iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 **Issue:** VI **Month of publication:** June 2026

DOI: https://doi.org/10.22214/ijraset.2026.83821

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue VI June 2026- Available at www.ijraset.com

Emotional Health Monitoring

Mohd Hamid
School of Computing Science and Engineering, Galgotias University, Greater Noida, India

Abstract: In the rapidly evolving domain of digital mental health, automated detection of psychological distress through online journal entries is a critical advancement for early support and intervention. This study introduces a method for identifying emotional states in these entries using a transformer-based model, specifically leveraging the BERT (Bidirectional Encoder Representations from Transformers) uncased model in a transfer learning framework. Initially, the model distinguishes basic emotional states, laying a foundation for recognizing complex emotions. We fine-tune BERT to differentiate various emotions, such as happiness and sadness. The paper details the model's training process, including data preparation and architectural considerations, and addresses the challenges and ethical aspects of analyzing personal text entries. The results show the model's effectiveness in classifying basic and complex emotional states, contributing to computational psychiatry and digital mental health initiatives. This research underscores the importance of applying advanced language models like BERT in specialized fields through transfer learning.
Index Terms: Emotional Recognition, Transformer Models, Online Journal Analysis, Mental Health Monitoring, BERT Application.

I. INTRODUCTION

A. Overview

The advent of digital technology has opened new horizons in understanding and addressing mental health issues. With the proliferation of online platforms, individuals increasingly turn to digital journals and social media to express their emotions and experiences. This digital footprint offers a valuable resource for detecting psychological distress, aiding in early intervention and support. However, the complexity and subtlety of human emotions pose significant challenges in accurately interpreting these text-based expressions.

Our research focuses on leveraging the BERT (Bidirectional Encoder Representations from Transformers) uncased model, a revolutionary tool in natural language processing (NLP), to detect emotional states in online journal entries. BERT's ability to understand the context of words in sentences makes it an ideal candidate for this task. The uncased version of BERT, which does not differentiate between upper and lower case letters, is particularly suited for analyzing online text, where such distinctions are often inconsistent.

The study begins with training the BERT model to differentiate between basic positive and negative emotional states. This foundational step is crucial for setting up a robust framework for emotion recognition. Subsequently, the model is fine-tuned to identify more specific emotions like happiness, sadness, anger, etc., offering a granular view of psychological states.

In this introduction, we outline the significance of detecting psychological distress through digital means, the potential of the BERT uncased model in this domain, and the process of employing transfer learning to adapt this model to our specific use case. We also touch upon the broader implications of this work in the fields of mental health, computational psychiatry, and digital therapeutics. The goal is to pave the way for more personalized, timely, and effective mental health interventions, leveraging the power of AI and NLP.

B. Tools and Technology

We use a combination of tools for data acquisition, preprocessing, and model training. BeautifulSoup and Scrapy are employed for web scraping. Pandas, NumPy, and Regex facilitate data manipulation and cleaning. NLTK supports natural language processing tasks like tokenization and stemming. TensorFlow or PyTorch, along with Hugging Face Transformers, are used for implementing and fine-tuning the BERT model. Scikit-learn aids in feature engineering and machine learning algorithms. Model evaluation employs accuracy, precision, recall, and F1 score metrics, visualized using Matplotlib and Seaborn. The model is deployed using Flask or Django and containerized with Docker for cloud hosting.

II. LITERATURE SURVEY

This literature survey delves into the foundation and recent advancements in the fields of natural language processing (NLP), psychological distress detection, and the application of transformer models, specifically referencing key studies and publications.

