TRUST-MH: Transparent and Responsible User-level Semantic Tagging for Mental Health Assessment

Sumit Dalal[1] Sarika Jain[2]

[1]Bennett University (The Times Group), India [2]National Institute of Technology Kurukshetra, India

Introduction

Motivation

- Mental health is a foundational pillar of individual well-being and sustainable social progress.
- Millions suffer from depression, but timely detection and intervention remain limited due to stigma, lack of resources, and limited access to mental health professionals (MHPs).
- With the increasing ubiquity of social networks, digital expressions now provide an unprecedented opportunity to identify early signs of mental distress and deploy AI-powered recommender systems for social good.

Problem

- Current systems must balance predictive power with explainability and ethical responsibility.
- Lexicon-based approaches:
  - Interpretable but miss contextual cues.
- Deep learning models (BERT, RoBERTa, etc.):
  - Provide nuanced understanding but lack transparency.
  - Their predictions are difficult for clinicians or users to trust.
- This creates a critical gap: a need for transparent, clinically interpretable AI systems for mental health detection and support.

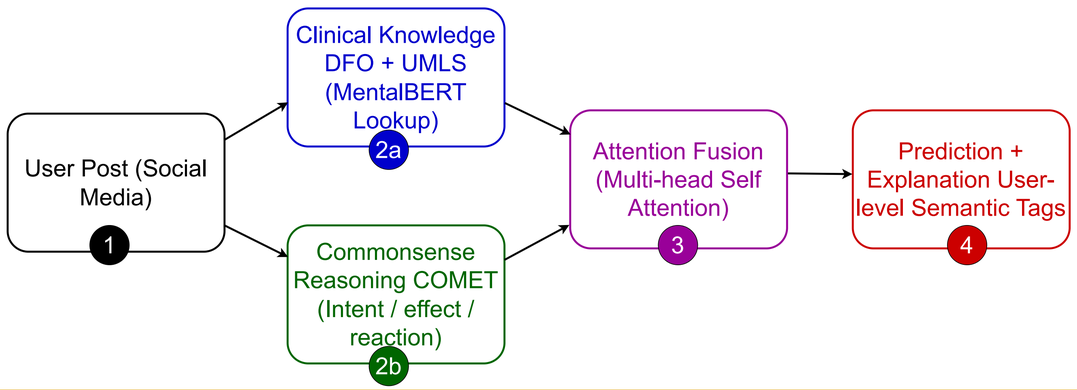
Architecture

Overview

- TRUST-MH = A knowledge-infused neural framework that combines:
  - Domain-specific knowledge (DepressionFeature Ontology, UMLS).
  - Commonsense reasoning (COMET).
  - Deep learning attention mechanisms.
- **Goal:** Produce transparent, clinically interpretable explanations at the user level.

Multi-level Knowledge Infusion

1. Clinical Phrase Tagging
   - User posts enriched with clinical phrases from:
     - DepressionFeature Ontology (DFO)
     - UMLS (Unified Medical Language System)
   - Contextual synonyms from MentalBERT if no direct match.
   - Enriched posts mapped to biomedical definitions and Concept IDs.
2. Commonsense Inference (COMET)
   - Extracts 5 aspects from each post:
     - Writer's intent, effect, reaction.
     - Effect and reaction on others.
   - Encodes social & emotional reasoning.
3. Attention Fusion
   - Clinical + commonsense representations jointly modeled.
   - Multi-head self-attention → balances domain knowledge with emotional cues.



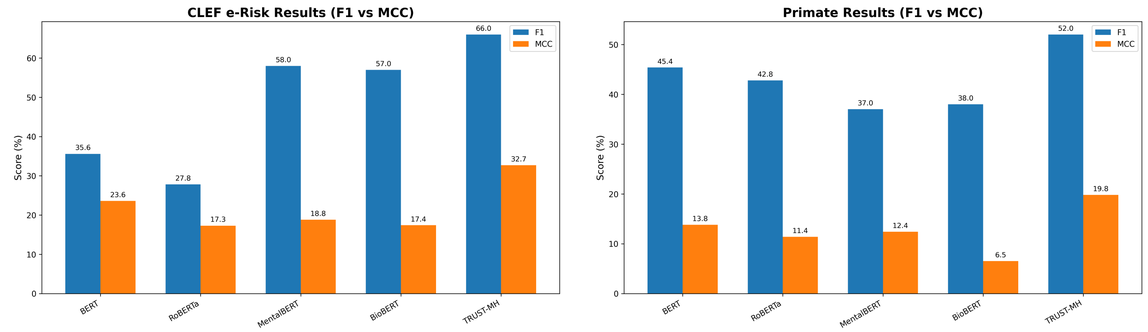
Architecture Contd.

Explanation Visualizer

- Uses attention weights to highlight clinically relevant phrases.
- Maps attention to ontology concepts for human-understandable explanation.
- Enhanced with GPT-3.5 prompts for natural language justifications.

Results & Evaluation

- **Datasets:** CLEF e-Risk (binary depression), PRIMATE (multi-label PHQ-9).
- **Metrics:** Precision, Recall, F1, MCC.
- **Key findings:**
  - TRUST-MH → +8% F1, +7% MCC over MentalBERT.
  - Qualitative explanations are richer & clinically grounded.



On CLEF e-Risk (binary depression detection) and PRIMATE (multi-label PHQ-9 symptoms), TRUST-MH consistently outperforms baselines in both F1 and MCC. The framework not only improves accuracy but also enhances reliability in mental health assessment.

So, I decided to go to psychiatrist, but now I feel like I'm not that bad? How bad is really bad? Maybe I just act dramatic I cry a lot without any specific reason, I want to kill myself quite often, but... I dont know if I really need therapy? Maybe its just who I am, just a lazy whiner? And when I will come to therapist I dont know what to say, what if that day I will feel fine and they will not believe me? I dont look like I struggle all the time, people around me don't even know that I have problems, I've always been a comedian in a company. I'm just a little bit afraid. I dont know if i really need therapy, and if I don't, so what do I need to feel normal?

(a) LIME Explanation

So, I decided to go to psychiatrist, but now I feel like I'm not that bad? How bad is really bad? Maybe I just act dramatic I cry a lot without any specific reason, I want to kill myself quite often, but... I dont know if I really need therapy? Maybe its just who I am, just a lazy whiner? And when I will come to therapist I dont know what to say, what if that day I will feel fine and they will not believe me? I dont look like I struggle all the time, people around me don't even know that I have problems, I've always been a comedian in a company. I'm just a little bit afraid. I dont know if i really need therapy, and if I don't, so what do I need to feel normal?

(b) TRUST-MH Explanation

Comparison of highlighted attention phrases between LIME and TRUST-MH for a sample user post. TRUST-MH identifies semantically rich, multi-word concepts related to depression, whereas LIME primarily highlights isolated unigrams without contextual grounding.

References