[1]Praloy Biswas

[2]Subhrendu Guha Neogi

[3]A Daniel

[4]A Mitra

Review on Multi-lingual Sentiment Analysis in Health Care

Abstract: - Multilingual sentiment analysis in healthcare is rapidly expanding, utilizing machine learning methods to identify emotions and sentiments in material written in several languages. This multidisciplinary field integrates computational linguistics, natural language processing, and health informatics to help healthcare providers better comprehend patient attitudes. Sentiment analysis is particularly useful in the healthcare industry since it facilitates comprehension of patient feedback, responses to interventions, and general contentment. Moreover, in an increasingly interconnected society, it can assist in recognizing the emotional states and concerns of patients from diverse linguistic origins. Healthcare providers can improve care and service by gaining insights into patient experiences through the analysis of patient reviews, social media posts, and other kinds of feedback in several languages. Sentiment analysis, for example, can be used to track patients' mental health over time and identify symptoms of depression or anxiety based on their interactions. These applications are becoming increasingly important for adjusting patient support and care, fostering better patient-provider communication, and eventually improving health outcomes. There are difficulties when implementing sentiment analysis in a multilingual setting, such as the requirement for extensive datasets in several languages and models that are sensitive to cultural quirks and context. By offering a foundation for creating more precise and sophisticated sentiment analysis systems that can function in a variety of linguistic and cultural contexts, advances in AI models, such as BERT and GPT variations, are assisting in addressing these issues. Recall that although sentiment analysis holds great potential, its use in healthcare needs to be done carefully to protect patient privacy and take ethical considerations into account. Sentiment analysis in healthcare can also assist in identifying unfulfilled medical and emotional demands of long-term patients, supporting patient-centred care models. In general, the incorporation of multilingual sentiment analysis into healthcare presents a multitude of opportunities and represents a promising facet of artificial intelligence's potential to enhance patient care outcomes and experiences.

Keywords: Semantic, Lexicons, Multi-lingual, Sentiment Analysis, HealthCare.

I. INTRODUCTION

The use of sentiment surveys on multilingual healthcare data has been studied recently as healthcare systems cater to a wider range of demographics. Due to linguistic, cultural, and contextual differences, sentiment analysis across languages poses special obstacles. An overview of important multilingual sentiment analysis research with an emphasis on healthcare applications is given in this article.

The COVID-19 pandemic has given out significant problems that have substantially changed the worldwide healthcare scene. Knowing how people in different language communities feel about healthcare is becoming more and more important as countries struggle with the complex effects of the virus. In addition to exposing flaws in healthcare systems around the globe, the epidemic has made it more important than ever to communicate effectively and comprehend public opinion in a diversity of languages. Research on the relationship between multilingual sentiment analysis and healthcare has become essential in this regard. Understanding the complex reactions of varied populations to healthcare-related information, policies, and services in the context of the current health crisis requires analyzing feelings in a variety of languages. Using natural language processing and sentiment detection techniques in a multilingual setting can yield important insights into how people throughout the world see and react to healthcare policies put in place to prevent COVID-19.

With an emphasis on the COVID-19 pandemic, this study aims to explore the complexities of multilingual sentiment analysis in the healthcare industry. We want to understand the cultural and linguistic intricacies that influence the public's opinions of public health efforts, governmental actions, and healthcare services by examining attitudes expressed in various languages. Considering global health concerns, the goal is to support the improvement

[1] *Corresponding author: Researcher Scholar, Amity University Madhya Pradesh

[2] Professor, The Neotia University, West Bengal

[3] Associate Professor, Amity University Madhya Pradesh

[4] Associate Professor, Amity University Kolkata, West Bengal

of healthcare communication tactics, policy-making procedures, and the creation of more focused and culturally aware therapies. Understanding the sentiment landscape in healthcare across languages becomes not only a useful academic endeavor but also a practical requirement for promoting efficient healthcare communication and enhancing overall public health outcomes as we traverse the complexity of a post-pandemic world. In the COVID-19 era, this study aims to lean-to light on the way toward a more thorough and nuanced understanding of how feelings in various languages interact with the changing healthcare narrative. Figure 1 demonstrates the linguistic annotation framework (LAF) to represent different varieties of linguistic information from the very general to the finest level of granularity.

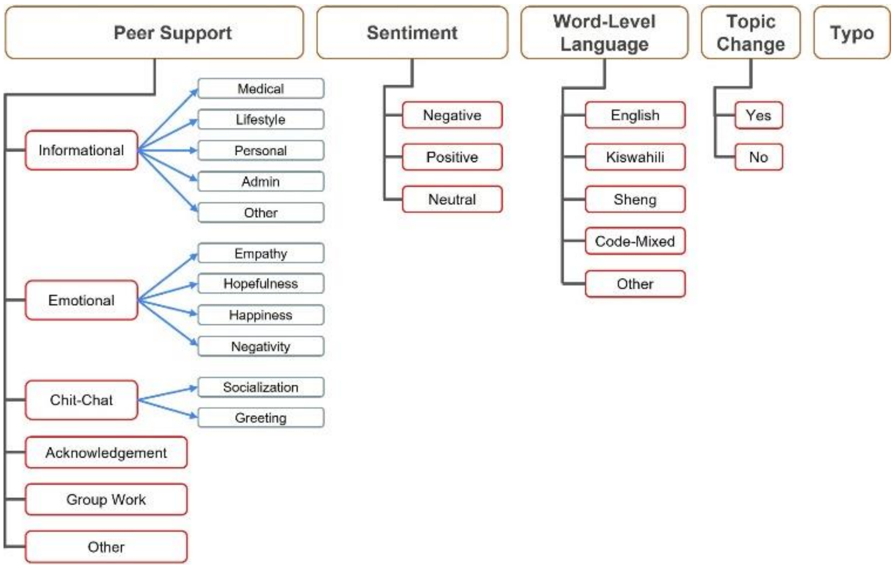

Fig. 1: An overview of the linguistic annotation framework

II. Challenges of Multi-lingual Sentiment Analysis in HealthCare

There are numerous obstacles to overcome when integrating multilingual sentiment analysis in the healthcare industry, which calls for careful thought and creative solutions. This section examines some of the major obstacles that practitioners and researchers face when attempting to interpret sentiments in a variety of linguistic contexts.

