet model captures subtle linguistic markers of severity. Robustness is enhanced by confidence-guided fusion during borderline situations. The findings lend credence to the notion that multi-stage and multimodal reasoning results in improved mental health prediction accuracy. Through careful hyperparameter tuning, capsule-based hierarchical learning, confidence-guided fusion, and multimodal feature integration, overall accuracy was improved. The contribution of each part leads to a prediction system that is reliable and efficient.

7. Conclusion and Future Scope

The proposed work detects early depression and assesses severity. The model combines clinical data and text analysis using multiple AI models to improve accuracy and reduce errors between stages. The proposed model achieves high performance with an accuracy of 98.21% and an AUC of 98.34%. The work also provides an AI assistant that improves real-time supportive responses to user interactions. The prototype of the proposed webpage demonstrates practical use. This approach can be applied in schools, colleges, workplaces, and hospitals since the system is accurate, reliable, and easy to use. In the future, the work can be developed by adopting more data types, supporting multiple languages, and enabling continuous mental health monitoring. Additionally, improvements could include mobile app integration for wider access and stronger privacy and data security measures to protect user information.

References

[1] Pradnyana, Gede Aditra, Wiwik Anggraeni, Eko Mulyanto Yuniarno, and Mauridhi Hery Purnomo. "Revealing Depression Through Social Media via Adaptive Gated Cross-Modal Fusion Augmented with Insights from Personality Traits." IEEE Access (2025).

[2] Joshi, Shantanu M., Hana R. Shaik, Shivam Rai Sharma, Philip Strong, Uma Srivatsa, Imo Ebong, Hyoyoung Jeong, Chen-Nee Chuah, and Lihong Mo. "Linking Electrocardiogram and Echocardiogram: Comparing Classical Machine Learning and

Deep Learning Neural Networks for the Detection of Regional Wall Motion Abnormalities." IEEE Access (2025).

[3] Martis, Gavryel, and Ryan McConville. "Federated Mental Wellbeing Assessment Using Smartphone Sensors Under Unreliable Participation." IEEE Access (2025).

[4] Li, Qiong, Xiaotong Liu, Xuecai Hu, Md Atiqur Rahman Ahad, Min Ren, Li Yao, and Yongzhen Huang. "Machine Learning-Based Prediction of Depressive Disorders Via Various Data Modalities: A Survey." IEEE/CAA Journal of Automatica Sinica 12, no. 7 (2025): 1320-1349.

[5] Galanina, Marina, Anna Rekiel, Anna Bączyk, and Bozena Kostek. "Depression Analysis and Detection Using Machine Learning: Incorporating Gender Differences in a Comparative Study." IEEE Access (2025).

[6] Dharma, Eddy Muntina, Harjanto Prabowo, Agung Trisetyarso, and Tjhin Wiguna. "CapsuleThermNet: A CNN-CapsuleNet Architecture for Early Detection of Suicide Risk in Depressed Patients Using Thermal Facial Imaging." IEEE Access (2025).

[7] Chen, Xin, Shihan Guan, Yici Liu, Zidong Liu, Qiang Chi, Regine Le Bouquin Jeannes, Jean-Louis Coatrieux, and Huazhong Shu. "GNMixer: A High-Density EEG Signal Processing Method Using P 2 GNN and MLP-Mixer for Depression Detection." IEEE Sensors Journal (2025).

[8] Liang, Zhen, Weishan Ye, Qile Liu, Li Zhang, Gan Huang, and Yongjie Zhou. "NSSI-Net: A Multi-Concept GAN for Non-Suicidal Self-Injury Detection Using High-Dimensional EEG in a Semi-Supervised Framework." IEEE Journal of Biomedical and Health Informatics (2025).

[9] Hossain, Md Mithun, Md Shakil Hossain, Sudipto Chaki, Md Saifur Rahman, and ABM Shawkat Ali. "Revolutionizing Mental Health Sentiment Analysis with Bert-Fuse: A Hybrid Deep Learning Model." IEEE Access (2025).

[10] Zhou, Longhui, Bin Hu, and Zhi-Hong Guan. "MDRA: A Multimodal Depression Risk Assessment Model Using Audio and Text." IEEE Signal Processing Letters (2025).

