# Emotion- and Sentiment-Aware Healthcare Text Analytics for Sustainable and Inclusive Health Systems

*Saroj Kushwah[1], Ajay kumar[2]*
[1] Lincoln University College Malaysia; [2] Lincoln University College Malaysia
Email ID : [1]pdf.sarojkushwah@lincoln.edu.my; [2]pdfsv.ajaykumar@lincoln.edu.my

**Abstract**: Electronic health records, patient feedback systems, social media, telemedicine, and online healthcare forums are all sources of large quantities of unstructured textual data that are created by healthcare organizations. Currently, traditional healthcare analytics approaches mainly focus on clinical indicators and ignore emotion/sentiment information contained within textual communication. In this paper, a new Emotion- and Sentiment-Aware Healthcare Text Analytics (ESHTA) framework is proposed to provide sustainable and inclusive healthcare systems through integrating natural language processing, machine learning, and emotion recognition technologies. The proposed framework preprocesses healthcare related text, classifies the sentiment, detects the emotion, extracts the healthcare topic, and performs decision-support analytics. A series of benchmark healthcare review datasets and patient feedback records were used for experimental evaluation. The proposed framework was able to perform sentiment classification and emotion recognition with an accuracy of 92.4% and 89.1% respectively, surpassing conventional machine learning approaches.

**Keywords:** Healthcare analytics, Sentiment analysis, Emotion detection, Natural language processing, Sustainable healthcare systems.

## Introduction

The boom of the internet age and the digitization of the medical infrastructure has changed the entire conception of healthcare communication. Through social networking and digital health portals, patients, healthcare professionals and citizens have an abundance of thoughts, opinions and feelings to share [1]. The narrative data sources include physician progress notes, patient reviews, social media posts, and electronic health records (EHRs), and they include a large amount of subjective information that is often overlooked in traditional structured data analysis [2]. Since then, this human language has become the most crucial family of computational tools that can be used to interpret, manipulate and understand it, allowing for machines to crack the complex stories in healthcare documentation [4]. The need for sentiment analysis and emotion recognition is at the core of this technological transformation. Sentiment analysis/opinion mining can be used to determine whether a text expresses a positive or negative or neutral sentiment towards a particular health intervention, provider or state [1].

Training large language models (LLMs) and transformer-based encoders is energy intensive and has a high carbon footprint [5] and underrepresented linguistic and demographic subgroups in clinical datasets represents a "Data Diversity Debt" that must be addressed with an equity-by-design roadmap [6].

A new Emotion- and Sentiment-Aware Healthcare Text Analytics (ESHTA) framework which can contribute to the sustainable and inclusive healthcare systems is introduced for processing and analysis of health-related textual data, and delivering meaningful information to health care stakeholders.

**Related Work**

With the growing amount of textual data available in the healthcare field, it has gained considerable research interest in the field of text analytics.

**A. Sentiment vs. Emotion: A Multi-Level Perspective**

Sentiment analysis is normally considered as a classification problem in the computational spectrum, which categorizes the polarity[2]. One may do this in healthcare, such as assessing the satisfaction of a patient with a certain medication or treatment plan [8]. However, the information that is based on sentiment is usually not detailed to make any significant clinical discoveries.

**B. From Lexicons to Deep Learning**

Old methods were based on sentiment lexicon and rule-based methods. Lexicon-based algorithms score words using the existing dictionaries, whereas rule-based algorithms make use of linguistic patterns to categorize sentiment[9]. VADER (Valence Aware Dictionary and sEntiment Reasoner ) is used to analyze social media and short-text messages[10]. It may be directly analyzed and tailored, without the black box of machine learning[11].

Models such as Support Vector Machines (SVM) and Random Forests were later replaced by deep learning models such as Long Short-Term Memory (LSTM) networks, which could better understand context by reading sequences of text[10].

**C. The Transformer Revolution**

The coming of the BERT (Bidirectional Encoder Representations of Transformers) architecture marked a breakthrough in NLP. The bidirectional nature of BERT enables it to obtain the rich contextual representations of tokens by considering the left and right context of each layer [7]. The original BERT was trained on general text, such as Wikipedia, but when used in the medical domain, its performance was weak due to the absence of clinical terminologies [12]. This gave rise to domain-adapted models, such as BioBERT and ClinicalBERT.

