Am I hurt?: Evaluating Psychological Pain Detection in Hindi Text using Transformer-based Models

RAVLEEN KAUR and M.P.S. BHATIA, Department of Computer Science and Engineering, Netaji Subhas University of Technology, New Delhi, India

AKSHI KUMAR[1], Department of Computing, Goldsmiths, University of London, United Kingdom

The automated evaluation of pain is critical for developing effective pain management approaches that seek to alleviate while preserving patients' functioning. Transformer-based models can aid in detecting pain from Hindi text data gathered from social media by leveraging their ability to capture complex language patterns and contextual information. By understanding the nuances and context of Hindi text, transformer models can effectively identify linguistic cues, sentiment and expressions associated with pain enabling the detection and analysis of pain-related content present in social media posts. The purpose of this research is to analyse the feasibility of utilizing NLP techniques to automatically identify pain within Hindi textual data, providing a valuable tool for pain assessment in Hindi-speaking populations. The research showcases the HindiPainNet model, a deep neural network that employs the IndicBERT model, classifying the dataset into two class labels {pain, no_pain} for detecting pain in Hindi textual data. The model is trained and tested using a novel dataset, दर्द-ए-शायरी (pronounced as *Dard-e-Shayari*) curated using posts from social media platforms. The results demonstrate the model's effectiveness, achieving an accuracy of 70.5%. This pioneer research highlights the potential of utilizing textual data from diverse sources to identify and understand pain experiences based on psychosocial factors. This research could pave the path for the development of automated pain assessment tools that help medical professionals comprehend and treat pain in Hindi speaking populations. Additionally, it opens avenues to conduct further NLP-based multilingual pain detection research, addressing the needs of diverse language communities.

CCS CONCEPTS
• **Computing methodologies** → Artificial Intelligence; Natural Language Processing;
• **Human-centered Computing** → social media.

Additional Keywords and phrases: Pain Detection, social media, word embeddings, transformer-based models, emotional pain

ACM Reference format:
Ravleen Kaur, M.P.S. Bhatia and Akshi Kumar. 2023. Am I hurt? Evaluation of psychological pain detection using transformer-based models in Hindi text.

Abbreviations to be used in the paper

Abbreviation	Description
IASP	*International Association for the Study of Pain*
EHRs	Electronic Health Records
NLP	Natural Language processing
DL	Deep Learning
BPL	Biopsychosocial
BERT	Bidirectional Encoder Representation from Transformers
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
DeBERTa	Decoding enhanced BERT with disentangled attention
XLNet	eXtreme Language Understanding Network
RoBERTa	Robustly optimized BERT approach

[1] Akshi.kumar@gold.ac.uk

1. INTRODUCTION

The IASP *(International Association for the Study of Pain)* described pain as “an unpleasant sensory and emotional experience associated with, or resembling that associated with, actual or potential tissue damage” [1]. It is challenging to manage pain since it is a personal, subjective sensation linked to several conditions and presentations [2]. It is often referred to as an unpleasant emotion or sensation that acts as a body’s alarm system, warning of impending damage or harm. Even with the extensive pain ratings and scales accessible in pain clinics, EHRs (Electronic Health Records) are a rich database with a large amount of granular pain information [3]. Acute or chronic pain can vary in severity, duration, and quality. A specific injury or disease can cause acute pain, which is often abrupt and brief [4]. On the other hand, chronic pain lasts for a long time- typically three months or longer. It might last longer than it usually would or appear for no apparent reason [5].

The perception and feeling of pain can be significantly influenced by psychological factors. Both sensory and cognitive processes are involved in the complicated phenomena of pain. Psychological variables can affect how people perceive pain and how they interpret and react to it. Pain has evident emotional and behavioural effects that have an impact on how chronic issues develop and how therapy works [6]. It is well recognized that psychological processes can trigger the onset of chronic pain issues [7]. Avoiding regular activities resulting in deterioration of functionality, worsens pain, and heightens fear of (re)injury. Anger, distress, fear, anxiety, remorse, disappointment, and depression are common emotional responses to pain [8]. Some of the psychological factors, including empathy, pain anticipation, affection, emotions, attachment, and learning can activate the neuropathological processes [9]. The patient's ability to control these emotions has an influence on how they affect pain. The various studies have demonstrated a bidirectional relationship between depression and pain: both chronic pain and depression are positively associated with each other's growth. Hope, pain acceptance and optimism are all positive psychological qualities that influence how one adjusts to chronic pain [10]. The semantic meaning of words is essential to how pain is perceived because pain perception in the brain is a sophisticated system of modulation and processing. One stimulus can influence several others according to the priming effect, a cognitive phenomenon. It is therefore possible that this effect is essential for controlling and sensing pain [11]. People commonly resort to social media as an outlet to express and vent their emotional pain [12]. NLP (Natural Language processing) is a computer science branch that examines unstructured free text and uses statistical algorithms and computational techniques to extract quantitative data from it.

NLP and Deep Learning (DL) have gained popularity, and there is a growing interest in investigating these technologies to aid in pain diagnosis utilizing textual data. An emerging field of research, pain detection using NLP, uses text data to identify and rate pain events. While traditional techniques of pain evaluation rely on subjective self-reporting by individuals, which are subjective and prone to biases, NLP presents a unique opportunity to extract pain-related information from numerous textual sources, including social media postings, patient records, online forums, and healthcare surveys. Medical keyword searches, diagnosis categorization and information extraction from clinical notes about disease type and symptoms have all been done using NLP approaches [13]. NLP algorithms may find phrases associated with pain, such as "headache" or "frontal," in clinical notes or other text sources. NLP models can recognize, and extract phrases and concepts associated with pain from unstructured text data by using methods like named entity recognition and keyword extraction. Automated pain detection utilizing NLP methods can offer insightful information and help psychoanalysts and NLP specialists spot probable psychological issues in patients. Models for emotion detection may be created using machine learning methods like logistic regression and deep learning. The emotions detected using NLP approaches can offer useful information for identifying depression, aiding in pain assessment, and enabling personalized medical care.

Social media platforms have grown into effective venues for people to express their feelings and share personal experiences with an extensive audience. The social media content often reflects the emotional state of users, including instances of pain and distress [14]. To discover linguistic patterns, the social media posts are assessed based on the frequency of specific words or phrases, the conversational tone, and the overall sentiment. When training machine learning models to predict instances of pain or identify depressive episodes, these patterns might be employed as useful features. This method is especially useful in the subject of psycholinguistics, which focuses on understanding how linguistic and mental variables interact. Psycholinguistics researchers have established that people's personalities may be inferred from their written language and writing styles [15], which can disclose information about their emotional health. This data can be used to spot depressive episodes or pinpoint those who are vulnerable to emotional pain. For instance, a person may be categorized as having a more depressed personality trait if they regularly utilize phrases like grief, sadness, hurt, or tears in their postings [16]. On the other hand, those with a more upbeat and outgoing attitude may use more positive language and expressions. Using labelled data with annotations indicating the presence or absence of pain or depression episodes, machine learning models may be developed. These models are capable of recognizing language patterns and relationships between certain words, phrases, or emotional expressions and the existence of pain or depressed symptoms. The algorithms may then be used to predict or detect the episodes of pain. There are various benefits of using social media data for pain detection. First off, it offers a non-intrusive and economical way to regularly check on people's health, especially in situations where direct observation or in-person assessments might not be practical. Additionally, it enables the analysis of large-scale datasets, which makes it easier to discover population-level trends and patterns. Finally, using machine learning models built on this data can assist healthcare practitioners in making informed decisions and enabling early detection of pain, intervention, and customized treatment planning. Most Indian languages are classified as low resource, implying that there is a scarcity of data for training NLP systems. The lack of data makes creating accurate and effective NLP systems for these languages difficult. Indian languages frequently have less resources, such as text corpora, labelled datasets, and linguistic tools, compared to the languages with vast resources, such as English. As a result, developing NLP systems for Indian languages requires innovative methods to overcome data scarcity and address the unique linguistic characteristics of these languages. Hindi is a widely spoken language in India and the other neighbouring countries which offers a unique context for pain detection research. Pain detection in local or regional languages like Hindi has received very little attention, even though multiple studies have concentrated on it using English textual data. However, the accessibility of large-scale textual data in Hindi, such as social media posts, patient records, healthcare forums and online surveys, presents chances to investigate NLP and DL strategies particularly adapted to the Hindi language.