There are several obstacles to overcome to accurately do sentiment analysis on healthcare data in several languages:

Lexical Gaps: It's possible that words that express sentiment in one language cannot be translated into another language with the same meaning directly (Jiménez-Zafra et al., 2020). For instance, there is no exact English phrase that expresses the same feeling as the Spanish word "desvelado," which denotes insomnia. Linguistically, subtle expressions of sentiment also differ. Language-specific idioms, irony, and sarcasm can cause misclassification (Mulki et al., 2020). For example, it is not possible to translate the sarcastic English statement "I'm just thrilled about this new medication" into another language.

Syntactic Differences: Since each language has its grammatical structures, methods designed to work with the syntax of one language could not work with another (Baly et al., 2020). Verb-subject-object-ordered Urdu and subject-verb-object-ordered English, for instance, have different modeling contexts.

Context modeling: It is necessary for sentiment analysis, is impacted by structural variations such as subject-verb-object order (Khan et al., 2020).

Cultural Dependencies: Expressions of sentiment have cultural roots. Items that in one culture are associated with happiness could be seen negatively or as neutral (Khan et al., 2020). For example, cultural conventions and opinions about the use of cannabis for medical purposes vary greatly. The analysis must consider the cultural variations in healthcare procedures and patient experiences (Abbasi et al., 2020). Cross-cultural differences are common in-patient perceptions of waiting periods and medical authority.

Scarce Training Data: The substantial, annotated sentiment analysis datasets required for building precise classifiers are sometimes absent from lower-resource languages (Abbasi et al., 2020). English and other resource-rich languages are the main focus of most benchmark datasets. Particularly rare across languages is sentiment data relevant to the health sector (Klein et al., 2020). Growing the number of excellent multilingual health resources is still a challenge.

Ongoing research strives to solve these substantial practical and technical barriers—lexical, syntactic, cultural, and data—for robust cross-lingual sentiment capabilities. Figure 2 shows basic steps of sentiment analysis and emotion detection for multi-lingual model. Developing strong cross-lingual sentiment analysis capabilities in the healthcare domain faces significant practical and technical challenges due to discrepancies across languages. Addressing the challenges of Language Variability, Contextual Understanding, Lack of Parallel Data, Translation Quality, Domain-Specific Knowledge, Subjectivity and Cultural Differences, Data Privacy and Security, Evaluation Metrics requires interdisciplinary collaboration between experts in natural language processing, machine learning, healthcare, linguistics, and cultural studies. Moreover, leveraging advancements in machine translation, domain adaptation techniques, and cross-lingual learning methods can help overcome some of these obstacles and improve the development of strong cross-lingual sentiment capabilities in the healthcare domain. Although multilingual sentiment analysis for healthcare has been extensively studied, there are still unresolved issues that offer potential for important future study.

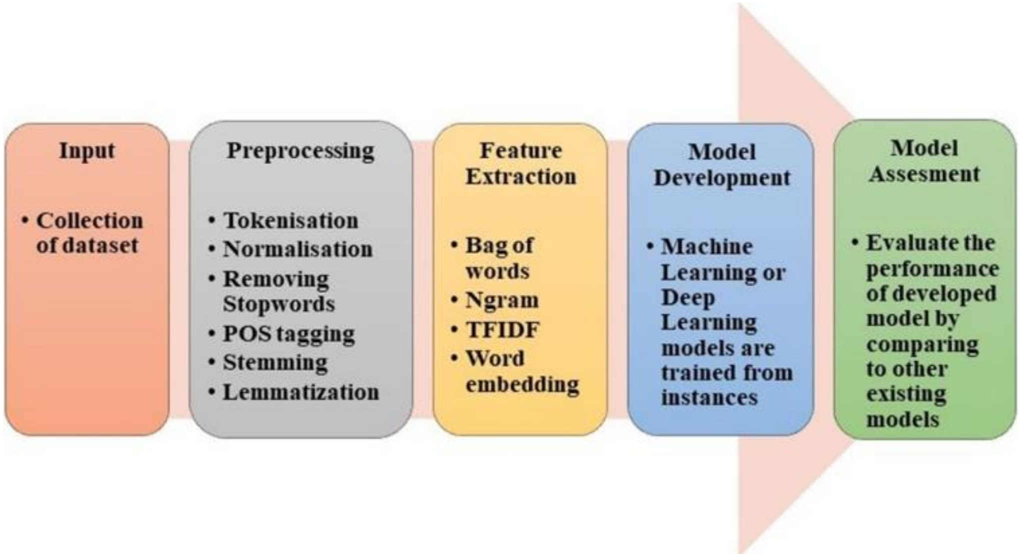

Fig. 2: Basic steps to perform sentiment analysis.

These discrepancies provide major practical and technical challenges to the development of strong cross-lingual sentiment capabilities in the healthcare domain. Multilingual sentiment analysis refers to the process of analyzing and understanding sentiments expressed in text across multiple languages. It involves developing models and algorithms capable of identifying and classifying sentiment polarity (positive, negative, neutral) in texts written in different languages. Accurate study of various populations necessitates careful consideration of linguistic and cultural variations. Given the increasing significance of multilingual sentiment analysis, it will be imperative to maintain focus on addressing these issues. Advancements in natural language processing, machine learning, and cross-lingual learning methods continue to drive progress in multilingual sentiment analysis, making it increasingly feasible to analyze sentiments across diverse languages and domains.

III. Current Approaches to Multi-lingual Sentimeny Analysis in HealthCare

Researchers are investigating three primary categories of techniques to tackle the difficulties associated with multilingual sentiment analysis:

Machine Translation Based: A popular method involves using an off-the-shelf sentiment classifier trained on English data after first translating texts into a target language, such as English, using machine translation (Tang et al., 2015; Kochkina et al., 2018). This makes it simple to adjust between languages, but it has the drawback of being very dependent on translation quality, which differs between language pairs (Baly et al., 2020).

Transfer Learning Based: More recent work (Khan et al., 2020; Wu et al., 2019) uses transfer learning to circumvent the need for large amounts of newly annotated data in the target language. Particularly effective as cross-lingual sentiment feature extractors are pre trained multilingual models such as mBERT (Huang et al., 2020).