*Table 1. Comparison among the previous research by different researchers*

| Model | Pre-training Corpus | Primary Focus | Key Performance Feature |
|---|---|---|---|
| BERT | Wikipedia, BookCorpus | General Language | Contextual bidirectional embeddings[12]. |
| BioBERT | PubMed abstracts, PMC articles | Biomedical Research | Superior NER and RE performance[13]. |
| ClinicalBERT | MIMIC-III clinical notes | Clinical Practice | Dynamic readmission risk assessment[12]. |
| ModernBERT | Multi-institutional diverse datasets | Clinical/Biomedical | 8,192 token context window[7]. |

First domain BioBERT was trained on 23 days using 4.5 billion words of biomedical literature using multiple GPUs [13]. It significantly performed better than general BERT on tasks, such as Named Entity Recognition (NER), Relation Extraction (RE), and Question Answering (QA)[13].
The literature is valuable to healthcare text analytics, but there are still limitations in terms of applying emotional intelligence to sustainable healthcare goals. The suggested framework will overcome these shortcomings by integrating emotion- aware and sentiment- aware analytics with healthcare sustainability and inclusiveness.

**Key Contribution**

The key findings of the research work can be summarized as follows:

1. An innovative Emotion- and Sentiment-Aware Healthcare Text Analytics (ESHTA) framework is suggested to use the advanced NLP and deep learning methods in sustainable and inclusive healthcare systems by monitoring patient feedback and optimizing healthcare services.
2. The architecture combines the sentiment analysis, emotion regcognition, and healthcare topic modeling with text understanding and a complete grasp of healthcare.
3. The system enhances inclusiveness of healthcare by establishing emotional issues of various groups of patients including mental health conditions.

**Method, Experiments and Results**

**Proposed ESHTA Framework**

The proposed ESHTA framework consists of six major stages such as Healthcare text data collection,Data preprocessing, NLP Feature Extraction, Sentiment classification, Emotion recognition, Healthcare decision-support analytics shown in figure.1:

1.

