Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 11, No. 03, September 2025.

FROM SURVEYS TO SENTIMENT: A REVIEW OF PATIENT FEEDBACK COLLECTION AND ANALYSIS METHODS

Ayushi Gupta, Anamika Gupta, Dhruv Bansal and Khushi

(Received: 15-May-2025, Revised: 21-Jul.-2025, Accepted: 4-Aug.-2025)

ABSTRACT

Patient feedback plays a crucial role in improving the quality, responsiveness and patient-centric approach of healthcare services. This paper presents a comprehensive review of both traditional and digital methods used to collect patient feedback, emphasizing their value in improving healthcare delivery, examines the tools and channels used, including surveys, interviews and multi-channel digital platforms. The review further explores sentiment-analysis techniques applied to patient feedback, focusing on how machine learning, deep learning and large language models are used to interpret and categorize unstructured text. The recent literature is systematically analyzed, with comparative tables that highlight feature-extraction methods, classification algorithms and performance metrics reported in various studies. Additionally, the paper addresses key challenges in feedback collection and sentiment analysis. Future research directions are proposed, such as automating feedback systems and incorporating patient perspectives into quality-improvement frameworks. This review is intended to assist Healthcare IT Professionals and medical Data Scientists who deal with healthcare delivery and computational analysis, whose target is to extract actionable insights from patient feedback using modern AI techniques.

KEYWORDS

Patient feedback, Sentiment analysis, Lexicon, Machine learning, Deep learning, Generative AI.

1. INTRODUCTION

Patient satisfaction is crucial for measuring the quality of healthcare services. It reflects how effective clinical care is and the broader experience of patients within the healthcare system. However, patient experiences are influenced by many different things, such as a person’s age, gender, education level and health condition. Traditionally, patient experience was viewed as a set of interactions that shape a patient’s point of view regarding care. Over time, in modern healthcare systems, the concept also includes the experiences of healthcare workers, families and the wider community. In [1], the authors stated that every interaction of a patient with healthcare-system matters, the values and behavior of the healthcare organization affect the care received by a patient, each patient’s personal feelings and background shape their views and patient experience changes throughout the entire treatment process. The authors highlighted the fact that the way healthcare workers feel and what they go through also affect the care they give to patients. The authors of [2] exhaustively reviewed 60 research papers from 1969 to 2019 to understand the factors that shape patient experiences and concluded that patient satisfaction is a complex topic and must be researched further to understand how thoughts and feelings of a patient affect his/her satisfaction. The authors of [3] developed a theory - Clinical Performance Feedback Intervention Theory (CP-FIT) to explain how patient feedback works and what makes it successful. The authors found that the feedback process involves goal setting, data collection, feedback delivery, interpretation, acceptance and behavior change. They identified 42 high-confidence factors that influence the success of feedback and concluded that feedback is most effective when it aligns with the values of healthcare professionals and results in clear and easy to implement improvements.

Feedback plays an important role in the growth and improvement of an organization. Taking feedback on a regular basis encourages an individual or an organization to engage in a culture of continuous learning and personal development. In the context of medicine, understanding patient feedback is crucial for enhancing healthcare services, as it provides insights into patient experiences and identifies

A. Gupta, Anamika Gupta (Corresponding Author), D. Bansal and Khushi are with Faculty of Computer Science, Shaheed Sukhdev College of Business Studies, Delhi, India. Emails: ayushigupta@sscbs.du.ac.in, anamikargupta@sscbsdu.ac.in, dhruv.23520@sscbs.du.ac.in and khushi.23531@sscbs.du.ac.in

areas for improvement.

Without any feedback mechanism, the quality of healthcare cannot be measured. Unstructured patient feedback full of useful information (from social media and online platforms) is growing quickly. However, it is not being used as much as it could be to improve healthcare services. Manually analyzing such large-scale data is not feasible due to time and resource constraints. The authors of [4] reviewed 19 studies that utilized natural language processing and machine-learning techniques for sentiment analysis and classification of patient feedback collected through surveys as well as social media. The selected studies employed supervised, unsupervised and semi-supervised learning methods that could categorize feedback into positive, negative or neutral sentiment and can be used for processing millions of such responses.

Figure 1 illustrates a structured workflow, used by various researchers, for classifying patient feedback into sentiments, incorporating both human annotation and artificial intelligence. AI mainly comprises of Natural Language Processing (NLP), Machine Learning (ML) and Deep Learning (DL) techniques. Initially, feedback of patients is collected through various mechanisms and stored in a database which follows pre-processing with several techniques, like Tokenization, Stemming, Lemmatization, Lowercasing …etc. to standardize the textual data. The standardized and processed textual data then undergoes two major pipelines, so that labels or sentiments can be generated for the data:

1) Traditional Machine Learning algorithms: Supervised, unsupervised, semi-supervised.
2) Large Language Models directly convert textual data and generate sentiment labels efficiently.

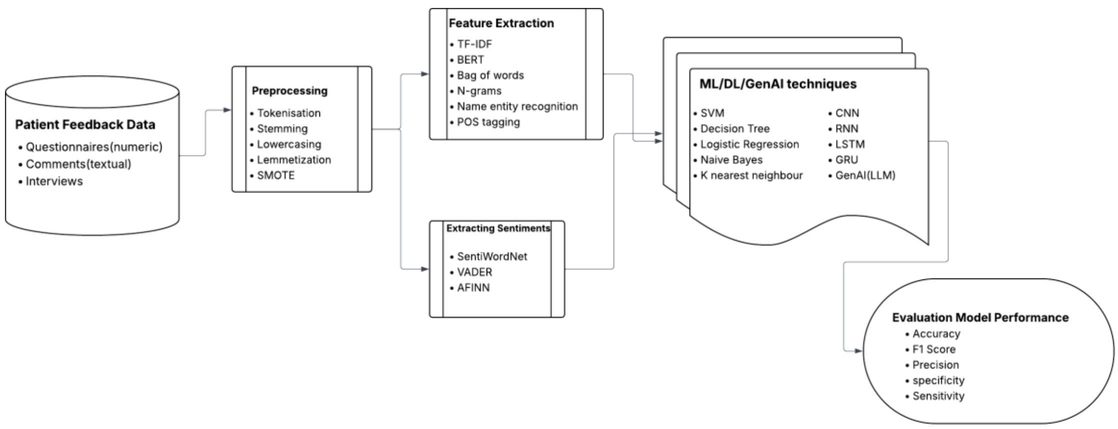

Figure 1. Methodology of sentiment analysis.

The labels are then manually checked for a sub-set of data by annotators ensuring consistency *via* Inter Annotator Agreement (IAA). When humans label data (e.g. tagging a comment as "positive", "neutral" or "negative"), their decisions can differ due to personal interpretation. IAA measures how consistently multiple human labelers agree when labeling or classifying data. The final human check ensures accurate sentiment analysis.

In this paper, our aim is to study the research space of sentiment classification in patient feedback. The initial focus is on the data-collection methods used by various researchers, followed by an analysis of the methods used for sentiment classification. Reliability and performance of sentiment-classification methods depend on the quality, accuracy and format of the collected feedback. Thus, it is crucial to study the data-collection mechanisms of the patient feedback. Various forms of inputs, such as surveys, interviews, questionnaires, and social-media content, yield different data types which will require different preprocessing and modeling strategies.

The Scopus database is chosen for literature reviews. The keywords "Patient Feedback" and ("Sentiment Analysis or Natural Language Processing or Machine Learning") are used. The documents are filtered from the last five years (2019-2024), including some studies from 2025 to focus on recent publications that reflect the latest advances and developments in this area. In this review are high-citation research papers related to feedback data-collection mechanism and sentiment-classification strategies.

Based on the motivation and scope of this review, the following research questions (RQs) are addressed.

1) **RQ1:** What are the current methods used for collecting patient feedback?
2) **RQ2:** How is sentiment analysis applied to patient feedback and what AI techniques (ML, DL, LLMs) are commonly used?
3) **RQ3:** What practical challenges arise when collecting and analyzing patient feedback, particularly at scale?

To address the above-mentioned RQs, various sections have been introduced. Section 2 details various methods that have been employed for collection and analysis of patient feedback without employing any AI techniques. Further, Section 3 provides a brief overview of how sentiment is analyzed using various ML and DL techniques and how generative AI is now being used for the same. This is followed by Section 4, which provides a review of recent studies that have performed sentiment analysis on patient feedback data. Moreover, the challenges associated with the collection and analysis of patient feedback are presented in Section 5. Lastly, Section 6 concludes the study along with future scope. This review is mainly for health-informatics researchers and IT professionals who want to develop or improve systems that can automatically analyze patient feedback. The goal is to help create tools that make it easier for healthcare teams to understand overall patient satisfaction and find areas that need improvement without reading thousands of comments manually. In addition, feedback-collection methods will help healthcare administrators and practitioners who need to implement them.

2. UNDERSTANDING AND COLLECTING PATIENT FEEDBACK

This section addresses RQ1 by discussing methods for understanding and collecting patient feedback. Recent research has explored various methods for collecting, analyzing and utilizing patient feedback effectively. Some of the recent studies that focus on data collection and highlight the challenges faced during the process are mentioned in this section. In [5], the authors explored different ways to collect patient feedback and followed a participatory research approach involving patients, general practitioners (GPs), medical receptionists and an advisory group. Semi-structured interviews were conducted, where a set of open-ended questions were prepared. The interviews were analyzed using Thematic Analysis, in which the responses were categorized by attaching keywords to them. The software that was used was MAXQDA software (version 2022). It was concluded that real-time feedback is the most effective way to capture patient experiences. Also, rather than continuous collection, periodic feedback was found to be more practical and manageable.

In study [6], the authors focus on whether collecting data in real time at multiple stages of hospitalization can identify areas for improvement more effectively than traditional satisfaction surveys. This research was carried out in the Orthopedics Department of an Italian university hospital. Patients were given two different paper-based questionnaires at two time points: at hospital admission and at discharge. The data collected covered four key categories - Patient-Reported Outcomes (PROs) to measure self-rated health, Patient-Reported Experiences (PREs) to evaluate the quality of care and efficiency of services, Patient-Reported Preferences (PRPs) to capture other aspects of care that patients value and Emotional State Tracking to measure patient emotions at different stages. The authors observed that capturing patient experiences at multiple points in the hospital journey provided better insights than a single post-discharge survey.

In [7], the authors studied a digital patient feedback platform Hospitalidee, where patients may post positive or negative feedback about hospitals that have partnered with the platform. They selected all the negative feedback from the platform for a single hospital called OSTI. A two-step analysis of 134 negative feedback comments was performed to reveal common themes in patient complaints. Firstly, complaints were classified into four categories based on the service provided. Further, complaints were classified according to departments in order to target the process of quality improvement to the areas where most needed. This was followed by thematic analysis of the feedback comments in order to identify important themes. The study concluded with the statement that digital patient-feedback platforms should be actively integrated into hospital decision-making processes.

In [8], the authors explored current practices of collecting feedback and utilizing it. The authors conducted semi-structured interviews with nine participants from three different hospitals. Four types of methods were identified to collect feedback, which are given in Table 1. The challenges faced during the process are also mentioned.

Table 1. Different methods of feedback collection [8].

Methods	Description	Challenges
Structured, Official Feedback	Standardized surveys distributed through web-based platforms, paper forms or automated systems.	Response rates are low. Feedback delayed post discharge. Limited depth due to structure.
Unstructured Feedback	Informal feedback through verbal conversations, emails or suggestion boxes.	Difficult to analyze, Underreported issues, Not documented systematically
Pilot Projects using Digital Tools	Hospitals experimenting with new feedback-collection technologies, such as mobile apps and real-time patient surveys.	Not widely implemented. Requires staff training. Cost and infrastructure barrier.
Occasional Studies and Research Projects	One-time research initiatives conducted by hospital staff, students or external organizations to assess patient experience.	Lack of continuity. Not integrated into daily operations. Results take time.

A study carried out in three large hospitals in Brazil is described in [9]. Nine semi-structured interviews were conducted and hospital documents, such as feedback forms, action plans and reports, were also analyzed. NVivo 11 software was used to organize and analyze the information. It was found that hospitals use structured quality-improvement (QI) tools to analyze patient feedback and make meaningful changes. Some of such tools are:

- Plan-Do-Check-Action: Identify a problem based on patient feedback, implement a small change, measure the impact and if successful, apply the change hospital-wide.
- Ishikawa (Fishbone) Diagram: A visual tool to identify root causes of a problem by categorizing potential reasons.
- Pareto Analysis (80/20 Rule): It follows the 80/20 rule, meaning, 80% of patient complaints come from 20% of the problems, fixing that 20% can solve most issues.

The authors of [10] focused on creating simple and short questionnaires suitable for hospital patients with varying literacy levels. The patient experience monitor had two versions that were adult inpatient (14 items) and adult outpatient (15 items), both of them included key aspects, like emotional support, waiting time, privacy, clarity of information, communication and family involvement. From this study, it was found that even patients with low literacy found patient experience monitor easy to understand. The short format improved response rate.

While feedback collection is an important step in improving healthcare services, it becomes valuable when it is interpreted. Most patient responses are in unstructured formats, like free-text surveys, interviews or online reviews, as seen above and contain implicit information that is not immediately assessable. Manual review of such comments is resource-intensive and inconsistent. This is where sentiment classification becomes important. Sentiment classification helps reveal the underlying emotional tone of patient comments, whether they are satisfied, frustrated, in fear or express gratitude. By categorizing feedback into sentiment, such as positive, negative or neutral, healthcare providers can identify problem areas more efficiently. The techniques used for sentiment analysis are presented in the next section. Table 2 describes the patient-feedback datasets that have been collected and analyzed further to derive useful insights.

3. Sentiment Analysis Techniques

Sentiment is an opinion influenced by emotions. Automating the extraction of sentiments in unstructured data, such as reviews, comments or feedback, is an area of study under Natural-language Processing. Its objective is to automate extraction and interpretation of sentiments or data from text, providing insights into public sentiment, customer satisfaction and market dynamics.

Due to digitization of processes and the increase in the use of social media, the amount of reviews or feedback is enormous, making it impossible to process them manually. Therefore, there is a growing need for the use of AI-driven approaches to identify and extract the sentiment.

Table 2. Summary of patient-feedback datasets used in the reviewed studies.

Ref.	Data-collection Period	Dataset Description	Record Type	Open Source
[6] (2021)	January-February 2019	Longitudinal survey: preferences, experience, outcomes at admission/discharge	Open-ended questions answered by 254 patients	Available upon request
[11] (2020)	January 2008-October 2019	Synthesized findings from studies on patient feedback and review of interventions	20 studies having patient feedback (qualitative & quantitative)	Available upon request https: //shorturl.at/ z4cxg (supplementary data)
[12] (2024)	2018-2021	Norwegian national patient-experience surveys conducted by the Norwegian Institute of Public Health (NIPH)	2250 patient comments	No
[13] (2020)	January 2018-January 2019	Patient surveys data collected at Geisinger Holy Spirit Hospital covering various aspects of care and labeled by sentiment	2830 records of un-structured free-text comments	No
[14] (2020)	2016-2020	Three survey questions with binary responses related to respect received, clarity of explanation and attentive listening	3134 patient responses to survey Questions	No
[15] (2021)	-	Patient reviews for specific medications along with a 10-star rating	232 K free-text drug reviews	https://surl.li/ wjvtwk
[5] (2025)	-	Qualitative study exploring patient-feedback methods for e-Health in general practice	Interview transcripts of 13 patients, 8 GPs, 2 receptionists	No
[16] (2023)	-	Cancer-patient stories	Study 1-14, 391 random posts, study 2-30,037 posts	https://www. cancerconnection ca/s/ https://surl.li/uirjeq
[17] (2022)	January 2017-July 2017	Friends and family test (FFT) free-text, Patient feedback	69,285 responses	No
[10] (2020)	-	Questionnaires, interviews, pilot study	28 interviews, pilot study and surveys	https://pmc.ncbi.nlm .nih.gov/articles/PM C7725101/table/t00 02/
[18] (2024)	-	Patient & family-member discussion posts on a medical forum	12,103 posts of patient narratives	https://patient.info/for ums
[7] (2023)	2018	Negative feedback data from a digital platform of one hospital	Analysis of 134 reviews.	No
[19] (2022)	-	Five questions based on information provided, personal approach, collaboration among healthcare professionals organization of care and general feedback	534 responses of open-ended questionnaire	No
[20] (2021)	2019-2023	Classifying the complaint records using ML and NLP	1465 records having different complaints describing communication	No
[21] (2025)	January 2014-December 2014	Analyzed sentiment in patient comments using natural-language processing	1117 comments and ratings from 1 (worst) to 5 (best)	https://surl.li/ zcxygz

Due to digitization of processes and the increase in the use of social media, the amount of reviews or feedback is enormous, making it impossible to process them manually. Therefore, there is a growing need for the use of AI-driven approaches to identify and extract the sentiment. Recent advancements in artificial intelligence, machine learning, deep learning and generative AI, particularly large language models (LLMs), have greatly enhanced the precision and scalability of sentiment-analysis systems, establishing sentiment analysis as a crucial tool for examining extensive unstructured data. Sentiment analysis traditionally classifies text into positive, negative or neutral categories. However, advances in the field have led to the identification of nuanced sentiments, such as anger, joy, fear, toxicity, sadness and surprise.

Techniques for Extraction of Sentiment

In recent years, multiple strategies have emerged to improve the precision and scalability of sentiment classification. Conventional methods, such as the lexical-based approach, use sentiment dictionaries to assign polarity scores to individual words. Meanwhile, machine-learning methods rely on labeled datasets to train models that can identify sentiment patterns. In recent years, large language models (LLMs) have revolutionized the domain by comprehending complex linguistic nuances and context on an unprecedented scale. This transition from rule-based methods to data-driven and neural approaches highlights the evolving landscape of sentiment analysis, offering a range of strategies to address the various challenges in text analysis.

Before applying any sentiment-analysis technique, pre-processing of the text needs to be carried out. Some of the text pre-processing techniques are listed below:

1) Data cleaning - removing/handling emojis, URLs, HTML Tags, stop words, punctuation marks, spell checking, normalization, number removal, and converting into lowercase are some of the common data-cleaning techniques
2) Tokenization breaks down text into smaller units called tokens. The tokens can be a single character, word, phrase, sentence, paragraph, …etc.
3) Stemming is a process to find the root of a word by removing suffixes.
4) Lemmatization is a process that considers the context and part of speech to reduce words to their base forms, called lemmas.

Further, the techniques for classification of text into various sentiments are classified as below:

1) Lexicon-based Approach

The lexicon-based approach to sentiment analysis relies on dictionaries of words that are pre-assigned sentiment values, typically categorized as positive, negative or neutral. This method estimates the overall sentiment by summing the sentiment scores of individual words within a text. Its simplicity and transparency make it a popular choice, especially for domains where interpretability is critical or when the labeled data for training machine-learning models is scarce. Tools, such as SentiWordNet [22], VADER [23] and AFINN [24], are widely used in research and industry.

2) Machine Learning-based Approaches

Machine learning (ML)-based approaches have transformed sentiment analysis by moving beyond simple keyword matching to more sophisticated algorithms that can automatically learn patterns from data. These models do not require pre-defined lexicons and are capable of handling larger datasets and more complex language patterns. The key strength of machine learning approaches lies in their ability to generalize from data and to adapt across different domains, making them highly effective for sentiment analysis in areas like social media, product reviews and customer feedback [25]. Supervised machine learning is a prevalent approach in sentiment analysis, where models are trained on labeled datasets to classify text as positive, negative or neutral. This process generally involves data pre-processing, feature extraction and model training.

Feature Extraction

Feature extraction is crucial in converting text data into numerical vectors that the machine-learning model can process. Common methods for feature extraction include Bag-of-Words [26], TF-IDF [27], Word Embeddings [28]-[29]. Bag-of-Words is a simple and easy method which represents text by counting word frequency. Context and semantic meaning are lost in this process. TF-IDF weighs terms by their importance across documents and highlights rare, but important, words. Though computationally expensive, the technique is widely used in many text-mining applications. Word Embeddings (Word2Vec, GloVe) map words to continuous vector space, capturing semantic meaning, context and word relationships.

Model Training

Model training involves feeding the features into a machine-learning algorithm, which learns to predict the sentiment label based on the training data. Some of the most commonly used algorithms for sentiment classification include:

- **Linear Regression**: A simple model for prediction of continuous outcome based on a linear combination of input features [30].
- **Decision Tree**: A tree-based model that chooses the feature as a node of the tree based on metrics, like Gini-index and Entropy [31].
- **Naive Bayes**: Simple and effective for high-dimensional data [32].
- **Support Vector Machines (SVMs)**: this technique finds optimal hyper-planes for classification, performing well in high-dimensional spaces [33].
- **Logistic Regression**: A linear model commonly used for binary classification, such as predicting whether a review is positive or negative [34].
- **K Nearest Neighbor**: A lazy learner technique that does not learn a model and matches the unseen tuple at the time of prediction. Classification of the sample is based on the majority label among its k nearest neighbors. [35].
- **Random Forest**: Ensemble method that combines multiple decision trees. Prediction is based on the majority voting of the output of all models [36].

3) Deep Learning-based Approaches

Building upon the foundation laid by traditional machine-learning approaches, deep learning has emerged as a transformative force in sentiment analysis. While traditional models rely heavily on feature engineering and handcrafted rules, deep-learning models automatically learn representations from data, capturing complex linguistic patterns and contextual information. This sub-section highlights the contributions of CNNs, RNNs, LSTMs and GRUs, illustrating the transformative impact of deep learning in extracting sentiment from textual data. Convolutional Neural Networks [57] are a fast and high-performance technique that applies convolutional filters to extract n-gram features from text. Recurrent Neural Networks (RNNs) represent a slow, moderately performing technique that processes sequential data by maintaining hidden states, especially suitable for time-series data. Long Short- Term Memory (LSTM) deals with memory cells for long-term dependencies, suitable for long text, emotion recognition, speech processing. Gated Recurrent Units (GRUs) constitute a technique that reduces the complexities of LSTM by combining gates, making it suitable for text classification and machine translation.

4) Generative AI-based Approaches

In recent years, the advent of Generative AI (GenAI) and Large Language Models (LLMs) has significantly transformed the landscape of sentiment analysis. Unlike traditional machine learning and deep-learning approaches that require extensive labeled data and task-specific architectures, LLMs leverage large-scale pre-training on diverse datasets, enabling them to generalize across multiple tasks, including sentiment classification, with minimal fine-tuning. Large Language Models, such as OpenAI’s GPT series, Google’s BERT and Meta’s LLaMA, have set new benchmarks in natural-language understanding (NLU) and generation [37]. Their transformer-based architecture allows them to handle long-range dependencies, outperforming traditional recurrent neural networks (RNNs) and convolutional neural networks (CNNs) in various NLP tasks [38].

Transformer Architecture, the Backbone of LLMs: The transformative power of LLMs lies in the underlying transformer architecture, introduced by [38]. This architecture is based on the self-attention mechanism, which enables models to weigh the significance of different words in a sentence, regardless of their position. Unlike RNNs, which process sequences step by step, transformers process entire sequences simultaneously, drastically improving efficiency and scalability. This parallelization allows transformers to model long-range dependencies more effectively, which is critical for capturing complex sentiment patterns in lengthy reviews or documents.

The self-attention mechanism facilitates context-aware sentiment analysis by dynamically adjusting attention to relevant words. For example, in a sentence like "The movie was surprisingly good despite its slow start," the transformer architecture can attribute higher attention weights to "surprisingly good," correctly identifying the overall positive sentiment.