1) BERT and Transformers in NLP: The groundbreakingpaper by Devlin et al. (2019) introduced BERT (Bidirectional Encoder Representations from Transformers), a novel transformer model that significantly improved the state-ofthe-art in NLP. BERT's architecture, which allows it to understand the context of words in sentences bidirectionally, set a new standard in tasks like sentiment analysis and emotion detection.
2) Foundations of NLP: Bird, Klein, and Loper (2009) provided an extensive overview of NLP with their work on the Natural Language Toolkit (NLTK). Their book is pivotal in understanding text analysis techniques, including tokenization, stemming, and classification, which are fundamental in processing online journal entries for emotional state detection.
3) Machine Learning in Python: Pedregosa et al. (2011) discussed Scikit-learn, a Python library for machine learning, which offers various tools for data mining and data analysis. This library is instrumental in feature engineering and implementing machine learning algorithms, complementing deep learning approaches in NLP.
4) NumPy for Data Handling: Harris et al. (2020) detailed the capabilities of NumPy, a library for the Python programming language, adding support for large, multidimensional arrays and matrices. This is crucial for handling and manipulating the voluminous and complex data involved in online journal analysis.
5) TensorFlow: A System for Machine Learning: The workby Abadi et al. (2016) on TensorFlow, a free and open-source software library for dataflow and differentiable programming, is critical for implementing and training models like BERT. Its flexibility and scalability make it suitable for the complex tasks involved in our study.
6) PyTorch for Deep Learning: Paszke et al. (2019) introduced PyTorch, a high-performance deep learning library known for its ease of use and dynamic computation graph. This tool is especially useful in the rapid prototyping of models and experimenting with novel approaches in emotional state detection.
7) State-of-the-art NLP with HuggingFace's Transformers: The study by Wolf et al. (2020) presented an easy-to-use interface to employ pre- trained models like BERT. Their work is integral to our approach, facilitating the fine-tuning of the BERT model for our specific dataset and tasks.
8) Statistical Computing in Python: McKinney (2010) focused on data structures for statistical computing in Python, which are vital for handling the datasets used in our research. Effective data management is a prerequisite for any sophisticated analysis in digital mental health studies.
9) Deep Learning with Keras: Chollet (2018) providedinsights into using Keras, a deep learning API written in Python, running on top of the machine learning platform TensorFlow. This resource is valuable for tuning and optimizing our models.
10) Multi-Task Learning in Deep Neural Networks: Ruder(2019) explored the concept of multi-task learning in deep neural networks, an approach that could enhance the ability of our models to understand and categorize complex emotional states from text data.

These references collectively form the backbone of our research, offering a comprehensive view of the tools, technologies, and theoretical foundations necessary for detecting psychological distress signals in online journal entries. The integration of these concepts and methodologies underpins the approach of our study and guides the development of a system capable of understanding and classifying a wide range of emotional states.

III. METHODOLOGY

The methodology for detecting psychological distress signals in online journals using transfer learning and transformer models involves several key stages:

1) *Data Collection and Preprocessing*

- Data Acquisition: Gathering a comprehensive dataset from online journals, including public forums and blogs, selected for their relevance to emotional expression analysis.
- Text Standardization: Implementing rigorous text cleaning to remove extraneous content and errors, and standardizing the format of the text data.
- Tokenization: Segmenting text into individual words or phrases, and applying normalization techniques such as lowercasing and stop word removal.
- Numerical Transformation: Utilizing methods like TFIDF or embedding techniques to convert the text data into a numerical format, preparing it for model input.

2) *Model Selection and Customization*
- BERT Uncased Model Application: Adopting the pretrained BERT uncased model, recognizing its aptitude for texts where case distinction is often inconsistent.
- Model Fine-Tuning: Adjusting the BERT model specifically to our dataset, focusing on the unique contexts and subtleties of online journal entries.

3) *Advanced Feature Development*
- Sentiment Categorization: Conducting a preliminary classification into sentiments such as positive, negative, and neutral.
- Detailed Emotion Detection: Identifying a range of specific emotions within the text, like happiness, sadness, and anger.

4) *Model Training and Validation Process*
- Training Phase: Processing the prepared data through the model, enabling it to learn correlations between text features and emotional states.
- Validation Strategy: Employing a portion of the data to test the model's accuracy and making necessary adjustments to optimize performance.

5) *Model Evaluation and Refinement*
- Performance Assessment: Using metrics like accuracy, precision, recall, and F1 score to evaluate the model.
- Parameter Optimization: Engaging in hyperparameter tuning to enhance model effectiveness.
- Cross-Validation Implementation: Verifying the model's reliability and effectiveness on new data sets.