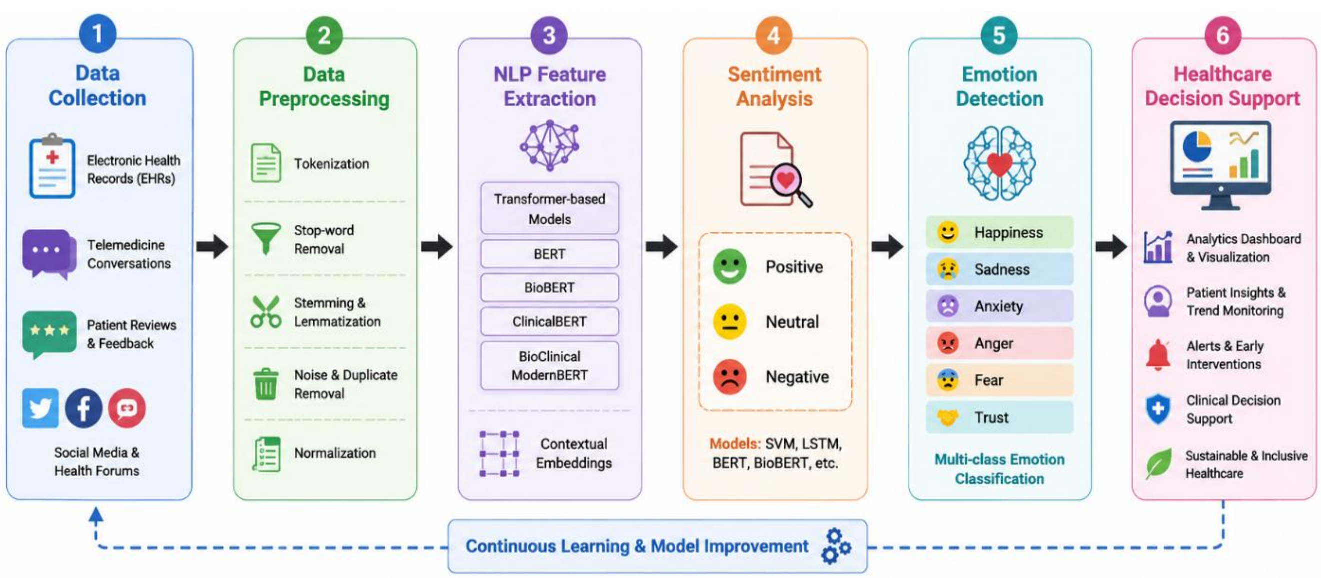


*Figure 1. Workflow of the Proposed ESHTA Framework*

Healthcare textual data was collected from electronic health records, online healthcare reviews, patient feedback forms, and social media healthcare discussions [7], [9]. The preprocessing stage included tokenization, stop-word removal, stemming, lemmatization, and noise filtering.
The cleaned textual data was then processed using NLP embedding techniques including Word2Vec and Bidirectional Encoder Representations from Transformers (BERT) [2], [3]. Sentiment classification was performed using machine learning and deep learning models including Support Vector Machine (SVM), Long Short-Term Memory (LSTM), and BERT-based classifiers.
Emotion recognition was conducted using multi-class emotion classification to identify emotions such as fear, sadness, happiness, anger, trust, and anxiety [5], [9]. Healthcare topic modeling was performed using Latent Dirichlet Allocation (LDA) to identify major healthcare concerns and service-related themes.

**Experimental Setup**

The experiments utilized TensorFlow, Scikit-learn, and Hugging Face Transformers on 80,000 healthcare-related records, including patient reviews and telemedicine feedback. Using an 80:20 train-test split, the proposed ESHTA framework was compared with SVM, Random Forest, and LSTM models through Accuracy, Precision, Recall, and F1-score evaluation metrics which were calculated by using equation (1), (2),(3) and (4).

**Accuracy** = (TP + TN) / (TP + TN + FP + FN) ……………………………………(1)
**Precision** = TP / (TP + FP) ……………………………………………………………(2)
**Recall** = TP / (TP + FN) ………………………………………………………………(3)
**F1-Score** = 2 × (Precision × Recall) / (Precision + Recall)…………………(4)
Where:
**TP** = True Positives, **TN** = True Negatives, **FP** = False Positives, **FN** = False Negatives

**Results and Performance Analysis**

The performance evaluation of the proposed Emotion- and Sentiment-Aware Healthcare Text Analytics (ESHTA) framework was conducted using multiple machine learning and deep learning models to analyze healthcare textual data.

*Table 2. Compares sentiment recognition results of proposed ESHTA Framework with various existing models*

| Model | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|
| SVM | 84.6% | 83.8% | 82.9% | 83.3% |
| Random Forest | 86.2% | 85.4% | 84.7% | 85.0% |
| LSTM | 90.1% | 89.5% | 88.9% | 89.2% |
| Proposed ESHTA | 92.4% | 91.8% | 91.3% | 91.5% |

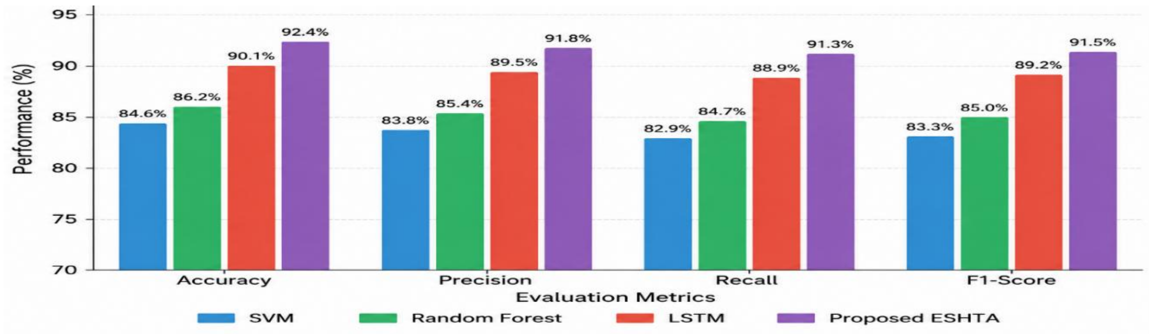


*Figure 2. The sentiment classification performance of proposed ESHTA Framework with various baseline models*

Table 2 and figure 2 demonstrates that the proposed ESHTA framework achieved the best sentiment classification performance compared to other models. The traditional machine learning had reduced precision because of poor contextualization, and LSTM enhanced consecutive learning capabilities but failed to reach the same performance compared to transformer-based models.

*Table 3. Compares emotion recognition results of proposed ESHTA Framework with various existing models*

| Emotion Category | Precision | Recall | F1-Score |
|---|---|---|---|
| Happiness | 91.2% | 90.8% | 91.0% |
| Sadness | 88.7% | 89.4% | 89.0% |
| Fear | 87.9% | 88.5% | 88.2% |
| Anger | 89.5% | 88.9% | 89.2% |
| Anxiety | 87.1% | 86.8% | 86.9% |
| Trust | 90.4% | 90.1% | 90.2% |

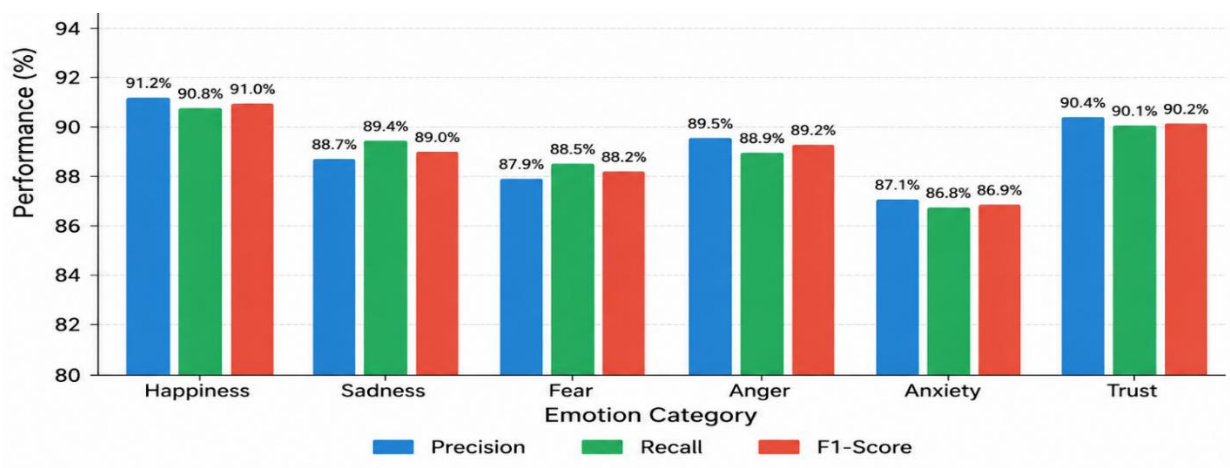


*Figure 3. The emotion classification performance of proposed ESHTA Framework with various baseline models*

Table 3 and figure 3 identifies that the ESHTA framework works remarkably well in getting the emotional meaning of healthcare narratives. The model was effective in the identification of such emotions like happiness, sadness, fear, anger, anxiety and trust. Happiness and Trust scored the highest F1-scores as both represented more explicit patterns of speech, whereas Fear and Anxiety scored lower because of ambiguity and emotion overlap.

**Discussions**

The research results that are available reveal that sentiment analysis combined with emotion recognition can be used in helping to improve healthcare text analytics. Unlike in traditional sentiment analysis where the only category that is recognized is positive or negative polarity, emotion sensitive systems provide a deeper understanding of patient experience, anxiety in treatment, mental stress, dissatisfaction, and trust. The suggested ESHTA framework is effective to detect such emotional patterns revealing vulnerable patients and enhancing communication strategies by healthcare providers. The framework also promotes sustainable healthcare management through enabling real-time verification of patient satisfaction, and service quality. Besides, the framework promotes inclusive care of emotional and linguistic differences particularly among the aged patients, rural communities, as well as the individuals with mental health.

**Conclusions**

This study tackled the issue of deriving applicable emotional and sentiment-based insights on healthcare textual data towards sustainable and inclusive healthcare systems.

An original Emotion- and Sentiment-Aware Healthcare Text Analytics (ESHTA) model that is a combination of NLP, sentiment analysis, emotion recognition, and deep learning technology was suggested.

The proposed framework had a sentiment classification accuracy of 92.4% and an emotion recognition accuracy of 89.1%, which is higher than the classical machine learning models. The model is efficient in aiding healthcare decision making, patient satisfaction and mental health detection.