The convergence of computational linguistics and emotional health has opened new avenues for understanding human sentiment, particularly through the lens of poetry. Our research introduces “दर्द-ए-शायरी: The Language of Pain,” a novel dataset of Hindi poetry that serves as a rich repository for detecting expressions of pain in textual form. Harnessing the capabilities of deep neural networks with static embeddings, we have pioneered a method that not only identifies but also interprets the nuances of pain with remarkable accuracy and efficiency. This breakthrough has profound implications across multiple domains, ranging from mental health surveillance to enriching cultural studies. Here are some of the real-time applications that exemplify the utility of our research:

- **Mental Health Monitoring**: Your model could be integrated into social media platforms to flag posts that indicate high levels of distress or pain, which can then be used by mental health professionals to provide timely support to individuals in need.
- **Crisis Intervention**: Organizations like suicide prevention hotlines could use the model to scan social media for signs of acute emotional pain, enabling them to reach out proactively to people who may be at risk of self-harm.
- **Cultural Studies**: Researchers in humanities and social sciences might apply the model to study the expression of pain in modern Hindi poetry as a reflection of societal issues, offering new insights into the collective psyche.
- **Healthcare Communication**: Hospitals and clinics could use the model to analyze patient interactions on their portals, helping to identify those who may be experiencing high levels of pain or discomfort and prioritizing their care.
- **Literary Analysis**: The tool could assist scholars and students in literary analysis by identifying and examining themes of pain and suffering in Hindi poetry, enhancing academic research and education.
- **Content Moderation**: Social media companies could use the model for content moderation, identifying posts that contain expressions of pain for further human review to ensure community guidelines around sensitive content are upheld.
- **Public Health Research**: Public health researchers could use the model to analyze social media data for patterns of pain expression related to physical or psychological conditions, contributing to large-scale health studies.
- **Artificial Intelligence in Therapy**: AI-driven therapy apps might integrate your model to track user progress or setbacks by analyzing their language use over time, personalizing therapeutic content accordingly.

These examples illustrate the potential of your research to extend beyond academic settings and provide practical, real-world benefits in various domains. Thus, the significant contributions to this research can be summarized as follows:

- Creation of a new dataset in Hindi called "दर्द-ए-शायरी: The Language of Pain dataset" (pronounced as *Dard-e-Shayari*). This dataset classifies the data into two class labels “pain” or “no_pain”. It will be useful for evaluating and benchmarking pain detection methods in Hindi.
- Developing a model called HindiPainNet that uses different deep learning techniques and word embeddings to detect pain attributes in Hindi poetry. We evaluated the model's performance and compared different word embeddings to find the most effective approach for pain detection in Hindi.
- A thorough examination of word embeddings in the Hindi dataset for pain using Psychometric NLP. This analysis helps us understand how different word embeddings perform in identifying pain-related content in Hindi text.

The rest of the paper is structured as follows: Section 2 discusses the recent advancements in pain detection, with a focus on the textual data. In Section 3, fundamental concepts related to the static and dynamic word embeddings and commonly used deep learning models in NLP are explored. In Section 4, we delve into the specifics of our curated Hindi pain detection dataset and our proposed model, HindiPainNet, designed for the purpose of pain detection in Hindi textual data. Section 5 summarizes and analyses our model's experimental outcomes. Section 6 concludes the paper by summarizing findings and proposing potential avenues for future research and improvements in this area.

2. RELATED WORK

The notion that information is important in the emergence of health and disease aligns with the primary principles of the Biopsychosocial (BPS) model. This paradigm proposes that an individual's well-being or sickness is impacted by psychological and social factors in addition to physical changes.