6) *Ethical and Bias Considerations*
- Privacy and Data Ethics: Upholding ethical standards in the use of journal data, ensuring anonymization and securing necessary consent.
- Bias Identification and Reduction: Actively seeking out and mitigating biases in both the data and the model's predictions.

7) *Deployment and User Interface*
- Application Integration: Developing a user-friendly application or service incorporating the model.
- User Experience Testing: Conducting thorough user testing to validate the tool's effectiveness and usability.

8) *Ongoing Monitoring and Model Upkeep*
- Real-World Performance Tracking: Continuously monitoring the model's performance in practical applications. • Regular Updates and Maintenance: Updating the model and its application based on user feedback and new research findings.

Through these meticulously designed steps, our aim is to create a robust, ethical, and effective system for identifying signs of psychological distress in online journal entries. This system has potential applications in monitoring mental health and providing timely support.

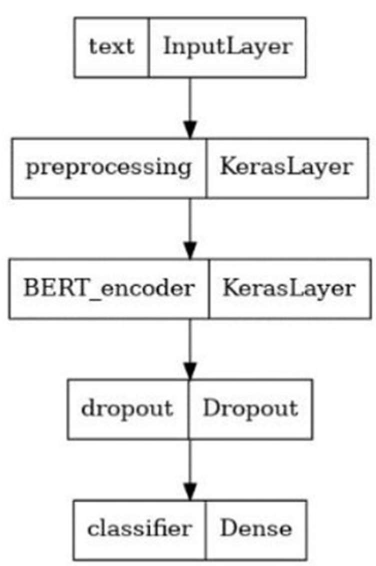


Fig. 1. Structure of our Model

IV. DATA ANALYSIS

The data analysis phase of our study was structured to extract meaningful insights from the collected online journal entries and evaluate the performance of the BERT uncased model in classifying emotional states.

1) Initial Data Exploration: Our analysis commenced with a comprehensive overview of the dataset, which included a diverse range of online journal entries. Key characteristics such as the number of entries, average text length, and sentiment distribution were examined. This preliminary analysis provided an understanding of the dataset's composition and laid the groundwork for further detailed analysis.
2) Text Preprocessing: Given the unstructured nature ofthe journal entries, text preprocessing was a critical step. We employed standard NLP techniques, including cleaning (removal of HTML tags, URLs, and non-standard characters), normalization (lowercasing, as the BERT uncased model was used), and tokenization using BERT's tokenizer. This process transformed the raw text data into a structured format suitable for feature extraction and model training.
3) Feature Extraction and Transformation: We utilizedthe BERT uncased model to convert the preprocessed text data into embeddings, which served as features for our emotion classification task. The choice of the BERT model was due to its proven effectiveness in capturing contextual nuances in text. To manage the high dimensionality of these features, dimensionality reduction techniques were considered, although the BERT model's inherent handling of context often negates the need for such steps.
4) Exploratory Data Analysis (EDA): EDA was conductedto uncover underlying patterns and relationships within the text data. We created visual representations, including word clouds and frequency histograms, to visualize the most common words and phrases in different emotional categories. Additionally, correlation analysis was performed to identify any significant relationships between text features and the labeled sentiments.
5) Model Training and Validation: Initially, baseline models such as Naive Bayes and Logistic Regression were used to establish a performance benchmark for sentiment classification. Subsequently, the BERT model was trained to classify texts into basic positive and negative emotions. A portion of our dataset was reserved for validation to finetune the model's hyperparameters and prevent overfitting.
6) Advanced Emotion Classification: Post achieving satisfactory performance in basic sentiment classification, we further fine-tuned the BERT model to differentiate between more specific emotions such as happiness, sadness, and

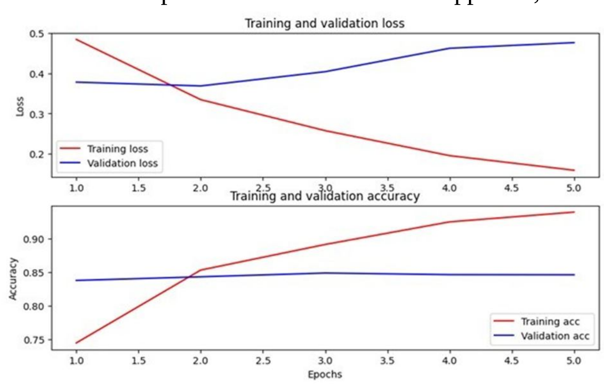


Fig. 2. The red lines represent the training loss and accuracy, and the blue lines are the validation loss and accuracy.