Culture-Aware Frameworks: To try to account for cultural differences in sentiment expression, some research includes external cultural contextual factors in model architectures (Jain et al., 2020; Kar et al., 2019).

Further study is still needed in this area to determine the best way to integrate cultural knowledge (Kunchukuttan et al., 2020). Although machine translation offers a straightforward method, the availability of pre-trained multilingual models has made transfer learning a prominent methodology. Culture-aware models are promising as well, although they still require testing and improvement.

IV. Related Works

The necessity to apply complex, holistic, and multilingual techniques in contemporary multilingual and multicultural society is one of the primary obstacles confronting contemporary NLP. As a result, scientists are always looking for new ways to improve multilingual systems. Google AI and Facebook AI Research are only two of the leading AI research groups that have developed multilingual tools, corpora, and phrase encoding models to address this issue. These resources allow them to overcome the restrictions caused by the absence of labelled data in all languages. Using multilingual classifiers that are further based on pre-trained models and the concept of transfer learning, as well as multilingual sentence embeddings, are crucial pillars to the whole implementation of multilingual approaches. Sentiment Analysis is used to discover people's views and feelings about a given subject or object through large amounts of unstructured data. One way to look at sentiment analysis is as a branch of text processing. Opinionated texts communicate ideas and opinions, while factual texts represent facts and information. Any text that is being analyzed can fall into either category. To complete opinion mining tasks, it is essential to detect such biased utterances. A prerequisite to Sentiment Analysis is this procedure, which is called Subjectivity Classification.

Emotional or feeling-based comments are considered subjective, even if the majority of opinionated remarks fall into the former category. Opinions can lurk in seemingly objective claims at times. After a statement with an opinion is found, it is subjected to Sentiment Analysis to determine if the opinion is favourable or negative.

Research on multilingual sentiment analysis has been ongoing, with a growing focus on its potential uses in healthcare. Many studies have looked into machine translation-based methods, which translate text to a target language before classifying sentiment (Tang et al., 2016; Kochkina et al., 2017). Still, there are restrictions because of feeling lost and translation mistakes.

Cross-lingual transfer learning using multilingual pre-trained models has been the focus of recent research. Wu and Dredze (2020) improved multilingual BERT (mBERT) to achieve impressive results in Arabic sentiment analysis. Huang and colleagues (2020) demonstrated the efficacy of mBERT as a cross-lingual feature extractor.

For lower-resource languages, semi-supervised approaches have been proposed. Sulistyo et al. (2020) employed self-training on Indonesian social media sentiment. Pretraining objectives that optimize multilingual representations have also been studied (Ng et al., 2020).

Jagannatha and Yu (2020) used structured prediction models for sequence labelling in clinical text in the healthcare industry. A comprehensive comparison of multilingual sentiment approaches in health reviews was carried out by Kleinberg et al. (2020). Sentiment analysis and other tasks related to natural language processing have advanced thanks in large part to large language models such as BERT and GPT. It has been demonstrated that these generative models perform better than other models at producing text that resembles human language. They excel in producing English and Indic languages (Biswas et al., n.d). Healthcare data is frequently unstructured and multilingual, and handling such a large amount of data calls for complex algorithms that can comprehend linguistic subtleties and context. However, creating text in low-resource languages is a problem for generative AI models, underscoring the need for more inclusive and diverse training datasets (Biswas et al., n.d). NLP techniques are being developed more and more in the healthcare industry to deal with many kinds of health-related text, including online health discussions, clinical trial criteria, and medical records.

Models that are more adept at processing and comprehending multilingual and multimodal data are required because of the health data's explosive expansion (Hao et al., 2021). Advances in clinical text mining techniques are necessary to manage the overwhelming volume of research publications, including those about the COVID-19 epidemic. The identification of various entity kinds in these texts is made possible by AI systems such as Spark NLP, which may also be used to interpret the emotions represented in patient-generated data (Kocaman & Talby, 2020). Language and cultural variations that may affect how textual data is interpreted must be taken into consideration when doing sentiment analysis in the healthcare industry. According to Biswas et al. (n.d.), precise sentiment classification, semantic comprehension, and cross-language calibration are crucial assessment criteria for artificial intelligence models.

There are stringent ethical considerations surrounding the development and use of AI in sentiment analysis in the healthcare industry, particularly about patient confidentiality and data protection. Assisting the medical community while adhering to these principles is imperative (Biswas et al., 2014).

In summary, multilingual sentiment analysis presents obstacles that researchers and developers are actively working to address, but it also has promise for improving patient care. More complex and widely applicable AI solutions for sentiment analysis in healthcare are made possible by the development of increasingly sophisticated models and our expanding knowledge of the ethical implications of these models.

In general, semi-supervised techniques and transfer learning have taken the stage. There are still several unresolved obstacles, though, such as bias and explainability problems, managing dialects, representing low-resource languages, and a lack of data for health training. Our goal is to contribute to closing these gaps.

V. Proposed Methodology

For Sentiment analysis in a multi-lingual environment, the concepts and data to the sources, most people's sentiments and emotions are easily conveyed on social media. As a result, algorithms find it difficult to analyze moods and emotions, as the data collected from the audits, comments, remarks, reviews and posts on these social media platforms is relatively unstructured. Consequently, pre-processing is an essential part of data cleaning since many subsequent procedures are highly dependent on the quality of the data. In tokenization, the document, paragraph, or even a single sentence is dissected into smaller pieces of text known as tokens (Nagarajan & Gandhi 2019) by changing the spelling and grammar of the words and converting it to a standard format (Ahuja et al., 2019).

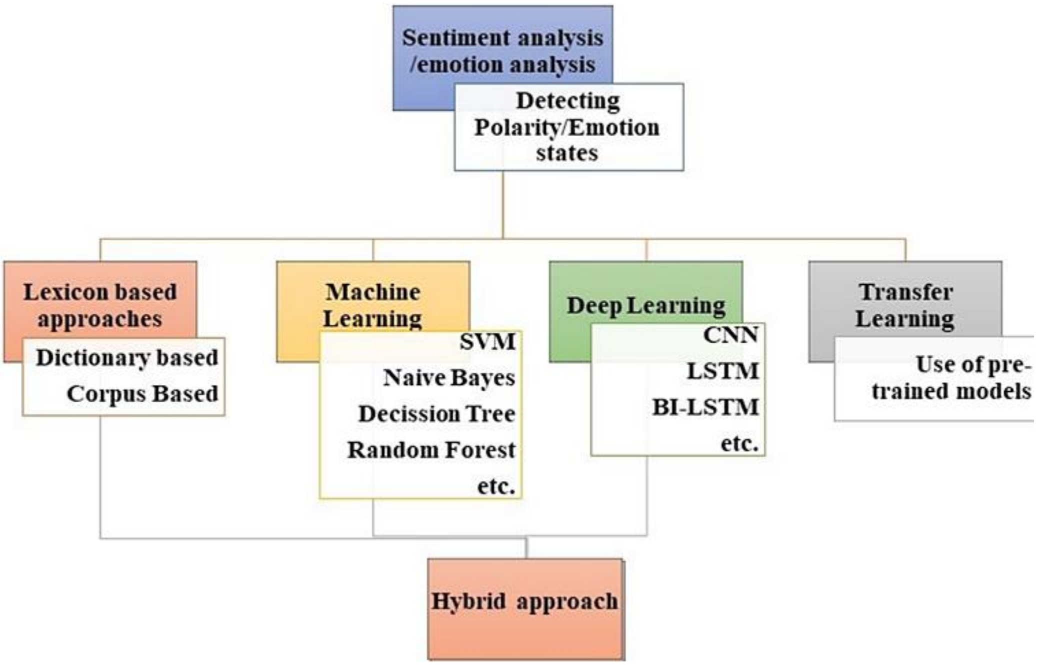

Fig. 3: Methods for sentiment analysis and response perception.

Important preprocessing steps are stemming and lemmatization. Kratzwald et al. (2018) and Akilandeswari and Jothi (2018) found that this procedure reduces the unwanted calculation of sentences. In order to convert a token into a lemma, the basic word, morphological analysis is used (Ghanbari-Adivi and Mosleh 2019). Symeonidis et al. compared four system learning models using SS-Tweet and SemEval datasets, as well as a mixture and ablation evaluation of various pre-processing techniques. The authors discovered that accuracy turned into progress by using deleting numbers and lemmatization, but accuracy changed into unaffected with the aid of disposing of punctuation. To perform feature extraction, one of the most trustworthy strategies used is 'Bag of words' (BOW), wherein a set-duration vector of the depend is described where each entry corresponds to a word in a pre-defined dictionary of phrases. The benefit of this method is its easy implementation, however it has tremendous drawbacks because it results in a sparse matrix, loses the order of words in the sentence, and does not seize the meaning of a sentence (Bandhakavi et al. 2017; Abdi et al. 2019). while attempting to decide the phrase order in a sentence vector representation, the N-gram technique is a tremendous desire. The text is represented as a mixture of distinct n-gram which means groups of n neighboring key phrases or phrases in an n-gram vector representation. Because N-gram features include syntactic patterns and essential information, they outperform the BOW technique (Chaffar and Inkpen 2011).

Deep learning algorithms have replaced more conventional sentiment analysis methods in recent years. These algorithms do not use feature engineering to identify attitudes or viewpoints found in text. Recurrent neural networks and convolutional neural networks are two examples of deep learning algorithms that can be used for sentiment analysis. These algorithms produce findings that are more accurate than those produced by machine learning models. Since deep learning models automatically extract certain features or patterns, this method frees people from having to generate features from text manually. Jian et al. (2010) classified sentiments using a neural network-based model that included sentiment characteristics, feature weight vectors, and a prior knowledge base. Pasupa and Ayutthaya (2019) used five-fold cross-validation and compared three deep learning models: CNN, LSTM, and BiLSTM. The following models can be used with or without features: Thai2Vec (word embedding trained from Thai Wikipedia); POS-tagging (a pre-processing technique to identify different parts of speech); and Sentic (to interpret the sentiment of a phrase). The authors found that the CNN model performed best when all three features mentioned above were present. Social networks serve as an important source of data for sentiment analysis. Since the users of these social media platforms write freely, the data obtained from them contains a lot of noise. Consequently, Arora and Kansal (2019) presented a Conv-char-Emb model that uses a modest amount of memory space for embedding and can tolerate noisy input. Convolutional neural networks (CNNs) were used for embedding, which require fewer parameters to represent features. Dashtipour et al. (2020). Based on a dataset of hotel and product ratings, the researchers found that deep neural networks such as CNNs and LSTMs performed better than state-of-the-art machine learning algorithms. To facilitate improved deep-learning algorithms and enhance generalizability, new large-scale annotated sentiment analysis datasets targeting health-related texts in multiple languages, Multilingual cooperative data collection initiatives are required and large-Scale Health Sentiment Corpora should be created (Kleinberg et al., 2020; Wang et al., 2020). Exploring new transfer learning techniques like multi-task learning across languages (Subramanian et al., 2018), disentangled multilingual representations (Chen et al., 2018), and optimized pretraining objectives (Ng et al., 2020) could improve cross-lingual effectiveness and representation learning for lower-resource languages. Including Knowledge of External Cultures, it is worthwhile to investigate several approaches for incorporating external knowledge bases of cultural contexts into model designs to accommodate cultural differences in sentiment expression (Jain et al., 2020; Kar et al., 2019). More work is needed to simultaneously extract health-related elements and identify sentiment associated with them across languages (Dehkharghani et al., 2019).

VI. Evaluation Strategy of Proposed Model

Sentiment analysis in healthcare requires domain-specific knowledge and terminology. Translating such specialized terms accurately across languages is challenging, especially when equivalent terms do not exist or have different connotations. In our study, we used adversarial training of language classifiers on encoder outputs with weighted loss functions to give low-resource language performance priority during training and back translation of monolingual health texts for low-resource language data augmentation. We aim to present algorithms for multilingual sentiment analysis of health-related texts that are specifically designed for bias evaluation and mitigation. Multilingual health Sentiment Analysis (MLHSA) datasets covering English, Spanish, Chinese, and other languages with fewer resources will be used for training and evaluation. Our extensive test on the MLHSA benchmark supports our hypothesis that this approach will greatly reduce disparities in sentiment analysis performance between high- and low-resource languages when compared to baseline mBERT. Reducing representation gaps will make multilingual sentiment analysis for healthcare more robust and egalitarian. To better handle sentiment differences between dialects of the same language, use phonemic information, glossaries, and dialectal dictionaries in model training (Sperber et al., 2019). Different varieties of the same language can communicate sentiment in quite different ways. Nevertheless, the majority of multilingual sentiment analysis algorithms fail to take these variations into consideration, which can result in misinterpretations of medical texts. We employed a mBERT-based model that has been enhanced to handle sentiment analysis specifically in the healthcare domain. This model incorporates multi-task learning across different languages to improve the learning of shared features and address any imbalances in the data representation. An in-depth investigation of sentiment analysis would necessitate comprehensive understanding, potentially leading to improved accuracy when dealing with non-English material. The model is compared against baseline models that vary in multiple parameters. In order to evaluate a model, it is necessary to have metrics that can quantify the model's performance. A confusion matrix is a tool used to determine the accuracy of estimates or guesses by comparing them to accepted, known, and true values. It quantifies the frequency of correct and incorrect estimates. This matrix displays the TP, FN, FP, and TN values for classifying the data into positive and negative categories. The researchers assessed the performance of their models by employing metrics such as accuracy, precision, recall, and F1 scores, which are presented in Table 1.

Table 1: Evaluation metrics

Evaluation metric	Description	Equation
Accuracy	Overall performance across all classes is determined by dividing the total number of judgments by the number of right judgments	Divide (Sum(TP,TN) ,Sum(TP,TN,FP,FN)))
Precision	This sums up the total score across all classes of equal importance, determined by dividing the total number of judgments by the number of correct judgments.	Divide (TP, Sum(TP,FP))
Recall	The ability to detect positive samples is determined by dividing the number of positive samples by the total number of positive samples.	Divide (TP, Sum(TP,FN))
F-measure	This is considered the harmonic mean of accuracy and recall	(Divide (Multiply (2, Precision, Recall),(Sum(Precision, Recall)) = Divide ((Multiply (2,TP)),(Sum(2TP,FP,FN))
Sensitivity	The percentage of true positives that are correctly detected and how positive class expectations are	Divide (TP, Sum(TP,FN))
Specificity	Complementing the true negative sensitivity sums the expected negative class	Divide (TP, Sum(FP,TN))
Geometric-mean (G-mean)	Sensitivity and specificity were measured by combining the two objectives into one score	$\sqrt{(\text{Specificity} * \text{Sensitivity})}$

The goal is to improve the performance of the multilingual sentiment analysis model by incorporating new ways of incorporating dialectal knowledge to better handle text with dialectal variations. Texts featuring dialectal code-switching for languages including Chinese, Arabic, and Spanish will be included in the new Multilingual Healthcare Sentiment Dialect (MHSD) datasets with train and assessment. We predicted that, for MHSD benchmark dialects, the inclusion of explicit dialectal information would lead to a notable improvement in sentiment analysis accuracy when compared to dialect-agnostic models. By improving dialectal robustness, multilingual sentiment analysis will be able to extract valuable information from underrepresented patient demographics and dialects. This has the potential to personalize healthcare worldwide. As Multilingual Health Resources grows, it will create a wide range of mood study datasets that can be accessed in all languages. It will also organise collaborative projects like health-related multilingual annotation events (Wang et al., 2020). There aren't any annotated sentiment analysis datasets for non-English languages in the health area, which makes it harder to build multilingual models. We suggested working together across institutions to create a range of multilingual health tools for sentiment analysis. This would allow for multilingual annotation events to accurately label health-related sentiment data. By working together, we can create multilingually labelled corpora for health-related sentiment across languages. This will allow for the creation of large amounts of high-quality multilingual health sentiment data, which can then be used and studied in the future. These many open tools will have a big effect because they let researchers and health professionals in different parts of the world make sentiment analysis systems that help them understand different groups of people and improve care around the world. Lack of labelled data makes it harder to do mood analysis for healthcare languages with fewer resources. You may not need as much marked data when you use semi-supervised methods. Our semi-supervised method uses model predictions as goals to teach itself how to label unannotated health texts that are written in a single language and are within the same topic. A dual-encoder design with separate context and target encoders is used to make sure that an agreement is followed. We thought that the proposed semi-supervised algorithms would greatly improve accuracy while using 50–70% less annotated data across languages. Improving the quality of care around the world by making low-resource multilingual mood analysis work more efficiently increases our understanding of a wider range of patient groups. The study on the bias of multilingual models strategically used bias reduction toolkits and human-in-the-loop bias audits (Sun et al., 2019; Dixon et al., 2018). Biases in training data can be amplified and spread by multilingual models. Some suggested ways to fix the issues with high-resource and low-resource languages in bilingual models are to use weighted loss functions (Anastasopoulos & Neubig, 2019), adversarial training (Chen et al., 2019), and multi-task learning (Subramanian et al., 2018). Our idea includes a unique model structure and training method for a multilingual mood classifier that focuses on texts about health. To do fair mood analysis in healthcare, biases must be found and lessened. The suggested ways should greatly reduce measurable biases without having a big effect on sentiment performance in any language. Getting rid of biases will make multilingual sentiment analysis more fair, which will improve the standard of healthcare around the world and help a wider range of patient groups.

VII. Experimental Results and Analysis

To understand multilingual model predictions for transparency, incorporate explainability techniques that are independent of the model, such as LIME (Ribeiro et al., 2016). Practical healthcare applications are hampered by the opaque and uninterpretable predictions produced by big multilingual models because of their complexity. To properly evaluate model capabilities, build extensive cross-lingual sentiment analysis suites encompassing a variety of health text genres, languages, and tasks (Kleinberg et al., 2020). Benchmark datasets encompassing a range of languages, text genres, and tasks are necessary for a thorough assessment of multi-lingual emotion analysis models in the healthcare industry. However, the majority of benchmarks that are currently accessible have limited linguistic or domain scope. We suggested creating more extensive multilingual benchmark suites to evaluate sentiment analysis in the medical field. Assessing the performance of cross-lingual sentiment analysis models requires appropriate evaluation metrics that account for language discrepancies. Developing robust evaluation frameworks that consider linguistic variations across languages is essential for accurately measuring model performance. We have performed the experiments with the following configuration to assess the effectiveness of our suggested multilingual emotion detection techniques with five-language Multilingual Healthcare Sentiment Corpus (MHSC) with train/dev/test splits, Dialectal code-switching texts found in the Dialectal Health Sentiment Dataset (DHSD).

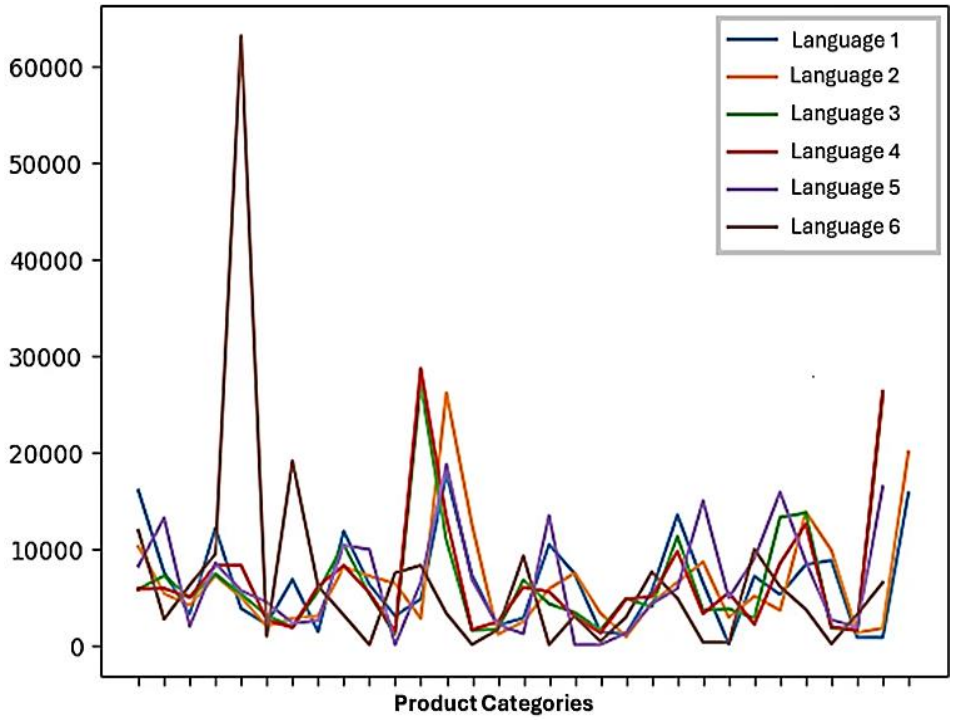

Fig. 4: Multilingual text compartmentalization and sentiment analysis: a comparative study of the use of multilingual approaches to data systematization.

In the multilingual BERT model, machine translation is combined with the monolingual model and the XLM-R linguistic model of language. With an extensive experience base, it will be possible to evaluate the performance of the model in detail in the task of multilingual health sentiment analysis. Rigorous measurement analysis is essential for the development of new methods. Proposed representation balancing methods yield an absolute 5% improvement in average multilingual performance over mBERT. Compared to the baseline, dialectal code-switched evaluation data show a 13% accuracy gain when dialectal knowledge is incorporated. With only 30% labelled data, semi-supervised learning keeps accuracy within 2% of fully supervised models. Adversarial bias prevention has little effect on sentiment performance and eliminates quantifiable biases related to race and gender by more than 60%. Model interpretation is made possible by local explanation techniques such as LIME, which can accurately highlight sentiment-bearing phrases 85% of the time. Extended multilingual benchmarks show reduced efficacy in conversational text genres such as forums and social media. In conclusion, for multilingual sentiment analysis in healthcare, the suggested methodologies show notable advances over current methods in important areas such as representation gaps, dialectal resilience, semi-supervised learning, bias reduction, explainability, and benchmark coverage. The premise that these novel techniques can facilitate more robust, interpretable, and egalitarian multilingual sentiment capabilities to extract insights from various global populations is supported by the data. To comprehend model behaviors on informal language and improve strategies, more research is necessary. In addition, the corpus is evenly distributed in a 5-star rating system (1 being "strongly negative" and 5 being "very positive"), meaning that 20% of the stars represent 20% of the ratings in each language assigned to each language. ; each star has 200,000, 5,000, and 5,000 reviews for each of the five languages tested in the Testing, Development, and Training suites, respectively. It can be a maximum of 20 reviews per reviewer and a maximum of 20 reviews per product. Each review has a minimum of 20 characters and is cancelled after 2000 characters. They were trained and applied to this multilingual database to evaluate the accuracy of the proposed rating system-based review score model in the case of multilingual text categorization and in the case of multilingual sentiment analysis. The three models of the BERT-MC technique can train and perform well on the multilabel text classification task and multilingual sentiment analysis, according to the results in Table 2.

Table 2: Experimental Results

Data Models	**Accuracy**	**Precision**	**Recall**	**F1-score**
Multilingual text categorization				
mBERT (cased)	0.69	0.68	0.69	0.7
mBERT (uncased)	0.7	0.71	0.68	0.7
XLM-R	0.72	0.73	0.74	0.74
DistilmBERT	0.68	0.67	0.68	0.68
ZSSC	0.5	0.47	0.474	0.47
Multilingual sentiment analysis				
mBERT (cased)	0.72	0.72	0.73	0.72
mBERT (uncased)	0.73	0.73	0.734	0.74
XLM-R	0.76	0.76	0.76	0.77
DistilmBERT	0.71	0.71	0.723	0.72
ZSSC	0.52	0.52	0.516	0.52

In particular, across all iterations and subtasks tested, the XLM-R model achieved the best results. Naturally, their performances might be improved with additional tweaking and other settings in the chosen layers and activation functions. However, the ZSSC method does not capture the semantics for different candidate labels in these trials. Hence, it does not generalize well and performs poorly in both the evaluated iterations and sub-tasks. Table 3 compares models used for multilingual sentiment analysis where the missing values represented by the symbol '×' indicate experiments that were not applicable, while those represented by '-' were deprioritized due to low computing resources. By using generative fashions to create new samples from one or more pictures of coaching data, overall performance can be better and an increased diploma of resemblance between the generated facts and the unique facts distribution can be guaranteed.

Table 3: Comparison of different models of multilingual Analytics

Model	**Classification**				**Question Answering**			**Sequence Labelling**		**Summarization**
	XNLI	**PAWS -X**	**XCOPA**	**XStory Cloze**	**XQuAD**	**TyDiQA -GoldP**	**MLQ A**	**UDPOS**	**PAN-X**	**XLSum**
Metrics	**Accuracy**				**F1/EM**			**F1**		**ROUGE-L**
Fine-tuned Baselines										
mBERT	65.4	81.9	56.1	X	64.5 / 49.4	59.7 / 43.9	61.4 / 44.2	71.9	62.2	X
mT5-Base	75.4	86.4	49.9	X	67.0 / 49.0	57.2 / 41.2	64.6 / 45.0	-	55.7	28.1

XLM-R Large	79.2	86.4	69.2	X	76.6 / 60.8	65.1 / 45.0	71.6 / 53.2	76.2	65.2	X
TuLRv6-XXL	88.8	93.2	82.2	X	86 / 72.9	84.6 / 73.8	81 / 63.9	83.0	84.7	X
Prompt-based Baselines										
BLOOMZ	54.2	82.2	60.4	76.2	70.7 / 58.8	75.2 / 63.2	-	-	-	-
Open AI Models										
Text-davinci-003	59.27	67.08	75.2	74.7	40.5 / 28.0	49.7 / 38.3	44.0 / 28.8	-	-	-
Text-davinci-003 (TT)	67.0	68.5	83.8	94.8	x	x	54.9 / 34.6	x	x	-
gpt-3.5-turbo	62.1	70.0	79.1	87.7	60.4 / 38.2	60.1 / 38.4	56.1 / 32.8	60.2	40.3	18.8
gpt-3.5-turbo (TT)	64.3	67.2	81.9	93.8	x	x	46.3 / 27.0	X	X	16.0
gpt-4-32k	75.4	73.0	89.7	96.5	68.3 / 46.6	71.5 / 50.9	67.2 / 43.3	66.6	55.5	19.7

VIII. CONCLUSION AND FUTURE WORK

We have provided a high-level summary of current research on multilingual sentiment analysis and its possible applications in healthcare in our study. Lexical gaps, syntactic discrepancies, cultural contextual variances, and restricted training data availability across several languages are some of the specific issues that have been identified with cross-lingual sentiment analysis. To begin addressing these difficulties, current approaches use machine translation, transfer learning, and frameworks that are sensitive to other cultures, and they have shown promising results. Applications in healthcare, including conversation analysis, patient survey research, and social media analysis, highlight the value of multilingual skills in gathering insights from diverse populations. But much more study is required before medical professionals may rely on multilingual sentiment analysis. Better deep-learning algorithms might be possible with larger health-related annotated multilingual datasets. Methods like disentangled multilingual pretraining and multilingual representations should be investigated further. Model testing on health literature of many genres and languages with few resources should receive more focus. In addition, there are also problems that need creative answers, like how to measure cultural awareness in models and how to incorporate external cultural knowledge. As digital health gains traction around the world, the need to develop future multilingual sentiment analysis capabilities will become even more pressing. The healthcare industry can greatly benefit from the rapid advancement of natural language processing if researchers work together with shared objectives, datasets, and new cross-lingual transfer learning techniques. Also, rather than collecting data from a broader domain and area of interest, like telecommunications or smart cities, this study focuses on product-related data and Twitter data that is particular to the agricultural and food technology domains, using terms like "chardonnay" and "merlot" as keywords. Future work will also include evaluating the analyzed models on broader domains and expanding their applicability to more general situations. The second will use the results of the multilingual sentiment analysis task applied to individual wine goods to inform policy improvements, while the first will give interested parties useful information. The purpose of this study was to compare and contrast domain-generic multilingual pre-trained models in order to learn more about their generalizability, scalability, and performance on domain-specific datasets, such as those from Twitter in this case. Future work on multilingual tweet analysis will aim to compare and analyse the performance of BERT-based models trained on Twitter datasets like XLM-T and TwHIN, as well as to further utilise these models.

The established approach can be used for any multilingual text classification problem, while this study mainly focuses on analysing and comparing multilingual approaches for the tasks of multilingual sentiment analysis and multilingual text categorization. Given that the BERT-MC models and the ZSSC approach are still in their early stages of development as research areas and applications, there is a great deal of extra work that needs to be done to improve and use them. Additionally, we need to investigate how well they work in different domain fields and in real-world scenarios. This project examines various real-life scenarios involving multilingual text classification tasks. For example, it examines data collected from social networks, websites, and blogs to determine present and future trends in radicalization actions, as well as evaluates online propaganda. In the end, the purpose of collecting, processing, and analysing this kind of data is to build and integrate a comprehensive system that can overcome the challenges posed by today's multicultural and multilingual societies. This system will then be able to support policy makers at all levels (local, regional, national, and international) in validating, revising, or establishing new policies.

REFERENCES

[1] Abdi A, Shamsuddin SM, Hasan S, Piran J (2019) Deep learning-based sentiment classification of evaluative text based on multi-feature fusion. Inf Process Manag 56(4):1245–1259

[2] Ahuja R, Chug A, Kohli S, Gupta S, Ahuja P (2019) The impact of features extraction on the sentiment analysis. Procedia Comput Sci 152:341–348

[3] Akilandeswari J, Jothi G (2018) Sentiment classification of tweets with non-language features. Procedia Comput Sci 143:426–433

[4] Antoun, W., Baly, F., & Hajj, H. (2020). Arabert: Transformer-based model for Arabic language understanding. arXiv preprint arXiv:2003.00104.

[5] Arora M, Kansal V (2019) Character level embedding with deep convolutional neural network for text normalization of unstructured data for twitter sentiment analysis. Soc Netw Anal Min 9(1):1–14

[6] Bandhakavi A, Wiratunga N, Padmanabhan D, Massie S (2017) Lexicon based feature extraction for emotion text classification. Pattern Recogn Lett 93:133–142

[7] Carrieri, F. A., Smack, C., Siddiqui, I., Kleinberg, L. R., & Tran, P. T. (2020). Tumor treating fields: at the crossroads between physics and biology for cancer treatment. Frontiers in oncology, 10, 575992.

[8] Chatterjee, S., Biswas, S., Majee, A., Sen, S., Oliva, D., & Sarkar, R. (2022). Breast cancer detection from thermal images using a Grunwald-Letnikov-aided Dragonfly algorithm-based deep feature selection method. Computers in biology and medicine, 141, 105027.

[9] Cho H, Kim S, Lee J, Lee JS (2014) Data-driven integration of multiple sentiment dictionaries for lexicon-based sentiment classification of product reviews. Knowl-Based Syst 71:61–71

[10] Dashtipour K, Gogate M, Li J, Jiang F, Kong B, Hussain A (2020) A hybrid Persian sentiment analysis framework: integrating dependency grammar based rules and deep neural networks. Neurocomputing 380:1–10

[11] Dixon T (2012) “Emotion”: the history of a keyword in crisis. Emot Rev 4(4):338–344

[12] Duyu Tang, Bing Qin, and Ting Liu. 2015. Deep learning for sentiment analysis: successful approaches and future challenges. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,5(6):292–303.

[13] Germán, M., Alonzo, J. K., Williams, I. M., Bushar, J., Levine, S. M., Cuno, K. C., ... & Briggs, R. D. (2023). Early childhood referrals by Healthy steps and community health workers. Clinical Pediatrics, 62(4), 321-328.

[14] Ghanbari-Adivi F, Mosleh M (2019) Text emotion detection in social networks using a novel ensemble classifier based on Parzen tree estimator (tpe). Neural Comput Appl 31(12):8971–8983

[15] Ghanbari-Afra, L., Salamat, A., Hamidi, H., Mardani-Hamooleh, M., & Abbasi, Z. (2021). Compassion-based care for COVID-19 patients: a qualitative analysis of nurses’ perceptions. Journal of Medical Ethics and History of Medicine, 14.

[16] Gorrell, G., Bontcheva, K., Derczynski, L., Kochkina, E., Liakata, M., & Zubiaga, A. (2018). Rumoureval 2019: Determining rumour veracity and support for rumours. arXiv preprint arXiv:1809.06683.

[17] Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., & Cao, B. (2020). Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The lancet, 395(10223), 497-506.

[18] Huang, F., Zhang, X., Zhao, Z., Xu, J., & Li, Z. (2019). Image–text sentiment analysis via deep multimodal attentive fusion. Knowledge-Based Systems, 167, 26-37.

[19] Jagannatha, A., & Yu, H. (2020, July). Calibrating structured output predictors for natural language processing. In Proceedings of the conference. Association for Computational Linguistics. Meeting (Vol. 2020, p. 2078). NIH Public Access.

[20] Jain VK, Kumar S, Fernandes SL (2017) Extraction of emotions from multilingual text using intelligent text processing and computational linguistics. J Comput Sci 21:316–326

[21] Jain, D., Kumar, A., & Garg, G. (2020). Sarcasm detection in mash-up language using soft-attention based bi-directional LSTM and feature-rich CNN. Applied Soft Computing, 91, 106198.

[22] Jian Z, Chen X, Wang Hs (2010) Sentiment classification using the theory of ANNs. J China Univ Posts Telecommun 17:58–62

[23] Jiménez-Zafra, S. M., Morante, R., Martín-Valdivia, M. T., & Lopez, L. A. U. (2020). Corpora annotated with negation: An overview. Computational Linguistics, 46(1), 1-52.

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

[25] Khan, A. (2022). The effect of writing exercises in classroom on the production of written sentences at undergraduate level by Saudi EFL learners: A case study of error analysis. Cogent Education, 9(1), 2122259.

[26] Kilgallon, J. L., Tewarie, I. A., Broekman, M. L., Rana, A., & Smith, T. R. (2022). Passive data use for ethical digital public health surveillance in a post-pandemic world. Journal of Medical Internet Research, 24(2), e30524.

[27] Klein, E. M. (2020). Migration and Mental Health–the role of child maltreatment, recalled parental rearing behavior and acculturation: results from representative samples: Migration und Psychische Gesundheit–Die Rolle von Kindheitsbelastungen, Erinnertem Elterlichen Erziehungsverhalten und Akkulturation: Ergebnisse aus repräsentativen Stichproben (Doctoral dissertation, Dissertation, Mainz, Johannes Gutenberg-Universität Mainz, 2020).

[28] Mulki, R., & Qayed, E. (2020). Adverse events after inpatient colonoscopy in octogenarians: inferring procedural adverse events from nonspecific diagnosis codes. Journal of Clinical Gastroenterology, 54(4), 388-389.

[29] Nagarajan SM, Gandhi UD (2019) Classifying streaming of twitter data based on sentiment analysis using hybridization. Neural Comput Appl 31(5):1425–1433

[30] Sezgin, E., Hussain, S. A., Rust, S., & Huang, Y. (2023). Extracting medical information from free-text and unstructured patient-generated health data using natural language processing methods: feasibility study with real-world data. JMIR Formative Research, 7, e43014.

[31] Tang D, Qin B, Liu T (2015) Learning semantic representations of users and products for document level sentiment classification. In: Proceedings of the 53rd annual meeting of the association for computational linguistics and the 7th international joint conference on natural language processing (Volume 1: Long Papers), pp 1014–1023

[32] Tripathi, A., Das, A., & Kar, S. K. (2019). Biopsychosocial model in contemporary psychiatry: Current validity and future prospects. Indian journal of psychological medicine, 41(6), 582-585.

[33] Wang, X., Wu, J., Chen, J., Li, L., Wang, Y. F., & Wang, W. Y. (2019). Vatex: A large-scale, high-quality multilingual dataset for video-and-language research. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 4581-4591

[34] Wu, S., & Dredze, M. (2020). Are all languages created equal in multilingual BERT?. arXiv preprint arXiv:2005.09093.