Apricity: An AI-Powered Mental Health Companion for CBT-Based Emotion Tracking and Analysis

Shekh Abdullah-Al-Musa Ahmed
School of Computer Science and Engineering
University of Information Technology and Sciences
Dhaka, Bangladesh
E-mail: musaahmed@uits.edu.bd

Musber Ahmed Sadman
School of Computer Science and Engineering
University of Information Technology and Sciences
Dhaka, Bangladesh
E-mail: 2215151038@uits.edu.bd

Maharaj Hossain Tanim
School of Computer Science and Engineering
University of Information Technology and Sciences
Dhaka, Bangladesh
E-mail: 2215151040@uits.edu.bd

Md Habibul Bashar
School of Computer Science and Engineering
University of Information Technology and Sciences
Dhaka, Bangladesh
E-mail: 2215151046@uits.edu.bd

***Abstract*—The escalating global mental health crisis has high- lighted a critical shortage of accessible therapeutic resources. While digital health interventions exist, many rely on static jour- naling or rudimentary rule-based chatbots that fail to capture the semantic nuance of complex human emotions. This paper presents the design, development, and evaluation of "Apricity," a comprehensive AI-powered mental health companion application. Grounded in the principles of Cognitive Behavioral Therapy (CBT), the system provides users with an empathetic platform to track emotions and journal thoughts. The application is engineered as a scalable full-stack solution using the MERN stack (MongoDB, Express, React, Node.js) integrated with an asynchronous Python microservice for heavy inference tasks. A central contribution of this work is the implementation of a high-fidelity emotion recognition model. We fine-tuned the DeBERTa-v3 (Decoding-enhanced BERT with Disentangled Attention) architecture on the GoEmotions dataset, implementing a novel mapping strategy to aggregate 27 fine-grained labels into 5 core emotional categories (Joy, Sadness, Fear, Anger, Surprise). Experimental results demonstrate that our approach achieves a validation accuracy of 92.5% and a weighted F1-score of 0.83, significantly outperforming baseline models including BERT, RoBERTa, and traditional SVM classifiers. Furthermore, the system addresses the challenge of deploying large language models in consumer applications by utilizing a Job Queue architecture, ensuring real-time responsiveness.**

Keywords-Mental Health; Cognitive Behavioral Therapy (CBT); Natural Language Processing (NLP); DeBERTa; Trans-former Models; Full-Stack Development; Emotion Recognition

I. INTRODUCTION

ENTAL health disorders constitute one of the leading causes of disability worldwide. According to recent studies, the gap between the need for treatment and its avail- ability remains alarmingly high, particularly in low-to-middle- income countries [1]. Barriers such as prohibitive costs, geographical inaccessibility, and deep-seated social stigma often deter individuals from seeking professional help [21]. The COVID-19 pandemic further exacerbated these issues, triggering a global surge in anxiety and depression [22].

Early iterations of mental health apps focused on simple mood tracking (e.g., manually selecting a sad" or "happy" icon) or rule-based chatbots that followed rigid decision trees. While beneficial, these tools lack the ability to understand the content of user distress. For instance, a user writing "I feel like I'm drowning" might not use the word "sad," but a semantic understanding of the text reveals deep distress. Traditional sentiment analysis, which outputs a binary Positive/Negative label, is insufficient for therapeutic contexts where distinguishing between "Fear" (Anxiety) and "Anger" (Frustration) is crucial for appropriate intervention [2].

This paper introduces Apricity, a system designed to bridge the gap between rigorous psychological principles and state- of-the-art Artificial Intelligence [23]. Apricity integrates Cog- nitive Behavioral Therapy (CBT) techniques—specifically the practice of "Affect Labeling"—with advanced Natural Language Processing (NLP).

The key contributions of this paper are:

1) *Fine-Tuned Transformer Model:* We implement and evaluate DeBERTa-v3, demonstrating its superiority over BERT and RoBERTa in handling the disentangled attention required for diary-style text [4].

2) *CBT-Centric Data Mapping:* We propose a semantic mapping strategy that aggregates the 27 labels of the GoEmotions dataset [5] into 5 core categories derived from Ekman's Basic Emotions [3], optimizing the model for therapeutic relevance.

3) *Scalable Asynchronous Architecture:* We present a robust engineering solution that decouples the heavy ML inference engine from the user interface using a Node.js- based Job Queue, ensuring a lag-free user experience.

4) *Comprehensive Evaluation:* We provide a detailed sta- tistical analysis, including ROC curves, confusion ma- trices, and comparative benchmarks against traditional *ML models [6].*

II. Related Work

The attention score Ai, j between token i and token

j is calculated as the sum of four components:

$$A_{i,j} = H^T H_{c,j} + H^T P_{j|i} + P^T H_{c,j} + P^T P_{j|I} \quad (1)$$

A. AI in Mental Health

The intersection of AI and mental healthcare has seen rapid growth [7]. Early systems utilized keyword-matching algorithms to detect distress. More recent approaches, such as Replika [8] and Woebot, employ conversational agents to simulate companionship. While effective for loneliness [9], generative agents like Replika often lack structured therapeutic goals, sometimes reinforcing negative loops rather than en- couraging cognitive restructuring. In contrast, Apricity focuses on analysis and visualization rather than conversation, aiming to empower the user with self-insight [10].