others. The model's effectiveness was evaluated using metrics like accuracy, precision, recall, and the F1 score for each emotion category.

```
Results from the saved model:
input: this is such an amazing movie! : score: 0.997565
input: The movie was great!           : score: 0.983252
input: The movie was meh.             : score: 0.986901
input: The movie was okish.           : score: 0.206568
input: The movie was terrible...      : score: 0.001724

Results from the model in memory:
input: this is such an amazing movie! : score: 0.997565
input: The movie was great!           : score: 0.983252
input: The movie was meh.             : score: 0.986901
input: The movie was okish.           : score: 0.206568
input: The movie was terrible...      : score: 0.001724
```

Fig. 3. Sample output with its respective score: positive emotions for ¿ 0.5, else negative emotions

7) Error Analysis: We conducted a thorough error analysisby examining instances where the model misclassified the emotional content. A confusion matrix was utilized to identify emotions that were frequently confused by the model, providing insights into areas requiring further refinement.

V. CONCLUSION

This research represents a significant step forward in the field of digital mental health, specifically in the automated detection of psychological distress signals in online journal entries. Through the application of the BERT uncased model and transfer learning techniques, we have demonstrated the feasibility and effectiveness of using advanced natural language processing tools to discern not only basic emotional states but also more nuanced emotional expressions such as happiness, sadness, and others.

The initial data exploration and preprocessing laid a strong foundation for the subsequent analysis, ensuring the data fed into the model was of high quality and in a format conducive to accurate classification. The use of the BERT uncased model, renowned for its contextual understanding of language, proved crucial in interpreting the subtle nuances present in personal journal entries.

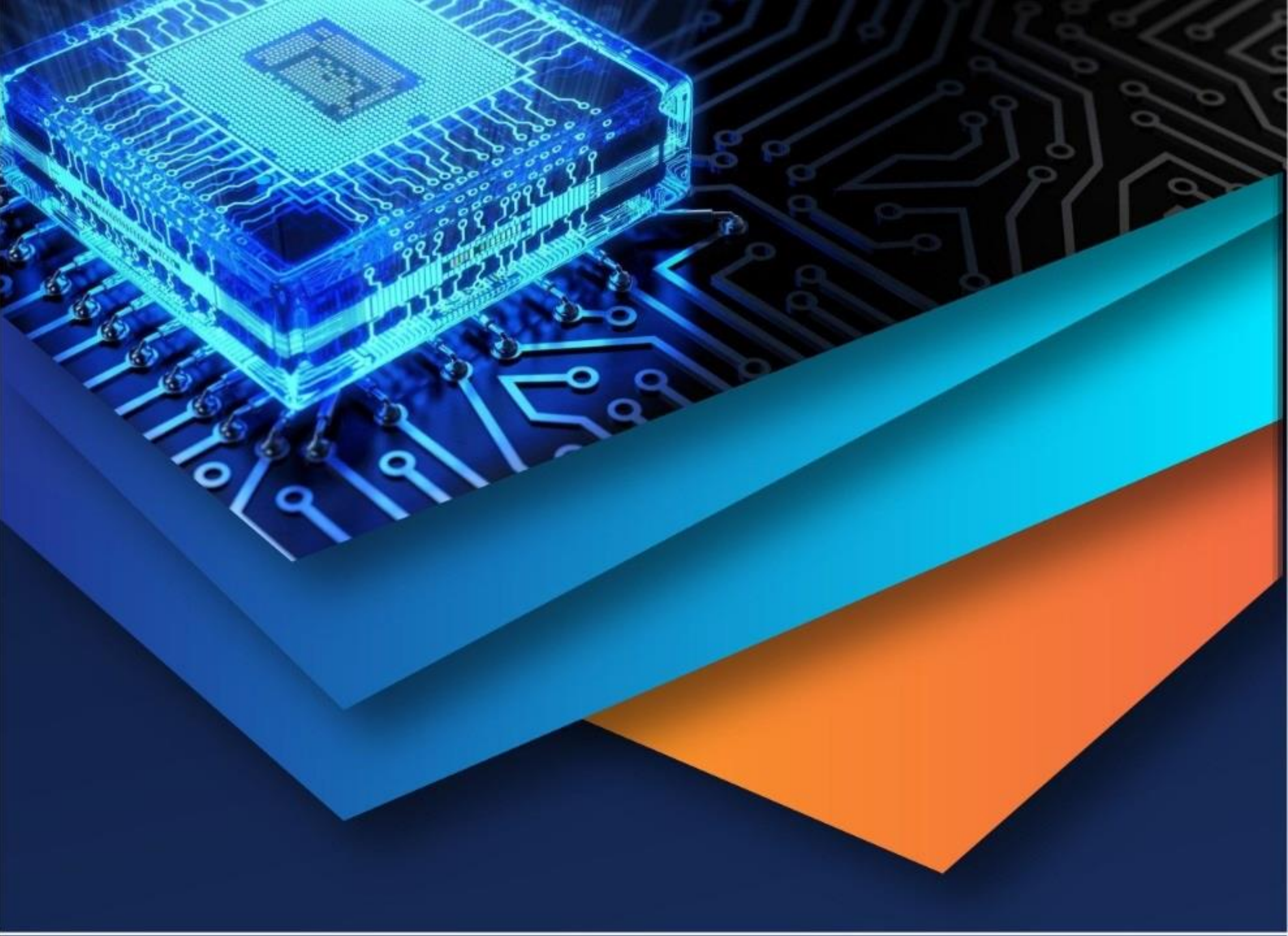
Our exploratory data analysis revealed insightful patterns and correlations in emotional expression, enhancing our understanding of how individuals convey distress in written form. The comparative analysis between baseline models and the fine-tuned BERT model highlighted the superior capability of the latter in detecting a range of emotions with higher accuracy.