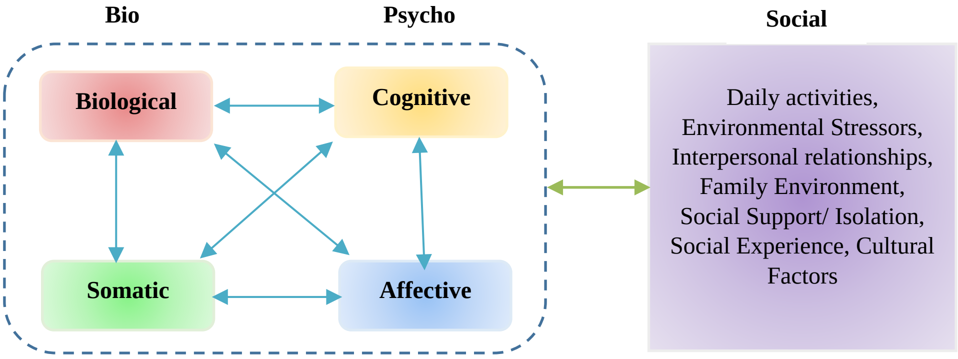

Fig. 1. Biopsychosocial model

As a result, the BPS model recognizes the interdependence of biological, psychological, and social factors in understanding and treating pain [17]. Language has lately been demonstrated to be useful in interpreting and quantifying a range of pain experience aspects other than qualia or severity by researchers. The psycholinguistic and affective characteristics of pain related words were examined in order to set appropriate guidelines for future research on painful emotions [18]. Detecting rumors in Arabic tweets poses challenges due to linguistic nuances. Gumaei A. et. al. [19] proposed XGBoost-based approach, leveraging diverse features, achieving a top accuracy of 97.18% on a public dataset, surpassing recent methods in rumor detection. Using a large-scale dataset comprising 2.5 million surveys and 1.8 million tweets, a study by Aggarwal A. [20] investigates the prediction of community-level pain using Twitter posts and reveals significant variations in pain expressions across different communities in the United States, highlighting the potential for Twitter-based interventions in community-focused pain management. Sawhney et. al. [21] provides SISMO, a hierarchical attention model that considers the ordinal structure of social media suicide risk assessment. It uses soft probability distribution to accommodate for different risk levels and shows good results on real-world Reddit posts that has been annotated by professionals. Patients with chronic pain were evaluated for their reaction to placebo using quantitative language patterns collected from semi-structured interviews [22], while pain disparities in underprivileged communities were discovered using comprehensive text mining of Electronic Health Records (EHRs) [23]. Social media posts have also been studied for a variety of pain-related purposes, including tracking patients over time and identifying new pain phenotypes [24], geographically monitoring and characterizing opioid use [25], exploring how pain is socially [26] and identifying population-level increases in pain conditions and symptoms [27]. A study conducted by Kumar A. et. al. [28] utilized XLM-R transformer with zero-shot transfer learning for sentiment analysis in resource-poor Indian languages. A study by Caldo D. et. al. [29] investigates the emotional patterns of effective spine pathology web pages, indicating their potential importance in comprehending chronic pain and affecting health-related behaviours. The findings highlight the necessity of examining the biopsychosocial components of pain and present ethical problems for digital health information providers. An Irish based researcher Mullins C.F. et. al. [30] examined the tweets about pain in Ireland over a two-week period and found that the most common terms were headache (90%) and migraine (66%). The majority of tweets were from women and identifying the dominant category of advice on back pain management. A longitudinal study by Deng H. et. al. [31] used natural language processing (NLP) tools to analyze tweets regarding migraines. User behaviour profiles were reported and examined, such as tweeting frequencies, popular words, and sentimental presentations. Many expressive tweets had a negative emotion, particularly those with a high frequency and severe sentiment, including the use of profanity. Guo Y. et. al. [32] examines the utilization of social networking platforms (Reddit and Twitter) as a valuable resource for studying migraine, including the availability of relevant discussions and the development of a text classification system. The use of deep learning neural networks and

NLP to text data from medical discharge summaries is investigated in the study proposed by Yang Z. et. al. [33], with an emphasis on patient phenotyping. The study emphasizes the importance of data quality, quantity, and token selection in obtaining higher performance, particularly in Chronic Pain classification. To examine headache and migraine discussions in Japan, Germany, and France, researchers analyzed social media data from several platforms, revealing linguistic trends, treatment references, and demographic information [34]. Table 1 demonstrates a thorough examination of the latest advancements in the field.

Table 1. The existing state-of-the-art

Study	Objective	Dataset
Aggarwal A. et. al., (2023) [20]	To predict community-level pain aligned with survey-based self-reports	Social media posts from Twitter
Sawhney R. et. al., (2021) [21]	To predict suicide ideation using Twitter	Reddit data
Guo Y. et. al., (2023) [32]	To create a text classification system that can automatically detect posts about self-reported migraines	Twitter & Reddit posts
Branco P. et. al., (2022) [22]	To predict placebo in chronic pain patients using NLP	Self-reported data
Deng H. et. al., (2020) [31]	To analyze tweets reporting migraine activity & explore their socio-behavioural aspects	Users' social profile data
Fiok K. et. al., (2021) [27]	To assess variations in the frequency of reported physical back pain complaints amid the Covid-19 epidemic	English Twitter posts
Goadsby P. et. al., (2023) [34]	To study the digital profiles of people suffering from headaches and migraines using social media data	Real time data from social media posts
Caldo D. et. al., (2023) [29]	To distinguish unique digital emotional patterns associated with web pages related to back pain	Sites with medical data
Heintzelman N. et. al., (2013) [24]	To conduct a longitudinal analysis of pain in individuals with metastatic prostate cancer using NLP	Medical records data
Mullins C. et. al., (2020) [30]	To analyze pain related tweets in Ireland	Twitter posts
Yang Z. et. al., (2020) [35]	Patient phenotyping integrating token selection with deep learning	Electronic Medical Records (EMRs)

In our research, we showcase the efficacy of using deep neural networks with static embeddings to identify pain in Hindi text data. To capture the temporal connections between words and discern the nuanced patterns related to pain characteristics in text, we employ the IndicBERT model. IndicBERT, a specialised variant of the BERT model for Indic languages such as Hindi, contributes to pain detection in social media posts by comprehending the contextual nuances and linguistic patterns specific to social media phrases. IndicBERT incorporates word embeddings as an integral part of its model architecture. It recognises pain indicators and emotional language in social media text by capturing the semantic meaning of words and phrases. The incorporation of these embeddings in pain detection models improves their performance by facilitating a better understanding of word meanings within the context of pain-related text.

3. PRELIMINARIES

Previously, NLP models relied on predetermined rules and characteristics to read text, which restricts when it comes to understanding natural language. However, due to recent advances such as word embeddings and deep learning, NLP models can now learn from large amounts of data and automatically detect essential patterns in language. NLP assists in analyzing the emotions portrayed in the text so that we can determine if someone is feeling good or negative about pain. It also helps with finding essential pain-related phrases, such as "pain", "agony", "hurt" which can help us anticipate and interpret pain from text. Overall, NLP facilitates the study and interpretation of text, providing helpful information for pain prediction and management.

3.1 WORD EMBEDDINGS: Word Meanings in Vectors

Word embeddings characterize words as vectorised representations with syntactic and semantic characteristics encoded. Word embeddings are classified into two types: static and dynamic. Static embeddings are pre-trained on huge corpora and remain fixed throughout the NLP process, efficiently capturing semantic and syntactic information. Examples include GloVe [36], Crawl [37], Wikipedia [38], BioWordVec [39], GoogleNews [40], PubMed [41], BERT [42], IndicBERT [43], RoBERTa [44], DeBERTa [45], and ELECTRA [46]. Dynamic embeddings, on the other hand, such as ElMo [47], FastText [48], and XLNet [35], are created during neural network training, adjusting to the specific NLP task, and incorporating recent linguistic trends. Using NLP approaches, we undertake a thorough evaluation of static and dynamic embeddings for pain detection in Hindi text.

3.2 DEEP LEARNING MODELS

Deep learning models have revolutionized Natural Language Processing (NLP) [49], enabling the development of accurate and resilient language models. These models utilize multi-layer neural networks to learn intricate patterns and representations from textual data. Although, transformer-based models have transformed natural language processing (NLP) by developing a new architecture based on self-attention processes. Transformer Based Models encompass a range of advanced neural network architectures used in NLP tasks. Some of the popular transformer-based models include:

- **BERT** (Bidirectional Encoder Representations from Transformers): To interpret and analyze natural language, this cutting-edge NLP model utilizes pre-training and fine-tuning approaches. Its contextualized word representations have transformed a variety of language-related activities, resulting in substantial advances in NLP research and industrial applications.
- **IndicBERT**: A multilingual ALBERT model, built on the BERT architecture, tailored for NLP tasks in Indian languages. It retains linguistic nuances and produces high performance in tasks like sentiment analysis and named entity recognition by pre-training on large Indic text corpus, helping NLP research in Indic languages.
- **ELECTRA** (Efficiently Learning an Encoder that Classifies Token Replacements Accurately): An advanced NLP model leverages a distinctive pre-training technique to generate replaced tokens and a discriminator to identify them. ELECTRA outperforms earlier models thanks to this novel technique, displaying higher efficiency and efficacy in language processing and generating tasks.
- **DeBERTa** (Decoding enhanced BERT with disentangled attention): It incorporates a disentangled attention mechanism to collect both content and position information of words, and an upgraded mask decoder for reliable prediction of masked tokens. For increased generalization, the model is further refined by virtual adversarial training.
- **XLNet** (eXtreme Language Understanding Network): For bidirectional context learning, a new pre training strategy maximizes probability across multiple word order permutations. XLNet outperforms prior models like BERT by merging autoregressive formulation with Transformer-XL ideas, resulting in increased language interpretation and higher performance in a variety of NLP tasks.
- **RoBERTa** (Robustly optimized BERT approach): It incorporates modifications including larger batch sizes, more training data, and a longer training time. RoBERTa outperforms other models on a wide range of Natural Language Processing (NLP) benchmarks and tasks by fine-tuning BERT on a bigger scale, proving its usefulness in language interpretation and representation learning.
- Attention: Attention is a vital component in NLP because it enables models to focus on specific input parts, assigning variable significance, and capturing contextual relationships. Attention, when used in transformer-based models such as BERT and GPT, improves language understanding and performance in tasks such as translation and sentiment analysis.