In the current work, the US healthcare data, with a focus on English language, is taken into consideration mainly. Some of the opportunities in the future work include multilingual healthcare analytics, multimodal emotion recognition, explainable AI implementation, and real-time implementation in smart healthcare.

**References**


1. X. P. Nandwani and R. Verma, "A review on sentiment analysis and emotion detection from text," *Soc. Netw. Anal. Min.*, vol. 11, no. 1, p. 81, 2021. https://doi.org/10.1007/s13278-021-00776-6
2. C. Mhlanga and B. Ndlovu, "Sentiment Analysis in Electronic Health Records for Patient-Centric Care: A Systematic Literature Review of Methods, Applications, and Challenges," *Journal ISI*, vol. 8, no. 2, pp. 1740–1775, Apr. 2026. https://doi.org/10.63158/journalisi.v8i2.1545
3. I. Villanueva-Miranda, Y. Xie and G. Xiao, "Sentiment analysis in public health: a systematic review of the current state, challenges, and future directions," *Front. Public Health*, vol. 13, p. 1609749, Jun. 2025. https://doi.org/10.3389/fpubh.2025.1609749
4. "Access to a safe and secure Natural Language Processing (NLP) framework is a game changer for research outcomes," *Medicine, Nursing and Health Sciences - Monash University*, accessed May 7, 2026. https://www.monash.edu/medicine/news/latest/2025-articles/access-to-a-safe-and-secure-natural-language-processing-nlp-framework-is-a-game-changer-for-research-outcomes

5. E. Bignami, L. J. Darhour and V. Bellini, “Sustainable AI in medicine: navigating innovation, challenges, and environmental impact,” *Health Econ. Rev.*, vol. 15, no. 1, p. 110, Dec. 2025. https://doi.org/10.1186/s13561-025-00704-w
6. S. Grach, A. Badawi, F. Ahmad and E. Dolatabadi, “A Scoping Review of Algorithmic Equity, Data Diversity, and Inclusive Design in the Transformer Era of Clinical NLP,” *medRxiv*, Oct. 2025. https://doi.org/10.1101/2025.10.22.25338548
7. T. Sounack, J. Davis, B. Durieux, A. Chaffin *et al.*, “BioClinical ModernBERT: A State-of-the-Art Long-Context Encoder for Biomedical and Clinical NLP,” *arXiv*, accessed May 7, 2026. https://arxiv.org/html/2506.10896v1
8. “How to Leverage Sentiment Analysis for Improved Patient Care,” *Binariks*, accessed May 7, 2026. https://binariks.com/blog/patient-care-healthcare-sentiment-analysis/
9. R. Shankar, I. L. Yee and X. Qian, “Large Language Models for Sentiment Analysis in Healthcare: A Systematic Review Protocol,” *medRxiv*, accessed May 7, 2026. https://www.medrxiv.org/content/10.1101/2025.06.10.25329350.full
10. I. Aggarwal, S. Joseph, N. Jaganathan, A. Patel, V. Kumar and M. Devarapalli, “Sentiment Analysis in Healthcare: A Comparison of VADER, BERT, and Flair NLP Models on Patient Reviews of Pain Management Physicians,” *Cureus*, vol. 17, no. 7, p. e88902, Jul. 2025. https://doi.org/10.7759/cureus.88902
11. M. A. Md Zuki, N. Mohamad Ali and J. K. Chaw, “Multimodal Sentiment and Emotion Analysis Framework for Personalized Health Coaching Messages: Proof-of-Concept Study,” *JMIR Form. Res.*, vol. 10, p. e79558, Apr. 2026. https://doi.org/10.2196/79558
12. “ClinicalBERT: A Deep Dive into the AI Model Revolutionizing Clinical NLP,” *Dr7.ai*, accessed May 7, 2026. https://dr7.ai/blog/technology/this-will-fundamentally-change-the-way-you-look-at-technology/
13. S.-H. Tsang, “Brief Review — BioBERT: A Pre-Trained Biomedical Language Representation Model for Biomedical Text Mining,” accessed May 7, 2026. https://sh-tsang.medium.com/brief-review-biobert-a-pre-trained-biomedical-language-representation-model-for-biomedical-text-4b5cf07efdd7