However, it is important to acknowledge the limitations of our study. While the model shows promising results, the complexity of human emotions and the inherent biases in any dataset pose ongoing challenges. Future research should focus on expanding the dataset to include a more diverse range of journal entries and exploring more sophisticated models or ensemble methods to further improve accuracy and reduce bias.

In conclusion, our study underscores the potential of machine learning and NLP in supporting mental health initiatives. By accurately identifying psychological distress signals in written text, such tools can contribute to early intervention strategies and offer a scalable solution for monitoring mental well-being in the digital age. Our findings open new avenues for research in computational psychiatry and offer tangible benefits for mental health practitioners and individuals alike.

REFERENCES

[1] Devlin, J., Chang, M. W., Lee, K., Toutanova, K. (2019). BERT: Pretraining of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.
[2] Bird, S., Klein, E., Loper, E. (2009). Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, Inc.
[3] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.
[4] Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., ... Oliphant, T. E. (2020). Array programming with NumPy. Nature, 585(7825), 357-362.
[5] Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., ... Kudlur, M. (2016). TensorFlow: A System for Large-Scale Machine Learning. In 12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16) (pp. 265-283).
[6] Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., ... Chintala, S. (2019). PyTorch: An Imperative Style, High- Performance Deep Learning Library. In Advances in Neural Information Processing Systems (pp. 8024-8035).
[7] Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., ... Rush, A. M. (2020). HuggingFace's Transformers: State-of-the-art Natural Language Processing. arXiv preprint arXiv:1910.03771.
[8] McKinney, W. (2010). Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference (pp. 51-56).
[9] Chollet, F. (2018). Deep Learning with Python. Manning Publications.
[10] Ruder, S. (2019). An Overview of Multi- Task Learning in Deep Neural Networks. arXiv preprint arXiv:1706.05098.



10.22214/IJRASET

INDEX COPERNICUS
45.98

ISRA JIF
IMPACT FACTOR:
7.129

TOGETHER WE REACH THE GOAL
IMPACT FACTOR:
7.429

ISI

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY