



**VIT<sup>®</sup>**  
**Vellore Institute of Technology**  
(Deemed to be University under section 3 of UGC Act, 1956)

**SCHOOL OF COMPUTER SCIENCE ENGINEERING AND  
INFORMATION SYSTEMS**

**FALL SEMESTER 2025-2026**

**COURSE NAME: NATURAL LANGUAGE PROCESSING**

**COURSE CODE: SWE1017**

**SLOT: E1**

**TITLE: Emotion Recognition Techniques for Hindi Language Text**

**REVIEW-1**

**TEAM MEMBERS:**

**Mamidi Smaran:22MIS0385**

**Akula Satish:22MIS0288**

## ABSTRACT:

In recent years, emotion detection from social-media text has become increasingly important for applications in consumer insight, psychology, human–computer interaction, and intelligent systems. Leveraging the growing volume of native Hindi content where affective expressions are frequent, this study investigates automatic emotion recognition for Devanagari Hindi text. It evaluates deep sequence models—CNNs, LSTMs, and Bi-directional

LSTMs—for multi-class emotion classification. The system emphasizes Hindi-specific preprocessing, including tokenization, normalization, and handling of negations (e.g., “नहीं”), to improve robustness across colloquial and formal registers. Current experiments demonstrate that LSTM/BiLSTM architectures provide strong, reliable baselines on the curated Hindi emotion corpus. Building on these results, ongoing work is extending the framework to transformer-based models, with a planned implementation using a BERT-family checkpoint suitable for Hindi, to assess potential gains in accuracy and generalization across domains and class distributions.

## Existing Methods:

- **Language Limitation:** Most emotion recognition systems are developed for English or other high-resource languages. These models often fail to capture the nuances of Hindi, which has rich morphology and frequent code-mixing with English.
- **Generalized Models:** Existing models for emotion recognition are usually trained on generic multilingual datasets and are not fine-tuned for Hindi text, leading to reduced accuracy in classification.
- **Shallow Context Understanding:** Traditional ML models (SVM, Naïve Bayes, etc.) and even some deep learning models (CNN, RNN) rely on surface-level features and struggle to capture deeper semantic context in Hindi sentences.
- **Dataset Constraints:** Very few publicly available datasets exist for Hindi emotion recognition. Most systems rely on small, imbalanced datasets, limiting model performance.

## Proposed methods:

- **Customized Hindi Dataset:** A labeled dataset of Hindi sentences is used for training, covering diverse emotions such as joy, anger, sadness, and fear. Preprocessing steps include Unicode handling, tokenization, stopwords removal, and managing code-mixed text.
- **Deep Learning Models:** LSTM and BiLSTM models are employed to capture sequential dependencies in Hindi text, improving emotion classification compared to traditional ML techniques.
- **Transformer-based Approach (BERT):** A BERT-based model is being implemented to enhance contextual understanding. Unlike LSTM/BiLSTM, BERT captures bidirectional dependencies and provides rich contextual embeddings, leading to better accuracy.
- **Improved Accuracy & Robustness:** By training specifically on Hindi data and using deep contextual embeddings, the proposed system is expected to outperform generalized models and provide more reliable emotion recognition for Hindi text.

## DATA COLLECTION/ANALYSIS:

### Dataset:

**Github link:** <https://github.com/smaranmamidi/hindi-emotion-classifier>

### Dataset:

```
1 Label,Sentence
2 neutral,आप अपने हाथ भरे होंगे।
3 neutral,कि मैंने किया। कि मैंने किया।
4 neutral,तो चलो अपने कर्तव्यों के बारे में थोड़ा बात करें।
5 surprise,मेरे कर्तव्य? ठीक है।
6 neutral,"अब आप एक पूरे डिजीजन का नेतृत्व करेंगे, तो आप बहुत सारे कर्तव्य होंगे."
7 neutral,मैं देखता हूँ।
8 neutral,लेकिन वहाँ शायद 30 लोग आपके नीचे होंगे तो आप उन पर एक निश्चित मात्रा फेंक सकते हैं।
9 neutral,जानना अच्छा है।
10 neutral,हम विस्तार में जा सकते हैं
11 fear,मैं तुमसे नहीं माँगता!
12 neutral,"ठीक है तो, हम तुम के लिए Pazartesi को एक निश्चित जवाब है, लेकिन मुझे लगता है कि मैं कुछ विश्वास के साथ कह सकता हूँ, आप यहाँ अच्छी तरह से फिट होंगे।"
13 surprise,सचमुच?
14 neutral,निश्चित रूप से. आप आराम कर सकते हैं
15 surprise,लेकिन फिर कौन? मैं पिछले महीने के साथ बाहर गया था रेस्तरां?
16 sadness,तुम्हें पता है? इसे भूल जाओ!
17 surprise,"नहीं-नहीं-नहीं, नहीं! कौन, तुम किस बारे में बात कर रहे थे?"
18 fear,"नहीं, मैं नहीं जानता, मैं वास्तव में नहीं जानता"
19 neutral,ठीक है!
20 neutral,"ठीक है, ठीक है..."
21 neutral,"हाँ, निश्चित है!"
22 neutral,"Hey, मोन."
23 neutral,Hey-hey-hey. आप कुछ सुनना चाहते हैं कि शून्य है.
24 joy,मैं कभी करता हूँ।
```

## References:

- 1) Wadhawan & Aggarwal, 2021 for code-mixed benchmarks and transformer comparison <https://aclanthology.org/2021.wassa-1.21.pdf>
- 2) IOSR “Hindi Text Emotion Recognition based on Deep Learning” for method parity with a Hindi LSTM pipeline, including preprocessing steps specific to Hindi. <https://www.iosrjournals.org/iosr-jmca/papers/Vol7-Issue3/Series-1/E07032429.pdf>
- 3) BiLSTM ABSA in Hindi (2024) to support BiLSTM effectiveness and training details in Hindi text classification tasks, even if task is different <https://thescipub.com/pdf/jcssp.2024.1438.1445.pdf>
- 4) J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,” in Proc. NAACL-HLT, 2019, pp. 4171–4186. <https://www.iosrjournals.org/iosr-jmca/papers/Vol7-Issue3/Series-1/E07032429.pdf>
- 5) T. Pires, E. Schlinger, and D. Garrette, “How Multilingual is Multilingual BERT?,” in Proc. ACL, 2019, pp. 4996–5001. <https://www.sciencedirect.com/science/article/pii/S1877050920311236>
- 6) Ghosh, Souvick, Satanu Ghosh, and Dipankar Das. (2017) "Sentiment identification in code-mixed social media text." *arXiv preprint arXiv:1707.01184*. <https://arxiv.org/abs/1707.01184>
- 7) R. Kumar et al., “Emotion Recognition in Hindi Text Using Multilingual BERT Transformer,” *Multimedia Tools and Applications*, 2023. <https://thescipub.com/pdf/jcssp.2024.1438.1445.pdf>
- 8) C. Sun, X. Qiu, Y. Xu, and X. Huang, “How to Fine-Tune BERT for Text Classification?,” in *Chinese Computational Linguistics*, 2019, pp. 194–206 <https://www.sciencedirect.com/science/article/pii/S1877050925017582>