Analyzing Weather-Induced Mental Health Risks in Online Forums using LLM

Rubaiyat Alam Ruhin
The University of Adelaide
Adelaide SA 5005
a1930497@adelaide.edu.au

1. Introduction

As climate change intensifies, its psychological toll on human well-being is becoming increasingly visible. While the physical impacts of extreme weather events—such as bushfires, droughts, or heatwaves—are well documented, their emotional and cognitive consequences often go unnoticed. Online mental health forums offer a unique window into these hidden experiences, as individuals frequently turn to platforms like Reddit or Beyond Blue to express distress, anxiety, or hopelessness linked to environmental stressors. This proposal focuses on the weather-induced dimension of psychological stress, aiming to detect and model the mental health effects of climatic variables such as temperature, humidity, and rainfall.

To capture these dynamics, this research proposes a machine learning pipeline that combines psycholinguistic analysis, temporal modeling, and large language models (LLMs). By extracting emotional features from user posts, aligning them with daily weather patterns in Australia, and classifying distress using LLM, the project aims to build a system that supports early detection of climate-related mental health risks. The outcomes may inform public health strategies and advance interdisciplinary approaches to environmental psychology, NLP, and computational social science.

2. Background

Existing research highlights the predictive power of language for identifying psychological states. Gu et al. developed a TextCNN-based classifier that used emotional lexicons and emoji features to detect cognitive change in support-seeking users [1]. Their method demonstrated that subtle shifts in tone and language use can indicate improvements or deteriorations in mental state. In another direction, Uddin et al. applied deep learning to structured survey data, showing that neural networks outperform traditional models in detecting depressive symptoms based on behavioral and contextual inputs [2].

Building on this, Xu et al. introduced Mental-LLM, a prompt-tuned language model that utilizes symptom-aware attention to classify mental health conditions from Reddit posts [3]. Their framework achieved superior performance by aligning LLM outputs with predefined clinical symptom prompts. Kaiser et al., meanwhile, applied network modeling to psychiatric symptom data, identifying key bridge symptoms that connect depression and anxiety disorders [4]. While these studies provide foundational methods for symptom detection, they do not integrate climate or weather context into mental health modeling, leaving a gap this project aims to fill.

This proposal extends prior work by linking weather data (temperature, humidity, rainfall) to psychological signals in online forums. It incorporates both dimensional emotion modeling—via the Circumplex Model of Affect—and the use of LLM for automated classification of climate-linked psychological distress. Together, these components offer a novel pipeline for understanding mental health in the face of environmental stressors.

3. Objectives

This project will achieve its overarching aim by pursuing three primary research objectives:

- **To extract psycholinguistic and emotional features from online sources and map these indicators onto the Circumplex Model of Affect for a dimensional understanding of expressed emotions.**
- **To conduct a temporal analysis correlating shifts in online mental health expressions with daily Australian weather data, including temperature, humidity, and rainfall etc.**
- **To use and validate a specialized Large Language Model (LLM) for the automated classification of weather-related psychological distress.**

4. Methodology

The research will begin by collecting textual data from online forums such as Reddit (e.g., r/Depression, r/Anxiety,

ChatGPT prompt *"Proof-read my writing"* was used to proof-read and fix grammar in the writing.

r/AusMentalHealth) and Beyond Blue, focusing on Australian users. Posts will be preprocessed using standard NLP techniques such as tokenization, lemmatization, stopword removal, and sentence segmentation. Emotion and symptom-related features will be extracted using LIWC, Empath, and custom emotion lexicons, following the approach used by Gu et al. to detect cognitive change [1]. These features will be mapped onto the Circumplex Model of Affect, allowing emotions to be represented along the dimensions of valence (positive–negative) for emotional profiling.

In the second phase, the study will incorporate daily Australian weather data, including temperature, humidity, and rainfall etc., obtained from online sources-such as keggle. Posts will be time-aligned with corresponding weather events to enable temporal correlation analysis. Following Xu et al.'s architecture [3], we will explore how psychological expression varies across pre-event, during-event, and post-event periods.

In the final stage, a specialized Large Language Model (LLM) will be used and validated for the automated classification of psychological distress associated with weather-related stressors. Rather than training a model from scratch, this phase will leverage an existing pre-trained LLM—such as RoBERTa or FLAN-T5—and adapt it to the mental health and climate context through task-specific instruction design and domain adaptation. The model will be evaluated on its ability to identify linguistic expressions of distress—such as anxiety, fear, and helplessness—in posts related to environmental conditions like extreme heat, humidity, or rainfall. A labeled evaluation dataset will be created from forum posts tagged by climate context. Model performance will be assessed using standard classification metrics (e.g., accuracy, F1 score, precision, recall), and its outputs will be analyzed for interpretability in relation to known emotional indicators. The result will be a validated LLM-based classification system capable of detecting weather-induced psychological distress in online discourse.

5. Project Timeline

- **Week 1**: Project topic finalization, objective refinement, and initial scoping discussions.
- **Weeks 2–6**: Literature review and data acquisition — including social media scraping (e.g., Reddit, Beyond Blue) and sourcing Australian weather datasets (temperature, humidity, rainfall).
- **Weeks 6–7**: Data preprocessing, cleaning, and exploratory data analysis (EDA). Begin initial feature extraction using LIWC, Empath, and emotion lexicons.
- **Weeks 8–10**: Core analysis phase — construct emotion embeddings using the Circumplex Model of Affect, align text data with weather records, and perform temporal correlation and changepoint analysis.
- **Weeks 11–12**: Initial evaluation of results. Visualize affective trends, conduct early analysis of weather-mental health correlations.
- **Weeks 13–18**: LLM selection, prompt formulation, and adaptation for automated classification of weather-related psychological distress. Prepare evaluation dataset and conduct benchmarking.
- **Weeks 19–24**: Final synthesis — validate LLM performance, analyze classification outputs for interpretability, and finalize results. Prepare visualizations, prototype (if applicable), and complete report writing.

References

[1] D. Gu, M. Li, X. Yang, Y. Gu, Y. Zhao, C. Liang, and H. Liu, "An analysis of cognitive change in online mental health communities: A textual data analysis based on post replies of support seekers," *Information Processing & Management*, vol. 60, 2023. [Online]. Available: https://doi.org/10.1016/j.ipm.2023.103192

[2] M. Uddin, M. Islam, M. Khondoker, et al., "Deep learning for prediction of depressive symptoms in a large-scale population-based survey," *BMJ Open*, vol. 9, no. 6, p. e026843, 2019. [Online]. Available: https://doi.org/10.1136/bmjopen-2018-026843

[3] X. Xu, T. Lin, Y. Zhang, Q. Chen, and H. Wang, "Mental-LLM: Leveraging Large Language Models for Mental Health Prediction via Online Text Data," *arXiv preprint*, arXiv:2401.04894, 2024. [Online]. Available: https://doi.org/10.48550/arXiv.2401.04894

[4] T. Kaiser, P. Herzog, U. Voderholzer, and M. Muehlhan, "Unraveling the comorbidity of depression and anxiety in a large inpatient sample using network analysis," *Psychological Medicine*, vol. 51, no. 13, pp. 2225–2233, 2021. [Online]. Available: https://doi.org/10.1017/S0033291720003077

ChatGPT prompt *"Proof-read my writing"* was used to proof-read and fix grammar in the writing.