In this study, we employed the IndicBERT model, along with performing a comparative analysis with other transformer-based models, including BERT, DeBERTa, RoBERTa, ELECTRA, XLNet.

4. METHODOLOGY

The methods used in the curation of the 'Dard-e-Shayari' dataset and the development of the 'HindiPainNet' model are described in depth in this section. Our approach is tailored to push the frontiers of Natural Language Processing (NLO), particularly focusing on Hindi language and make substantial contributions to the domain of pain detection.

4.1 "दर्द-ए-शायरी: THE LANGUAGE OF PAIN" (pronounced as *DARD-E-SHAYARI*) DATASET

To create a Dard-e-Shayari dataset, where 'dard' means 'pain' and 'shayari' means 'poetry', we collected poetries in Hindi (Devanagari script) posted by people on social media platforms, focusing on pain-related poetries that encompass words synonymous with pain, such as 'दर्द', 'दुःख', 'उदास', 'आंसू', 'निराश', and 'ग़म', (pronounced as *dard, dukh, udaas, aansu, nirash* and *gham* respectively) with the aim of detecting pain based on the psychological state expressed in the poetries. Expert annotators manually annotated the dataset and classified it into two class labels, namely, 'pain' or 'no_pain'. The accuracy of the annotations was assured by a thorough validation procedure collaboration with post-graduate students specializing in Psychology from a prominent public university in Delhi, India. Table 2 illustrates the statistical information of the dataset.

Table 2. दर्द-ए-शायरी (*Dard-e-Shayari*) Dataset Statistics

Description	**Statistics**
Overall count of instances	1000
Total count of “pain” occurrences	500
Total count of “no_pain” occurrences	500

The dataset used in this study consists of 1000 rows, evenly distributed with 500 entries of painful shayari and 500 entries of neutral or positive shayaris. It has 2 columns, with the first row serving as the header. The shayari text is included in the first column, while the second column indicates the corresponding label for each utterance, either 'pain' or 'no_pain'. Table 3 showcases a selection of sample shayaris along with their corresponding categories. This newly proposed dard-e-shayari dataset can serve as a reliable benchmark for pain detection in the Hindi language, as well as other text classification or NLP applications related to the Hindi language.

Table 3. A brief layout of दर्द-ए-शायरी (*Dard-e-Shayari*) Dataset

Shayari (Hindi)	**Label**
हमें नही आता अपने **दर्द** का दिखावा करना, बस अकेले रोते हैं और सो जाते हैं !!	pain
We do not know how to show our **pain**, Just cry alone and fall asleep !!	
हमने सोचा था की बताएंगे सब **दुःख** दर्द तुमको पर तुमने तो इतना भी न पूछा की खामोश क्यों हो	pain
We thought I will tell you all my **grief pain,** but you didn't even ask why are you quiet	
आँसू भी आते हैं और दर्द भी छुपाना पड़ता है ये जिंदगी है साहब यहां जबरदस्ती भी मुस्कुराना पड़ता है।	pain
Tears also come and **pain** also has to be hidden This is life, sir, one has to smile forcefully here too.	
ग़म सलीके में थे जब तक हम खामोश थे, जरा जुबान क्या खुली दर्द बे-अदब हो गए।	pain
Sadness was well mannered when we were silent, Just opened the mouth and the **pain** became disrespectful.	
जब कभी फुर्सत मिले, मेरे दिल का बोझ उतार दो, मैं बहुत दिनों से **उदास** हूँ, मुझे को शाम उधार दो.	pain
Whenever you get free time, help me take off my heart’s burden, I have been **depressed** for many days, lend me an evening.	
हमको मिटा सके वो ज़माने में दम नहीं, हमसे ज़माना खुद है ज़माने से हम नहीं।	no_pain
No one has the power to destroy us, World is from us, not us from the world.	
तुझसे मांगी हर दुआ एक ख्वाब बना है ए खुदा तेरी रहमतो से ही मैं इंसान बना है.!!	no_pain
Every prayer is created into a dream, Oh God, a human is made with your blessing.!!	
असत्य के आगे कभी झुकना नही और हार कर कभी रुकना नही.!!	no_pain
Never bow down to the untruth And never stop even after losing.!!	
समय हर समय को बदल देता है, बस समय को थोड़ा समय चाहिए।	no_pain
Time changes all the time, Just time needs a little time.	

In order to preprocess the Hindi posts data, which may include special characters, we employ standard text-cleaning techniques commonly utilized in Natural Language Processing (NLP). To address this, we utilize the iNLTK library [50], which offers dedicated functionalities for Indian languages, including Hindi. By leveraging this library, we can effectively manage and remove non-pertinent characters, thereby enhancing the processing of the Hindi text in our dataset.

4.2. THE HindiPainNet MODEL

The proposed method utilizes the IndicBERT model for detecting pain in social media textual data specifically in Hindi. IndicBERT, a variant of the BERT model tailored for Indic languages, offers enhanced language understanding and contextual representation. By leveraging IndicBERT, we aim to capture the nuanced patterns and linguistic cues related to pain in social media posts. This approach enables us to effectively analyze and interpret the Hindi text, facilitating accurate detection of pain-related content in social media. This specialized application of the IndicBERT model signifies a significant stride in harnessing machine learning for health informatics, leveraging the nuanced capabilities of natural language processing to interpret and provide insights into the well-being of Hindi-speaking individuals based on their social media interactions.

IndicBERT relies on contextualized word embeddings that are learned during the pre-training phase. These embeddings obtain semantic and contextual information from words in Indic languages such as Hindi. During pre-training, IndicBERT is exposed to an extensive corpus of text-data, enabling it to learn dense vector representations of words that consider the surrounding context. These conceptualized words embeddings empower IndicBERT to comprehend word meanings accurately within the specified context of pain detection in social media textual data, facilitating precise analysis and interpretation. The architecture of the proposed HindiPainNet model is depicted in Fig. 2.

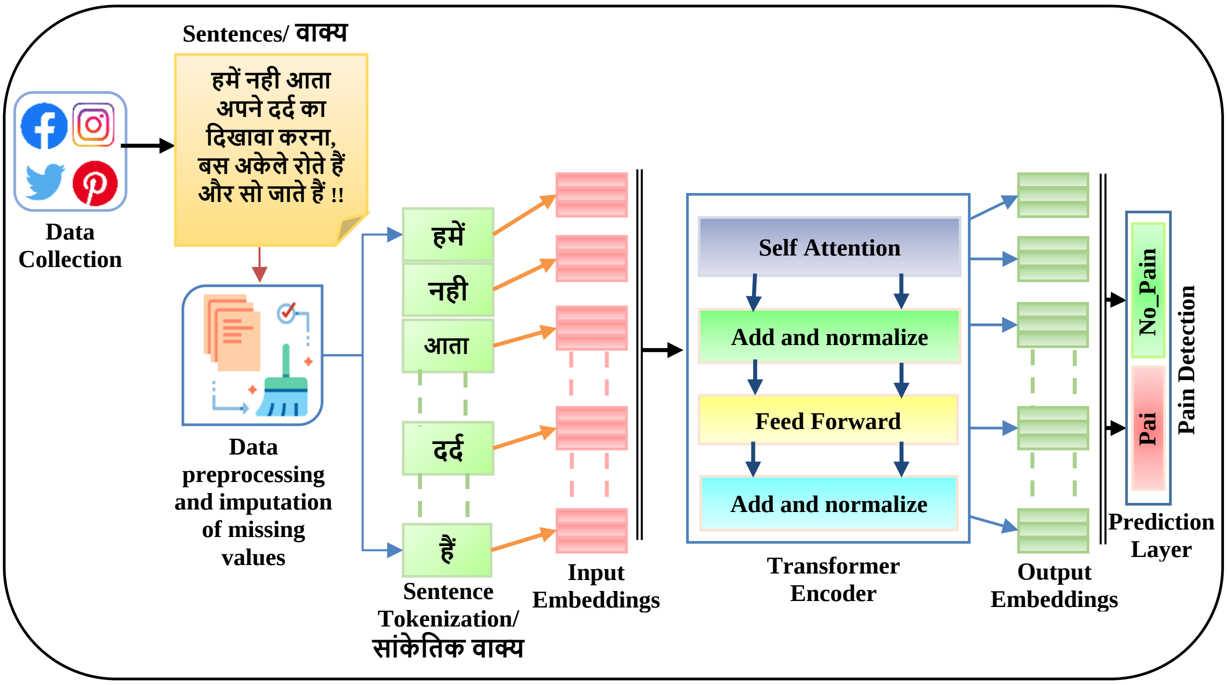

Fig.2. The HindiPainNet Model

The depicted model outlines the implementation of the IndicBERT model, tailored for detecting expressions of pain within Hindi textual data sourced from social media. The process begins with the collection of textual content from a variety of platforms, including Facebook, Instagram, Twitter, and Pinterest. Following this, the data undergoes preprocessing to clean and impute any missing values, ensuring the integrity and completeness of the dataset. Subsequently, the text is tokenized into sentences, a step that prepares it for numerical transformation. The sentences are then converted into input embeddings, numerical vectors that encapsulate the semantic properties of the text. At the heart of the model lies the Transformer encoder, consisting of self-attention mechanisms that evaluate the relevance of different words in the sentence to understand context more effectively, and feedforward networks that discern intricate patterns in the data. The processed data emerge as output embeddings, which carry the contextual information identified by the Transformer. These embeddings feed into the prediction layer, which is designed to assess the likelihood of pain expression in the analyzed text.

Incorporating cultural nuances into pain detection in Hindi texts is crucial due to the language's rich expression forms. Hindi utilizes metaphors and idioms extensively, which are deeply rooted in cultural and historical contexts. These expressions often convey emotions in a manner that is not straightforward, requiring nuanced interpretation. Regional dialects add another layer of complexity, as words and phrases can have different connotations across regions.

- **Metaphors and Idioms:** For example, a phrase like "दिल पर पत्थर रखना" (literally 'placing a stone on the heart') metaphorically expresses enduring pain or sorrow. The model must interpret such phrases beyond their literal meaning to understand the underlying emotional state.
- **Regional Variations:** Hindi's diverse dialects can affect expression. A phrase expressing sorrow in one region might be used differently in another, impacting the model's interpretation.
- **Historical and Literary Influences**: Hindi literature, influenced by various cultural epochs, uses unique expressions for pain and suffering. The model should recognize these historical styles of expression.
- **Social and Cultural Norms**: In some Hindi-speaking cultures, openly expressing pain might be less common than in others. The model must discern subtle cues in such contexts.

Our model, HindiPainNet, accounts for these variations through its training on a diverse dataset. This dataset includes classical and contemporary Hindi literature, social media content, and colloquial speech, encompassing a wide range of cultural expressions. The model leverages IndicBERT's transformer-based architecture, enabling it to understand context and subtleties in language. Additionally, we have incorporated linguistic expert input to refine the model's ability to interpret culturally specific metaphors and idioms accurately. This approach ensures that HindiPainNet is not just linguistically adept but also culturally sensitive in detecting expressions of psychological pain in Hindi texts.

It is imperative to note that transformer-based models can still be beneficial for pain detection in social media posts, even with a limited dataset. Transfer learning can be employed by pre-training the model on a larger, general-purpose dataset and fine-tuning it on the limited pain detection dataset. To enlarge the training dataset artificially, data augmentation techniques can be applied. For text data, this can involve techniques such as adding noise, synonym replacement, or random word masking. By generating additional variations of the available data, it helps in preventing overfitting and improving the model's capacity to generalise to previously unknown cases. Instead of training a transformer model from scratch, consider utilizing domain-specific language models that have been previously trained on the larger corpus of data specific to the pain domain. These models may have learned pain-related language patterns, making them well-suited for pain detection tasks even with limited data. Additionally, an ensemble of multiple transformer models can further enhance performance. While limited data poses challenges, these approaches can leverage the power of transformer models for effective pain detection.

5. RESULTS & DISCUSSIONS

The performance evaluation of HindiPainNet model encompasses a variety of measures, including accuracy, F1-score, precision, and recall. These measures jointly evaluate the model's performance, with F1-score addressing unbalanced datasets explicitly and reflecting both specificity and sensitivity. To construct a benchmark, a comparison with the most recent model is performed. Table 4 presents a comprehensive overview of the proposed model's performance results, demonstrating its effectiveness.

Table 4. Comparative analysis of the proposed HindiPainNet model with the existing state-of-the-art

Study	Model	Evaluation Metrics			
		Accuracy	Precision	F1-Score	Recall
Aggarwal A. et. al., (2023) [20]	LR, RFC, ETC	79%	-	-	-
Sawhney R. et. al., (2021) [21]	SISMO model	-	66%	64%	59%
Guo Y. et. al., (2023) [32]	BERTweet (Twitter data)	89%	88%	91%	90%
	RoBERTa (Reddit data)	90%	91%	95%	93%
Branco P. et. al., (2022) [22]	Linear Regression	67%	68%	65%	65%
Fiok K. et. al., (2021) [27]	RoBERTa, XGBoost	95.44%	95.36%	-	95.16%
Caldo D. et. al., (2023) [29]	SVM	90%	-	-	-
Yang Z. et. al., (2020) [33]	CNN	71.27%	68.29%	55.96%	48.91%
Proposed Model	IndicBERT	70.5%	68.447%	72.037%	76%

LR: Logistic Regression, **RFC**: Random Forest Classifier, **ETC**: Extra Trees Classifier, **SVM**: Support Vector Machine, **XGBoost**: Extreme Gradient Boosting, **SISMO**: Suicide Ideation detection on social media using Ordinal formulation, **BERTweet**: Bidirectional Encoder

Representations from Transformers for Twitter, **RoBERTa**: Robustly Optimized BERT pretraining approach, **CNN**: Convolutional Neural Network

Since this is pioneer research, it is crucial to acknowledge that subsequent studies have primarily concentrated on English datasets gathered from diverse sources, including Electronic Medical Records (EMRs) and social networking platforms data. The focus of these studies is on employing machine learning techniques and natural language processing to detect and analyse pain. These subsequent studies advance our knowledge of pain patterns, improve pain management tactics, and improve healthcare decision-making by using textual material available in EHRs and social media. While the current research focuses on psychological pain detection using Hindi social media textual data, it provides a valuable foundation for future research in similar domains and languages, offering insights that can be applied and extended to other low source textual datasets.

The performance of the model yielded impressive results, achieving a peak accuracy of 70.5%, having a recall of 76% indicating its strong ability to correctly identify positive instances. Additionally, the model exhibited an F1-score of 70.037%, indicating the precision of 68.44%. These outcomes highlight the effectiveness of the HindiPainNet model, equipped with IndicBERT, in capturing the crucial patterns and features required for accurate predictions in the specific task at hand. Fig. 3 displays the Confusion Matrix of the formerly discussed model.

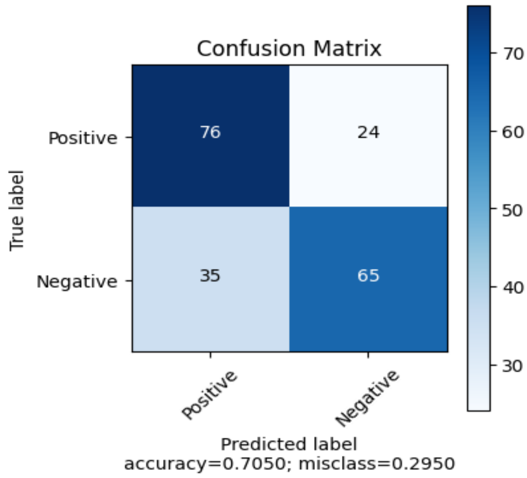

Fig. 3. Confusion Matrix of HindiPainNet model

The confusion matrix visualizes the performance of a model designed to detect pain from a set of predictions. In this context:

- True Positives (TP): The model correctly identified 76 instances of pain.
- False Negatives (FN): The model failed to detect pain in 35 cases where it was present.
- False Positives (FP): The model incorrectly flagged 24 instances as pain when there was none.
- True Negatives (TN): The model correctly recognized 65 instances where pain was not present.

The matrix indicates an accuracy of 70.5%, meaning the model correctly identified pain 70.5% of the time. The misclassification rate is 29.5%, reflecting the proportion of instances where the model's pain detection was incorrect. The matrix is a valuable tool for evaluating the effectiveness of pain detection algorithms, highlighting their diagnostic reliability.

Table 5 provides the comparative analysis of various transformer-based models, including BERT, ELECTRA, DeBERTa, XLNet, and RoBERTa with the proposed IndicBERT model. The table showcases the results obtained by each model, allowing for a comprehensive assessment of their effectiveness in the given task.

Table 5. Comparing HindiPainNet model with other Transformer-based models

Model	Accuracy	F1-score	Precision	Recall
IndicBERT	70.5%	72.04%	68.45%	76%
BERT	64%	64.36%	63.73%	65%
ELECTRA	49%	7.27%	40%	4.0%
DeBERTa	59%	61.68%	57.89%	66%
XLNet	50.5%	7.47%	57.14%	4.0%
RoBERTa	48.33%	23.5%	32%	49%

The table presents a comparative analysis of various Transformer-based models in the context of the HindiPainNet model, specifically focusing on their performance in pain detection from Hindi text. The models are evaluated based on standard classification metrics: accuracy, F1-score, precision, and recall.

- **IndicBERT**: This model leads in performance with an accuracy of 70.5%, indicating that it correctly predicts pain detection 70.5% of the time. The F1-score, which is a harmonic mean of precision and recall, is 72.04%, suggesting a balanced performance between precision and recall. Its precision is at 68.45%, meaning that when it predicts pain, it is correct about 68.45% of the time. The recall is the highest among all models at 76%, indicating that it correctly identifies 76% of all actual pain instances.
- **BERT**: Shows moderate performance with 64% accuracy and a corresponding F1-score of 64.36%. Its precision is slightly lower than its recall, suggesting that while it's somewhat reliable in its positive predictions, it doesn't capture as many actual pain instances as IndicBERT.
- **ELECTRA**: Notably underperforms in this task with the lowest accuracy of 49%. The F1-score is significantly low at 7.27%, indicating a substantial imbalance between precision and recall. The precision of 40% is relatively higher compared to its recall of just 4.0%, pointing towards a conservative model that predicts fewer instances of pain, but a large portion of those predictions are incorrect.
- **DeBERTa**: Shows a fair accuracy of 59% and an F1-score of 61.68%. The precision is 57.89%, and the recall is 66%, suggesting that it's better at identifying true cases of pain than ELECTRA but less precise than IndicBERT and BERT.
- **XLNet**: Similar to ELECTRA, it has a low accuracy of 50.5% and an F1-score of 7.47%. The precision is relatively higher at 57.14%, but like ELECTRA, the recall is very low at 4.0%. This indicates that XLNet is not effective in this task, missing many actual cases of pain.
- **RoBERTa**: Has an accuracy of 48.33% and an F1-score of 23.5%. The precision is 32% and the recall is 49%, indicating that while it identifies nearly half of the actual pain cases, its predictions are not very precise.

In summary, IndicBERT outperforms other models in all metrics, suggesting it is the most suited for detecting pain in Hindi text. BERT and DeBERTa offer moderate results, while ELECTRA, XLNet, and RoBERTa seem to struggle with this specific task. The low F1-scores of ELECTRA and XLNet, in particular, highlight the challenges these models face in balancing precision and recall for pain detection in Hindi. Fig. 4 displays the comparative analysis of the proposed model with different transformer-based models.

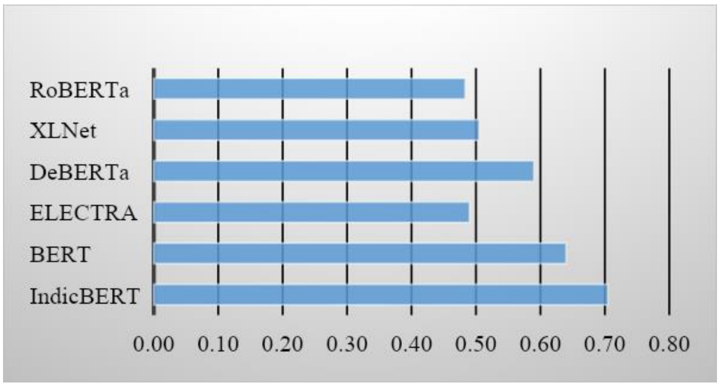

Fig. 4. Comparison of IndicBERT with other transformer-based models

While this research has made significant contributions to NLP in the Hindi language, there are some limitations to our study:

- **Dataset Bias**: The Dard-e-Shayari dataset, primarily sourced from literary and social media texts, may not comprehensively represent everyday language, potentially leading to biases in model training.
- **Model Generalizability**: While HindiPainNet shows promising results, its effectiveness outside the specific context of Hindi poetry and social media texts is not fully established.
- **Cultural Diversity**: The model may not fully capture the vast cultural and dialectical diversity within the Hindi-speaking population, impacting its accuracy across different Hindi-speaking regions.
- **Complex Expressions**: Detecting nuanced emotional expressions, especially those embedded in metaphors and idioms unique to Hindi culture, remains challenging and may not always be accurately interpreted by the model.
- **Technology Limitation**: The reliance on transformer-based models may limit the ability to capture deeper, context-specific emotional states that require human-level understanding and empathy.

6. CONCLUSION

In recent times, significant advances have been witnessed in the field of Natural Language Processing (NLP) in various areas and languages. This progress has also been extended to the analysis of the Hindi language, where the availability of comprehensive datasets and the development of efficient models have become vital for enabling sophisticated NLP applications. These datasets and models lay the groundwork for advanced language analysis tasks in Hindi, empowering researchers as well as practitioners to explore and leverage on the potential of NLP in this particular language. This study introduces two significant contributions: the establishment of the 'Dard-e-Shayari' dataset and the 'HindiPainNet' model. The 'Dard-e-Shayari' dataset is composed of Hindi social media posts that have been carefully classified into 'pain' and 'no_pain' categories, providing a useful resource for identifying the occurrences of pain based on psychological states of the social media users. This dataset, which contains 1000 instances, serves as a benchmark for pain detection from Hindi textual data. The 'HindiPainNet' model, leveraging IndicBERT architecture, effectively captures semantic relationships and contextual information. It performs well in pain detection tasks and exhibits the potential for textual data analysis. These contributions significantly advance NLP in the Hindi language and offer valuable resources for researchers and practitioners in the field.

While this research has made significant contributions to NLP in the Hindi language, there are various avenues for further research and development. Expanding the 'Dard-e-shayari' dataset by gathering additional textual data from diverse sources, such as hospitals, therapists etc. and annotations from different domains, can enhance its representativeness and applicability. Additionally, exploring data collection in other regional languages would further broaden the scope of the research. Furthermore, the 'HindiPainNet' model can be enhanced by making architectural modifications, exploring different deep learning models, or incorporating advanced techniques such as attention mechanisms or upgrading transformer-based architectures. Integrating multiple modalities, such as audio and visual data, in addition to text, can result in comprehensive pain detection models, allowing for a better understanding of human pain influenced by psychological factors. In terms of pain classification, future research could focus on developing models capable of detecting and distinguishing more complex pain features and dimensions. Lastly, the practical deployment of the HindiPainNet model in real-world applications, such as integrating them into healthcare systems, chatbots, virtual assistants, or social media platforms. This will provide insights into the effectiveness, usability, and ethical considerations of using these models in clinical and everyday contexts.

REFERENCES

[1] S. N. Raja, D. B. Carr, M. F. N. Cohen, H. Flor, S. Gibson and K. Vader, "The revised IASP definition of pain: Concepts, challenges, and compromises.," *Pain,* p. 161(9), 2020.

[2] R. B. Fillingim, "Individual differences in pain: understanding the mosaic that makes pain personal.," *Pain,* pp. 158(Suppl 1), S11, 2017.

[3] S. J. Fodeh, D. Finch, L. Bouayad, S. L. Luther, L. H. R. D. Kerns and C. Brandt, "Classiying clinical notes with pain assessment using machine learning," *Medical & biological engineering & computing,* pp. 56, 1285-1292, 2018.

[4] Q. Johnson, R. R. Borsheski, J. L. Reeves-Viets, "Pain management mini-series. Part I. A review of management of acute pain.," *Missouri Medicine,* pp. 110(1), 74-79, 2013.

[5] S. E. Mills, K. P. Nicolson, B. H. Smith, "Chronic pain: a review of its epidemiology and associated factors in population-based studies.," *British journal of anaesthesia,* pp. 123(2), e273-e283, 2019.

[6] S. J. Linton and W. S. Shaw, "Impact of psychological factors in the experience of pain.," *Physical therapy,* pp. 91(5), 700-711, 2011.

[7] M. K. Nicholas, S. J. Linton, P. J. Watson, C. J. Main and ". o. t. Flags", "Working Group Early identification and management of psychological risk factors ("yellow flags") in patients with low back pain: a reappraisal.," *Physical therapy,* pp. 91(5), 737-753, 2011.

[8] G. G. Gilam, "What is the relationship between pain and emotion? Bridging constructs and communities.," *Neuron,* pp. 107(1), 17-21, 2020.

[9] N. Corsi and L. Colloca, "Placebo and nocebo effects: the advantage of measuring expectations and psychological factors.," *Frontiers in psychology,* pp. 8, 308, 2017.

[10] N. Vadivelu, A. M. Kai, G. Kodumudi, K. Babayan, M. Fontes and M. M. Burg, "Pain and psychology—a reciprocal relationship.," *Ochsner Journal,* pp. 17(2), 173-180, 2017.

[11] F. Lena, M. Pappaccogli, M. Santilli, M. Torre, N. Modugno and A. Perrotta, "How does semantic pain and words condition pain perception? A short communication.," *Neurological Sciences,* pp. 43(1), 691-696, 2022.

[12] A. Arora, P. Chakraborty, M. P. S. Bhatia and P. Mittal, "Role of emotion in excessive use of Twitter during COVID-19 imposed lockdown in India.," *Journal of Technology in Behavioral Science,* pp. 6, 370-377, 2021.

[13] Z. Zeng, Y. Deng, X. Li, T. Naumann and Y. Luo, "Natural language processing for EHR-based computational phenotyping.," *IEEE/ACM transactions on computational biology and bioinformatics,* pp. 16(1), 139-153, 2018.

[14] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks.," *IEEE 60th international midwest symposium on circuits and systems (MWSCAS),* pp. 1597-1600, 2017.

[15] L. A. Carlson and W. M. Hooten, "Pain—linguistics and natural language processing.," *Mayo Clinic Proceedings: Innovations, Quality & Outcomes,* pp. 4(3), 346-347, 2020.

[16] P. J. Tighe, B. Sannapaneni, R. B. Fillingim, C. Doyle, M. Kent, B. Shickel and P. Rashidi, "Forty-two million ways to describe pain: topic modeling of 200,000 pubmed pain-related abstracts using natural language processing and deep learning-based text generation.," *Pain Medicine,* pp. 21(11), 3133-3160, 2020.

[17] G. L. Engel, "The need for a new medical model: a challenge for biomedicine.," *Science,* pp. 196(4286), 129-136, 1977.

[18] E. Borelli, D. Crepaldi, C. A. Porro and C. Cacciari, "The psycholinguistic and affective structure of words conveying pain," *PloS one,* p. 13(6), 2018.

[19] A. Gumaei, M. S. Al-Rakhami, M. M. Hassan, V. H. C. De Albuquerque and D. Camacho, "An effective approach for rumor detection of Arabic tweets using extreme gradient boosting method.," *Transactions on Asian and Low-Resource Language Information Processing,* pp. 21(1), 1-16, 2022.

[20] A. Aggarwal, S. Rai, S. Giorgi, S. Havaldar, G. Sherman, J. Mittal and S. C. Guntuku, "A Cross-Modal Study of Pain Across Communities in the United States," *Companion Proceedings of the ACM Web Conference 2023,* pp. 1050-1058, 2023.

[21] R. Sawhney, H. Joshi, S. Gandhi and R. R. Shah, "Towards ordinal suicide ideation detection on social media.," *Proceedings of the 14th ACM international conference on web search and data mining,* pp. 22-30, 2021.

[22] P. Branco, S. Berger, T. Abdullah, E. Vachon-Presseau, G. Cecchi and A. V. Apkarian, "Predicting placebo analgesia in chronic pain patients using natural language processing: a preliminary validation study.," *Pain,* pp. 10-1097, 2022.

[23] E. Pierson, D. M. Cutler, J. Leskovec, S. Mullainathan and Z. Obermeyer, "An algorithmic approach to reducing unexplained pain disparities in underserved populations.," *Nature Medicine,* pp. 27(1), 136-140, 2021.

[24] N. H. Heintzelman, R. J. Taylor, L. Simonsen, R. Lustig, D. Anderko, J. A. Haythornthwaite and G. S. Bova, "Longitudinal analysis of pain in patients with metastatic prostate cancer using natural language processing of medical record text.," *Journal of the American Medical Informatics Association,* pp. 20(5), 898-905, 2013.

[25] A. Sarker, G. Gonzalez-Hernandez, Y. Ruan and J. Perrone, "Machine learning and natural language processing for geolocation-centric monitoring and characterization of opioid-related social media chatter.," *JAMA network open,* p. 2(11), 2019.

[26] K. D. Craig and N. E. MacKenzie, "What is pain: Are cognitive and social features core components?," *Paediatric and Neonatal Pain,* pp. 3(3), 106-118, 2021.

[27] K. Fiok, W. Karwowski, E. Gutierrez, M. Saeidi, A. M. Aljuaid, M. R. Davahli and B. D. Sawyer, "A study of the effects of the COVID-19 pandemic on the experience of back pain reported on Twitter® in the United States: a natural language processing approach," *International Journal of Environmental Research and Public Health,* pp. 18(09), 4543, 2021.

[28] A. Kumar and V. H. C. Albuquerque, "Sentiment analysis using XLM-R transformer and zero-shot transfer learning on resource-poor indian language.," *Transactions on Asian and Low-Resource Language Information Processing,* pp. 20(5), 1-13, 2021.

[29] D. Caldo, S. Bologna, L. Conte, M. S. Amin, L. Anselma, V. Basile and G. De Nunzio, "Machine learning algorithms distinguish discrete digital emotional fingerprints for web pages related to back pain.," *Scientific Reports,* pp. 13(1), 4654, 2023.

[30] C. F. Mullins, J. Lane and T. O'Connor, "Sharing the pain: an observational analysis of Twitter and pain in Ireland.," *Regional Anesthesia & Pain Medicine,* pp. 45(8), 597-602, 2020.

[31] H. Deng, Q. Wang, D. P. Turner, K. E. Sexton, S. M. Burns, M. Eikermann and T. T. Houle, "Sentiment analysis of real-world migraine tweets for population research.," *Cephalalgia Reports,* p. 3, 2020.

[32] Y. Guo, S. Rajwal, S. Lakamana, C. C. Chiang, P. C. Menell, A. H. Shahid and A. Sarker, "Generalizable Natural Language Processing Framework for Migraine Reporting from Social Media.," *AMIA Summits on Translational Science Proceedings,* p. 261, 2023.

[33] Z. Yang, M. Dehmer, O. Yli-Harja and F. Emmert-Streib, "Combining deep learning with token selection for patient phenotyping from electronic health records.," *Scientific reports,* pp. 10(1), 1432, 2020.

[34] P. Goadsby, E. Ruiz de la Torre, L. Constantin and C. Amand, "Social Media Listening and Digital Profiling Study of People With Headache and Migraine: Retrospective Infodemiology Study.," *Journal of Medical Internet Research,* p. 25, 2023.

[35] Z. Yang, Z. Dai, Y. Yang, J. Carbonell, R. R. Salakhutdinov and Q. V. Le, "Xlnet: Generalized autoregressive pretraining for language understanding.," *Advances in neural information processing systems,* p. 32, 2019.

[36] J. Pennington, R. Socher and C. D. Manning, "Glove: Global vectors for word representation.," *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP),* pp. 1532-1543, 2014.

[37] P. K. Sarma, Y. Liang and W. A. Sethares, "Domain adapted word embeddings for improved sentiment classification.," *arXiv preprint arXiv: 1805.04576,* 2018.

[38] E. Sheehan, C. Meng, M. Tan, B. Uzkent, N. Jean, M. Burke and S. Ermon, "Predicting economic development using geolocated wikipedia articles.," *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining,* pp. 2698-2706, 2019.

[39] Y. Zhang, Q. Chen, Z. Yang, H. Lin and Z. Lu, "BioWordVec, improving biomedical word embeddings with subword information and MeSH.," *Scientific data,* pp. 6(1), 52, 2019.

[40] Y. Wang, S. Liu, N. Afzal, M. Rastegar-Mojarad, L. Wang, F. Shen and H. Liu, "A comparison of word embeddings for the biomedical natural language processing.," *Journal of biomedical informatics,* pp. 87, 12-20, 2018.

[41] B. Chiu, G. Crichton, A. Korhonen and S. Pyysalo, "How to train good word embeddings for biomedical NLP.," *Proceedings of the 15th workshop on biomedical natural language processing,* pp. 166-174, 2016.

[42] G. Wiedemann, S. Remus, A. Chawla and C. Biemann, "Does BERT make any sense? Interpretable word sense disambiguation with contextualized embeddings.," *arXiv preprint arXiv: 1909. 10430,* 2019.

[43] D. Kakwani, A. Kunchukuttan, S. Golla, N. C. Gokul, A. Bhattacharyya, M. M. Khapra and P. Kumar, "IndicNLPSuite: Monolingual corpora, evaluation benchmarks and pre-trained multilingual language models for Indian languages.," *Findings of the Association for Computational Linguistics: EMNLP 2020,* pp. 4948-4961, 2020.

[44] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach.," *arXiv preprint arXiv:1907.11692,* 2019.

[45] P. He, X. Liu, J. Gao and W. Chen, "Deberta: Decoding-enhanced bert with disentangled attention.," *arXiv preprint arXiv:2006.03654,* 2020.

[46] K. Clark, M. T. Luong, Q. V. Le and C. D. Manning, "Electra: Pre-training text encoders as discriminators rather than generators.," *arXiv preprint arXiv:2003.10555,* 2020.

[47] K. Ethayarajh, "How contextual are contextualized word representations? Comparing the geometry of BERT, ELMo, and GPT-2 embeddings.," *arXiv preprint arXiv:1909.00512,* 2019.

[48] B. Athiwaratkun, A. G. Wilson and A. Anandkumar, "Probabilistic fasttext for multi-sense word embeddings.," *arXiv preprint arXiv:1806.02901,* 2018.

[49] M. M. Lopez and J. Kalita, "Deep Learning applied to NLP.," *arXiv preprint arXiv:1703.03091,* 2017.

[50] G. Arora, "inltk: Natural language toolkit for indic languages.," *arXiv preprint arXiv, 2009.12534.,* 2020.