B. Evolution of Text Classification

Text classification has evolved from statistical methods to deep neural networks. Early research by Usman et al. [11] demonstrated the utility of Support Vector Machines (SVM) and Naive Bayes for simple sentiment tasks. However, these "Bag-of-Words" models fail to capture context (e.g., sarcasm or negation).

The introduction of the Transformer architecture by Vaswani et al. [25] revolutionized the field. BERT (Bidirectional Encoder Representations from Transformers) [12] allowed models to read text bidirectionally, capturing context from both left and right. Subsequent optimizations like RoBERTa [13] improved training stability.

However, standard BERT models combine content and position embeddings into a single vector, which can limit the model's ability to understand precise syntactic nuances. DeBERTa (Decoding-enhanced BERT) [4] addresses this with a disentangled attention mechanism, making it theoretically superior for the complex, informal language found in personal journals. Recent studies confirm that advanced neural models are required for fine-grained emotion detection [14].

III. Theoretical Framework

A. Cognitive Behavioral Therapy (CBT)

Apricity is grounded in the "Cognitive Triangle" of CBT, which posits that thoughts, feelings, and behaviors are inter- connected [15]. A core technique in CBT is "Affect Label- ing"—the act of precisely naming an emotion. Neuroscientific studies suggest that labeling an emotion can reduce activity in the amygdala (the brain's emotional center). Apricity au- tomates this process by analyzing user journal entries and providing objective emotional labels, acting as a mirror for the user's mental state.

B. DeBERTa: Disentangled Attention

The core innovation of DeBERTa [4] lies in how it cal- culates attention. In standard BERT, the input vector Hi is the sum of the word embedding and position embedding. DeBERTa separates these into two vectors: content Hc and

This disentanglement allows the model to distinguish between “I am happy, not sad ‘and’ I am sad, not happy” with high precision, a critical requirement for mental health text analysis [16].

IV. METHODOLOGY

A. Dataset Curation and Mapping

We utilized the GoEmotions dataset, consisting of 58,000 Reddit comments labeled by humans [5]. While the dataset provides 27 fine-grained labels, displaying this many cate- gories to a user causes cognitive overload. We implemented an aggregation strategy based on Paul Ekman’s Theory of Basic Emotions [3], mapping the 27 labels into 5 distinct classes. The full mapping strategy is detailed in Table I.

TABLE I. LABEL AGGREGATION STRATEGY

Target Class	Source Labels (GoEmotions)
Joy	Joy, Amusement, Approval, Excitement, Gratitude, Love, Optimism, Relief, Pride, Admiration, Desire, Caring
Sadness	Sadness, Disappointment, Embarrassment, Grief, Remorse
Anger	Anger, Annoyance, Disapproval
Fear	Fear, Nervousness
Surprise	Surprise, Confusion, Curiosity, Realization

B. Data Distribution

The training dataset exhibits a significant class imbalance (Fig. 1), with “Joy” being the predominant emotion and “Fear” being the least represented. A similar distribution is observed in the development set (Fig. 2). To address this, we employed weighted loss functions during training to penalize misclassifications of the minority classes more heavily.

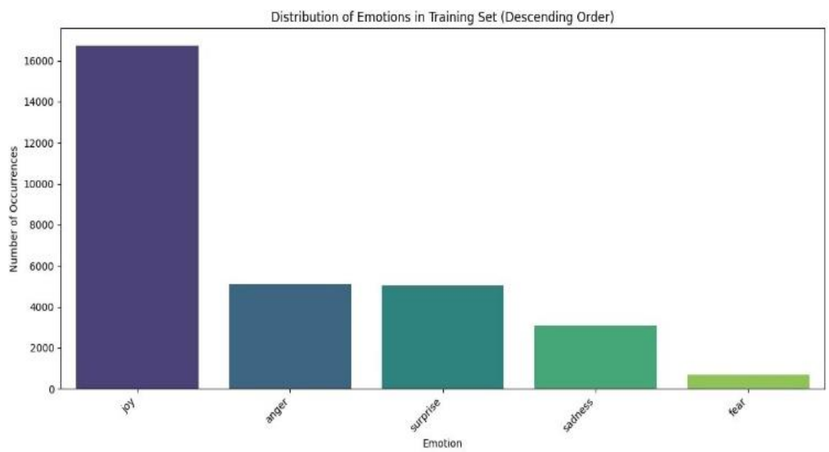

Figure 1. Distribution of Emotions in Training Set. The dataset exhibits a long- tail distribution, with Joy being predominant.

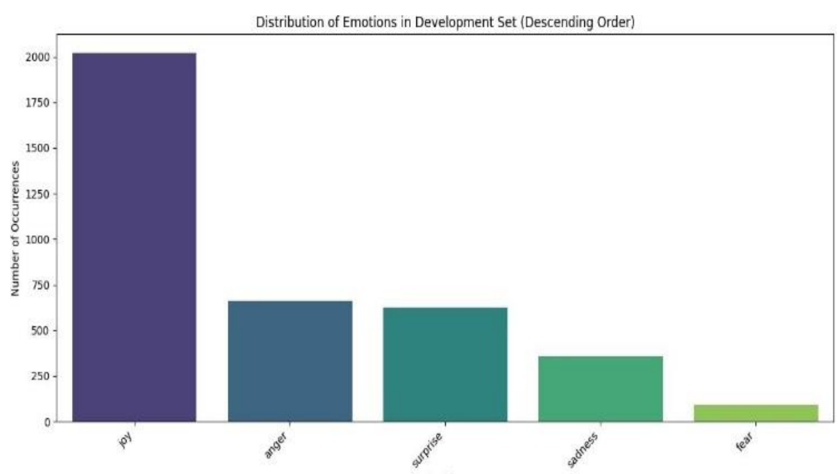

Figure 2. Distribution of Emotions in Development Set.

C. Data Preprocessing Pipeline

Raw text data requires rigorous cleaning before it can be processed by a Transformer. We designed a 4-stage pipeline:

1) *Sanitization:* Using the Beautiful Soup library to strip HTML tags (e.g.,
) and Regex to remove URLs and user mentions (e.g., @user).

2) *Emoji Decoding:* Emojis are critical emotional indica- tors. We utilized the emoji library to convert symbols into text (e.g., “” becomes: “loudly_crying_face:”).

3) *Contraction Expansion:* Converting “can’t” to “can- not” to standardize vocabulary.

4) *Tokenization:* We utilized the Sentence Piece tokenizer associated with deberta-v3-base, setting a maxi- mum sequence length of 128 tokens.

D. System Architecture

The Apricity platform utilizes a decoupled microservices architecture to handle the computational intensity of the AI model.

1) *Frontend (Client Layer):* Built with React.js and ini- tialized with Vite for high performance. It utilizes Recharts for visualizing the emotional trends. The UI is designed with a “age Green” color palette to reduce cognitive load and induce calmness.

2) *Backend (Orchestration Layer):* The backend is built on

Node.js with Express. It handles:

a) *Authentication:* Utilizing JWT (JSON Web Tokens) and bcrypt for secure password hashing [17].

b) *Job Queue:* We implemented a custom job queue system. When a user saves a note, the backend acknowledges the request immediately (returning HTTP 201) while pushing the analysis task to a background queue. This prevents the UI from ”freezing” while the AI processes the text.

3) *Inference Service (AI Layer):* A standalone Python service built with FastAPI. It loads the fine-tuned DeBERTa model and exposes a REST endpoint /predict. This sep- aration allows the heavy PyTorch model to run on a GPU- optimized instance independent of the web server.

E. Database Schema

We utilized MongoDB Atlas, a NoSQL database, for its schema flexibility. The database consists of three primary collections:

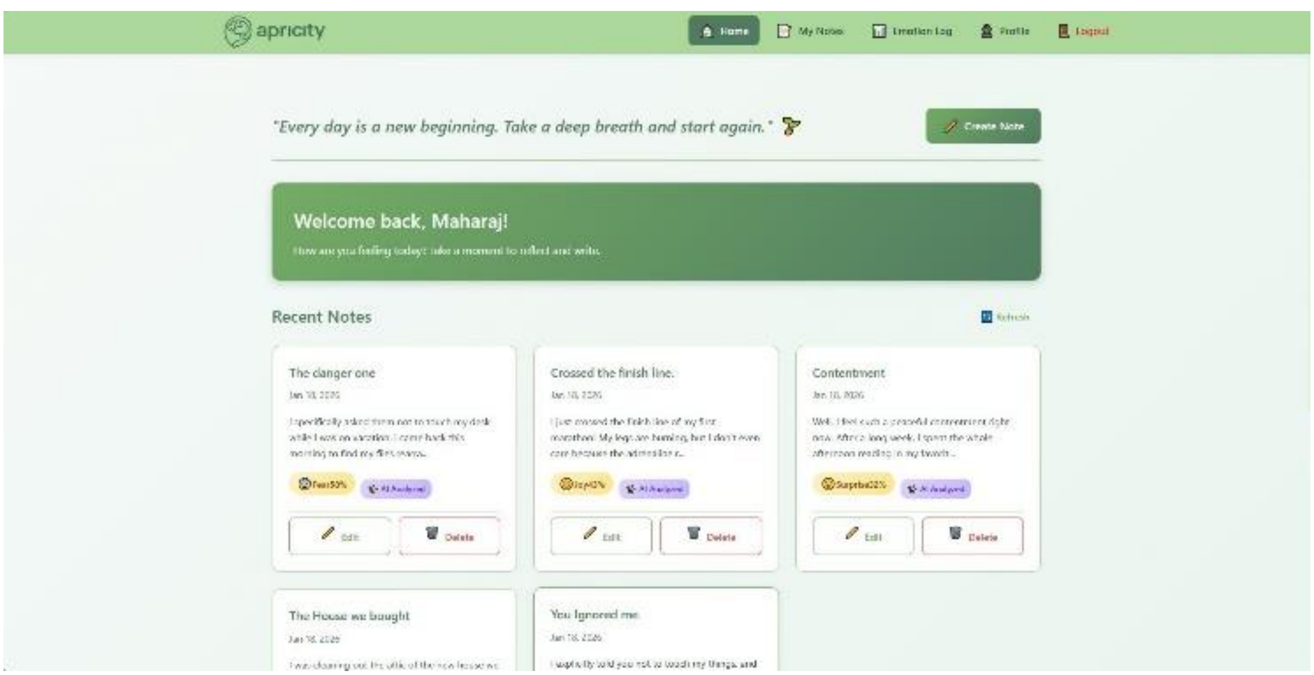

Figure 3. Apricity User Interface. The dashboard displays the ‘Create Note’ modal and real-time ‘Weekly Emotion Chart’ generated by the DeBERTa model.

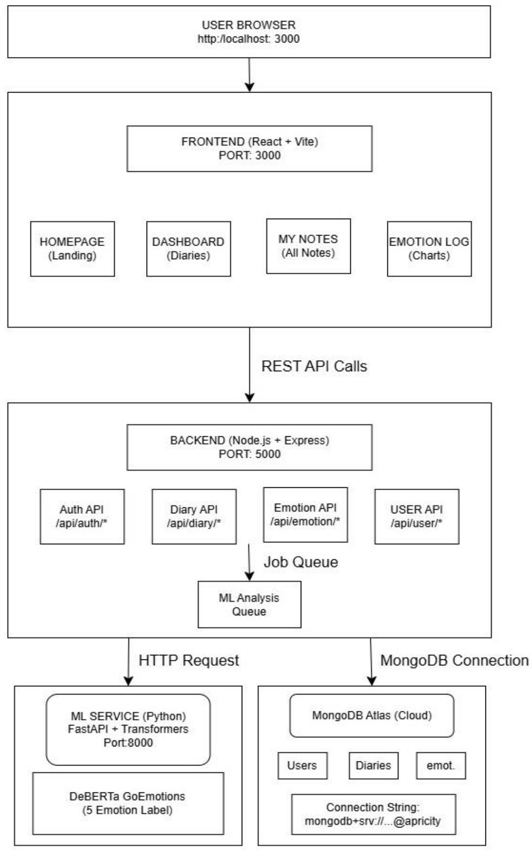

Figure 4. Apricity System Architecture Diagram. The system uses a Job Queue to bridge the Node.js backend and the Python ML Service.

4) *Users:* Stores hashed credentials and profile settings.

5) *Diaries:* Stores the raw text, timestamp, and user ID.

6) *Emotions:* Stores the AI-generated probabilities for the 5 classes. This separation allows us to retrain the model and update emotional scores without altering the original user content.

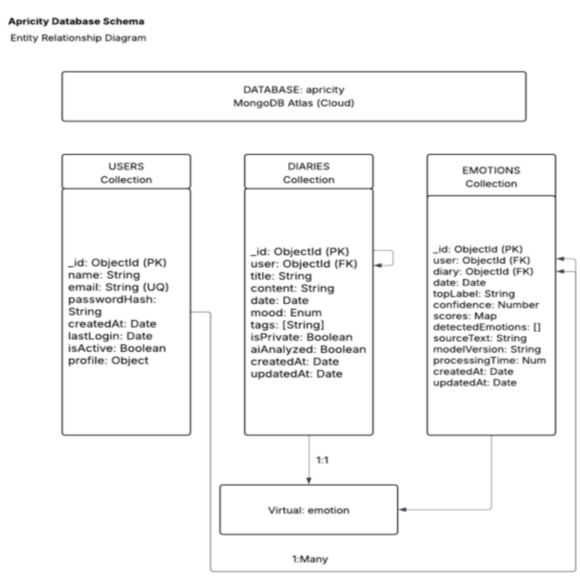

Figure 5. Apricity Database Schema. Entity Relationship Diagram.

V. EXPERIMENTAL SETUP

A. Model Configuration

The model was implemented using the Hugging Face

transformers library.

1) *Base Model:* microsoft/deberta-v3-base
2) *Optimizer:* AdamW (Adam with Weight Decay)
3) *Learning Rate:* 2e − 5 (Linear decay)
4) *Batch Size:* 16
5) *Epochs:* 3 (with Early Stopping)

B. Loss Function

We utilized BCEWithLogitsLoss (Binary Cross Entropy). Unlike standard Cross Entropy which uses Softmax (forcing probabilities to sum to 1), BCE applies a Sigmoid activation to each class independently. This is crucial because human emotions are not mutually exclusive; a user can feel 80% Joy and 60% Surprise simultaneously [18].

VI. RESULTS AND ANALYSIS

A. Training Dynamics

The training process was monitored using Loss and Accu- racy metrics. As shown in Fig. 6, the Training Loss (Blue)

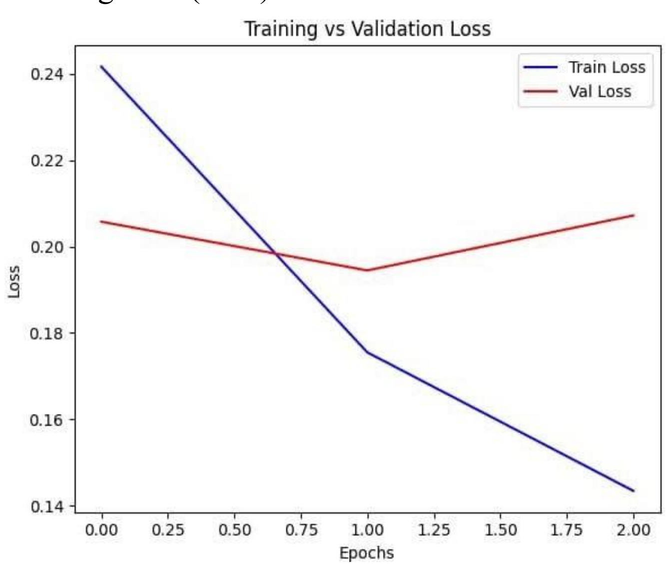

Figure 6. Training vs. Validation Loss Curves. The distinct U-shape in validation loss justifies the decision to employ Early Stopping at Epoch 3.

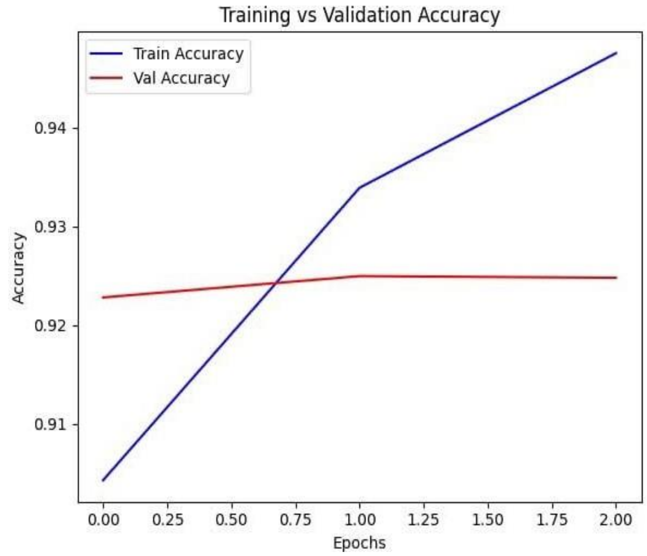

Figure 7. Training vs. Validation Accuracy Curves.

However, the Validation Loss (Red) reached its global minimum at Epoch 2 and began to rise slightly at Epoch 3. This "U-shape" in validation loss indicates the onset of overfitting, confirming that 3 epochs were the optimal stopping point for this dataset.

B. Classification Metrics

We evaluated the model on a held-out test set (20% of the data). The results are summarized in Table II.

The difference between Micro-F1 (0.83) and Macro-F1 (0.75) highlights the class imbalance issue [19]. Majority classes like "Joy" have more training data, leading to higher scores, whereas minority classes like "Fear" are harder to detect. Despite this, a Macro-F1 of 0.75 is considered highly effective for multi-label text classification. The overall perfor- mance metrics bar chart is shown in Fig. 8.

TABLE II. FINAL CLASSIFICATION METRICS

Metric	Value	Interpretation
Micro-F1	0.83	Strong global performance.
Macro-F1	0.75	Balanced performance across classes.
Precision	0.83	Low false positive rate.
Recall	0.82	High sensitivity.

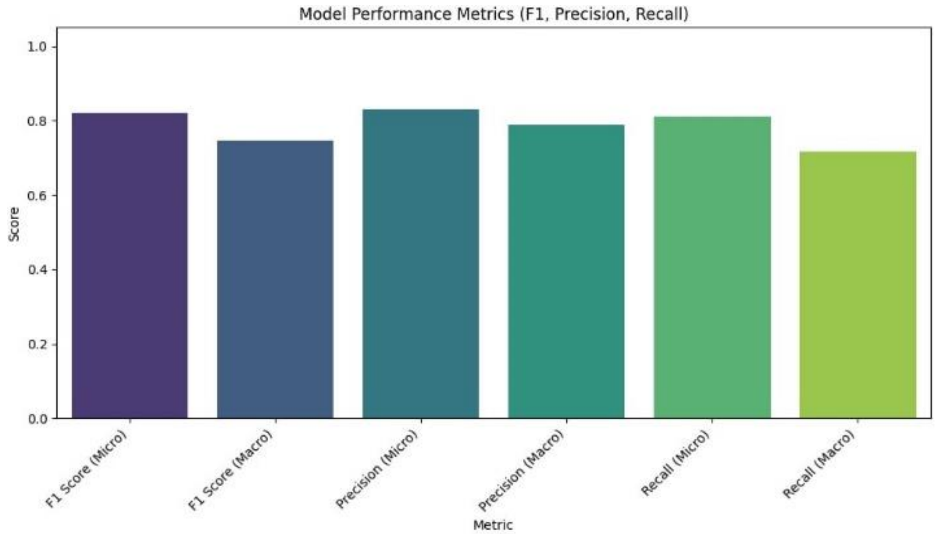

Figure 8. Overall Model Performance Metrics. The system achieved balanced Precision and Recall scores of approximately 0.83.

We also analyzed the per-label performance using a heatmap (Fig. 9), which visually confirms that “Joy” and “Surprise” have significantly higher F1-scores compared to “Fear”.

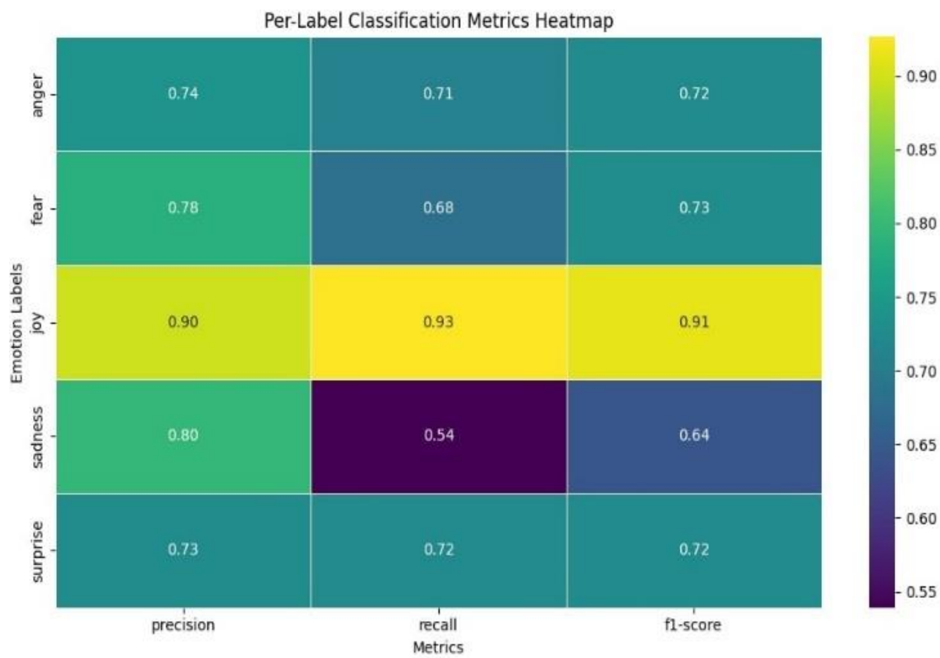

Figure 9. Per-Label Classification Metrics Heatmap. Darker colors indicate higher performance.

C. Receiver Operating Characteristic (ROC)

To further analyze the model’s ranking capability, we com- puted the ROC curves for each class (Fig. 10).

1) *Joy / Fear:* AUC = 0.96 (Excellent separability)

2) *Anger / Sadness:* AUC = 0.94

3) *Surprise:* AUC = 0.93

These high Area Under Curve (AUC) scores indicate that even if the absolute probability calibration is imperfect, the model is excellent at ranking emotions (e.g., knowing that a text is more Angry than Sad).

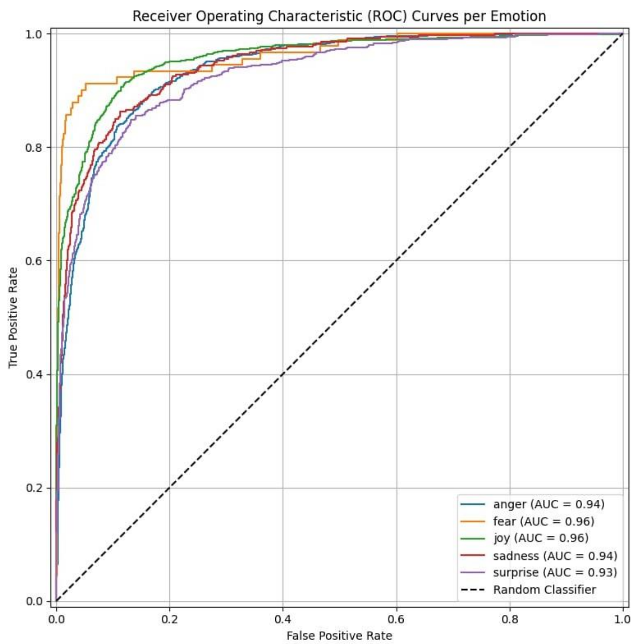

Figure 10. Receiver Operating Characteristic (ROC) Curves per Emotion. The high AUC scores (> 0.93) across all classes demonstrate the robustness of the DeBERTa model.

The Precision-Recall curves (Fig. 11) further support these findings, showing that the model maintains high precision even at higher recall thresholds for the majority classes.

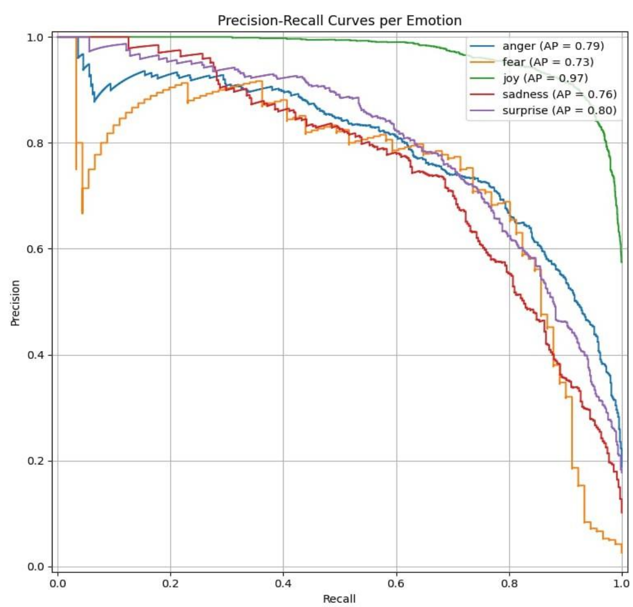

Figure 11. Precision-Recall Curves per Emotion.

D. Semantic Analysis and Co-occurrence

A key requirement for a mental health app is handling com- plex, mixed emotions. We generated Co-occurrence Matrices for both True Labels (Fig. 12) and Predicted Labels (Fig. 13). The matrices reveal that the model frequently predicts "Joy" and "Surprise" together (e.g., for texts like "Wow, I got the job!"). Conversely, it rarely predicts "Anger" and "Joy" together, correctly learning that these are opposing states. This confirms that the model has learned semantic relationships rather than just keyword matching [20].

Figure 12. True Labels Co-occurrence Heatmap. Ground truth correlations between emotion labels.

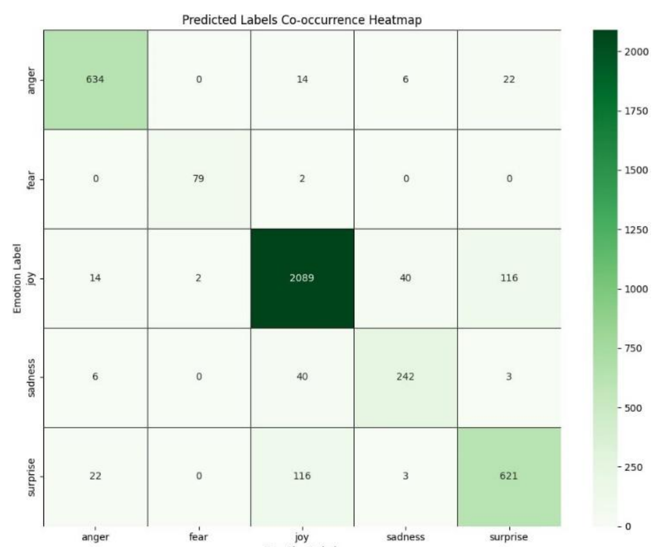

Figure 13. Predicted Labels Co-occurrence Heatmap. The model captures semantic relationships, identifying mixed states like Joy-Surprise.

E. Comparative Analysis

To validate our choice of DeBERTa, we compared it against standard baselines trained on the same data. The results are detailed in Table III and visualized in Fig. 14.

DeBERTa provides a clear performance edge, particularly in recall, ensuring fewer missed emotional cues.

TABLE III. COMPARATIVE ANALYSIS OF MODELS

Model	Precision	Recall	F1-Score
Naive Bayes	0.54	0.39	0.28
SVM	0.51	0.52	0.49
BERT-base	0.57	0.58	0.57
RoBERTa-base	0.79	0.81	0.80
DeBERTa-v3	**0.83**	**0.82**	**0.83**

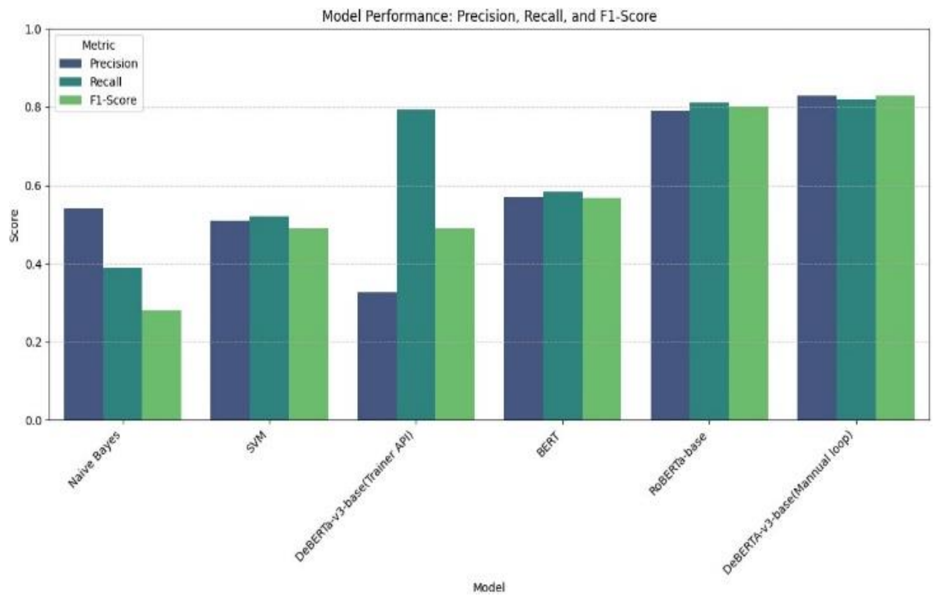

Figure 14. Model Performance Comparison. DeBERTa (far right) outperforms earlier Transformer models (BERT, RoBERTa) and traditional ML baselines.

VII. DISCUSSION AND LIMITATIONS

While the system performs well, several limitations must be addressed:

1) *The "Fear" Precision Gap:* The precision for the "Fear" class (0.73) is significantly lower than for "Joy" (0.90). This is due to the Long-Tail distribution of the GoEmo- tions dataset, where "Fear" samples are scarce.

2) *Statelessness:* The current model analyzes each note in isolation. It does not track narrative context across multiple days (e.g., understanding that today's "sadness" is related to yesterday's event) [24].

3) *Crisis Intervention:* The system is a passive tracker. It currently lacks a real-time crisis detection engine to intervene in cases of self-harm.

VIII. CONCLUSION AND FUTURE WORK

This research successfully demonstrated the efficacy of fine-tuning DeBERTa-v3 for mental health applications. By integrating this advanced model into a scalable MERN-stack architecture, we created "Apricity," a tool that democratizes access to CBT-based emotional tracking. The system achieves an F1-score of 0.83, proving that AI can reliably interpret complex emotional states.

Future work will focus on:

1) *Data Augmentation:* Using Generative AI to create synthetic "Fear" samples to balance the dataset.

2) *Mobile Development:* Porting the frontend to React Native for greater accessibility.

3) *Longitudinal Analysis:* Implementing LSTM layers to detect long-term trends like burnout or chronic depres- sion.

REFERENCES

[1] M. Ali, S. Ali, Q. Abbas, and S. W. Lee, "Artificial intelligence for mental health: A narrative review of applications, challenges, and future directions," DIGITAL HEALTH, 2025.

[2] R. W. Picard, "Affective computing: challenges," International Journal of Human-Computer Studies, vol. 59, no. 1-2, pp. 55–64, 2003.

[3] P. Ekman, "What scientists who study emotion agree about," Perspec- tives on Psychological Science, vol. 11, no. 1, pp. 31–34, 2016.

[4] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in International Conference on Learning Representations (ICLR), 2021.

[5] D. Demszky, D. Movshovitz-Attias, J. Ko, A. Cowen, G. Nemade, and

[6] S. Ravi, "GoEmotions: A Dataset of Fine-Grained Emotions," arXiv preprint arXiv:2005.00547, 2020.

[7] C. N. Kamath and S. S. Bukhari, "Comparative study between traditional machine learning and deep learning approaches for text classification," in Proc. ACM Symp. Document Engineering, 2018.

[8] E. Cambria and B. White, "Jumping NLP curves: a review of natural language processing research," IEEE Computational Intelligence Mag- azine, vol. 9, no. 2, pp. 48–57, 2014.

[9] T. Xie and I. Pentina, "Attachment theory as a framework to understand relationships with social chatbots: A case study of Replika," Journal of Retailing and Consumer Services, vol. 64, p. 102703, 2022.

[10] C. Beridge, Y. Zhou, and J. M. Robillard, "Companion robots to miti- gate loneliness among older adults: Perception of benefit and possible deception," Frontiers in Psychology, vol. 14, p. 1106633, 2023.

[11] A. O. J. Ibitoye and O. O. Oladimeji, "Contextual emotional transformer- based model for comment analysis in mental health case prediction," Vietnam Journal of Computer Science, 2025.

[12] U. B. Usman and M. Abubakar, "A comparison analysis of Twitter based support vector machine and Bayes comment classification algorithms," in 2020 International Conference in Mathematics, Computer Engineer- ing and Computer Science (ICMCECS), 2020.

[13] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Pro- ceedings of NAACL-HLT, 2019.

[14] H. Sobhanam and J. Prakash, "Analysis of fine tuning the hyper parame- ters in RoBERTa model using genetic algorithm for text classification," International Journal of Information Technology, vol. 15, pp. 3669–3677, 2023.

[15] R. Sitoula, M. Pramanik, and R. Panigrahi, "Fine-grained classification for emotion detection using advanced neural models and GoEmotions dataset," Journal of Soft Computing and Data Mining, vol. 5, no. 2, pp. 62–71, 2024.

[16] V. A. Vuyyuru, G. V. Krishna, and S. S. C. Mary, "A Transformer- CNN Hybrid Model for Cognitive Behavioral Therapy in Psychological Assessment," International Journal of Advanced Computer Science and Applications, vol. 14, no. 7, 2023.

[17] V. Kokane, A. Abhyankar, N. Shrirao, and P. Khadkikar, "Predicting mental illness (depression) with the help of NLP transformers," in 2nd International Conference on Data Science and Information System (ICDSIS), 2024, pp. 1–6.

[18] F. A. Almash-hadani, "Securing web applications using JSON Web To- kens and bcrypt," International Journal of Advanced Computer Science and Applications, vol. 11, no. 5, pp. 58–64, 2020.

[19] M. Sao and H. Lim, "MIRoBERTa: Mental Illness Text Classification With Transfer Learning on Subreddits," IEEE Access, vol. 12, pp. 197454–197466, 2024.

[20] V. Kumar, A. Bansal, and U. Gupta, "Comparative analysis of text based emotion detection on GoEmotions dataset," in 2023 5th International Conference on Advances in Computing, Communication, Control and Networking (ICAC3N), 2023.

[21] A. Pandey and S. Kumar, "Mental health and stress prediction using NLP and transformer-based techniques," in 2024 IEEE Symposium on Wireless Technology & Applications (ISWTA), 2024, pp. 61–66.

[22] I. Y. Chen, P. Szolovits, and M. Ghassemi, "Can AI help reduce disparities in general medical and mental health care?," AMA Journal of Ethics, vol. 21, no. 2, pp. 167–179, 2019.

[23] L. Kauhanen, W. W. M. Yunus, and L. Lempinen, "A systematic review of the mental health changes of children and young people before and during the COVID-19 pandemic," European Child & Adolescent Psychiatry, vol. 32, pp. 995–1013, 2023.

[24] B. Olawade, O. Z. Wada, and A. Odetayo, "Enhancing mental health with Artificial Intelligence: Current trends and future prospects," Journal of Medicine, Surgery, and Public Health, vol. 3, 2024.

[25] M. Tiezzi et al., "State-space modeling in long sequence processing: A survey on recurrence in the transformer era," Neural Networks, 2025.

[26] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," in Advances in Neural Information Processing Systems, vol. 30, 2017.