[11] AS, Kavitha Bai, and J. Somasekar. "Early Prediction of Heart Disease in Diabetic Patients using TabNet's Dynamic Attention Mechanism." In 2025 International Conference on Machine Intelligence and Smart Innovation (ICMISI), pp. 393-396. IEEE, 2025.

[12] Asal, Burçak. "Interpretable Deep Learning for Pulsar Star Classification with Explainable AI Techniques: A Comparative Analysis of TabNet and FT-Transformer Models." In 2025 33rd Signal Processing and Communications Applications Conference (SIU), IEEE, 2025, 1-4.

[13] Anand, Ankit, Syed Wali Ahmad Rizvi, Sanjiith Ravindhran, Rishi Ajith, Gokula Kumaran KG, and Divij Goyal. "Breaking Down Barriers: Next-Generation Techniques for Segmenting Medical Abstract Text using DeBERTa-V3." In 2024 4th Asian Conference on Innovation in Technology (ASIANCON), IEEE, 2024, 1-5.

[14] Liao, Xiao, Wei Cui, Min Zhang, Aiwu Zhang, and Pan Hu. "Optimized Two-Stage Anomaly Detection and Recovery in Smart Grid Data Using Enhanced DeBERTa-v3 Verification System." Sensors 25, no. 13 (2025): 4208.

[15] Lei, Jierui, Qingyi Yang, Bo Li, and Wenjian Zhang. "Ccfn: Depression Detection Via Multimodal Fusion with Complex-Valued Capsule Network." In 2024 International Joint Conference on Neural Networks (IJCNN), IEEE, 2024, 1-6.

[16] Hindi, Mahd, Linda Mohammed, Ommama Maaz, and Abdulmalik Alwarafy. "Enhancing the Precision and Interpretability of Retrieval-Augmented Generation (Rag) In Legal Technology: A Survey." IEEE Access (2025).

[17] Song, Minchae. "Enhancing Rag Performance by Representing Hierarchical Nodes in Headers for Tabular Data." IEEE Access (2025).

[18] Singh, Nongmeikapam Thoiba, Harkamal Kaur, Jyoti Dhiman, Ayush Aryan, Jyoti Rani, and Manoj Wadhwa. "AI-Driven Document Analysis: Employing Streamlit, Faiss, Nvidia Nemo." In 2025 3rd International Conference on Inventive Computing and Informatics (ICICI), IEEE, 2025, 314-322.

[19] Gayathri, S., P. Synthan, K. Vishnu Karthikeyan, and V. Gokulnath. "An AI-Powered Chatbot for Advanced Scan Interpretation of Brain Tumors, Pneumonia, and Lung Cancer." In 2025 3rd International Conference on Artificial Intelligence and Machine Learning Applications Theme: Healthcare and Internet of Things (AIMLA), IEEE, 2025, 1-6.

[20] Milišić, Dane, Dinu Dragan, Dušan Gajić, and Veljko Petrović. "Visualization of Neural Networks Training in Real-Time: A Web Based Example." In 2025 12th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN), IEEE, 2025, 1-5.

[21] Kamra, Vikas, Tanishq Saini, Nishant Shishodia, Manish Singh, and Pratyaksh Sehrawat. "A Novel Approach Towards Full Stack Comprehensive Expense Monitoring System." In 2025 International Conference on Next Generation Information System Engineering (NGISE), vol. 1, IEEE, 2025, 1-5.

[22] Culjak, Gordana. "Access, Awareness and Use of Internet Self-Help Websites for Depression in University Students." In 2012 45th Hawaii International Conference on System Sciences, IEEE, 2012, 2655-2664.

[23] Saidi, Afef, Slim Ben Othman, and Slim Ben Saoud. "Hybrid CNN-SVM Classifier for Efficient Depression Detection System." In 2020 4th international conference on advanced systems and emergent technologies (IC_ASET), IEEE, 2020, 229-234.

[24] Marriwala, Nikhil, and Deepti Chaudhary. "A Hybrid Model for Depression Detection Using Deep Learning." Measurement: Sensors 25 (2023): 100587.