Zero-shot, Few-shot and Fine-tuning Approaches: LLMs have the capability of classifying sentiments based on the prompts given. Various types of prompts, such as zero-shot and few-shot can be used for learning. For example, models such as GPT-3 can classify sentiments even without direct training by utilizing prompt engineering techniques. By presenting the model with instances of positive, negative and neutral sentiments, researchers can steer the model toward producing precise predictions [39]. This versatility minimizes the necessity for labeled datasets and greatly speeds up the implementation in practical scenarios. Further, fine-tuning BERT on social-media datasets having informal and noisy data improves the sentiment-classification accuracy [40] and RoBERTa, a variant of BERT, optimizes the pertaining techniques and works on larger datasets [41].

4. SENTIMENT ANALYSIS ON PATIENT FEEDBACK

This section addresses RQ2: How is sentiment analysis applied to patient feedback and what AI techniques (ML, DL, LLMs) are commonly used. The reviewed literature has been organized by approach type — ML, DL and LLMs. The feature-extraction and classification techniques employed in the reviewed studies are presented in Tables 3 and 4. Table 3 outlines the ML and DL approaches used for feedback analysis, while Table 4 summarizes the techniques applied in LLMs, respectively. The tables also give the performance achieved by different techniques. The following observations can be made from Table 2:

1) Approximately 43% of the datasets used in the reviewed studies were unstructured, while about 29% were structured and 29% were based on survey responses.
2) Majority of the studies categorized the sentiments as positive, negative and neutral. Maehlum et al. [12] used four sentiment categories - positive, negative, neutral and mixed, where mixed indicates sentences containing both positive and negative polarity. Similarly, Cho et al. [49] also defined positive aspects as care and kind and negative aspects as pain and rude.
3) Data cleaning was also observed to be an important part of all studies to improve model performance. Moreover, text cleaning and pre-processing techniques, such as tokenization, lemmatization, stop-word removal, stemming and lowercasing have been utilized in majority of the studies.

The bar chart in Figure 2 represents the different feature extraction techniques that have been used in the reviewed studies along with the study count. It can be observed that TF-IDF is the most widely used feature extraction technique in analyzing patient feedback data.

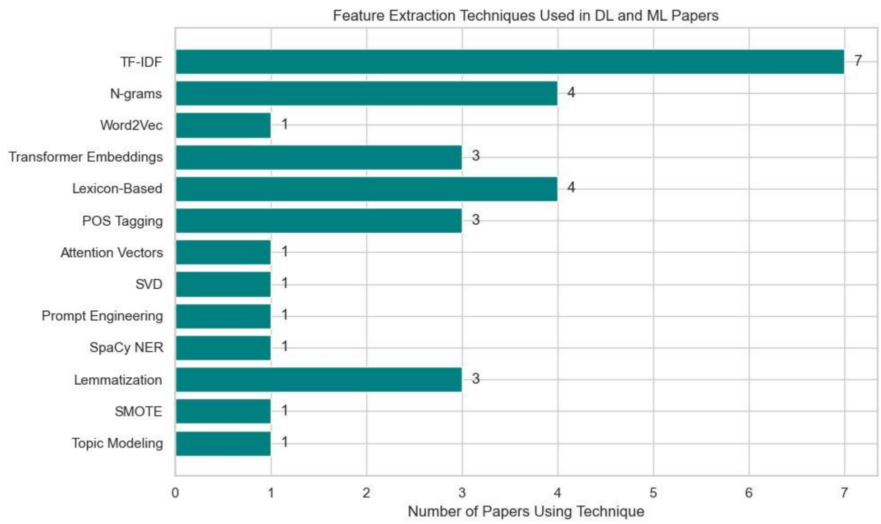

Figure 2. Feature-extraction techniques used in the analysis.

While Tables 3 and 4 summarize a wide range of studies applying various NLP techniques to patient

Jordanian Journal of Computers and Information Technology (JJCIT), Vol. 11, No. 03, September 2025.

Table 3. Summary of NLP, ML and DL techniques used in patient-experience analysis.

Ref.	Feature-extraction Techniq.	Classification Techniques	Performance Metrics
[4]	TF-IDF	Supervised (Support Vector Machine (SVM), Naive Bayes (NB)), Unsupervised (Linear Discriminant Analysis (LDA), Factorial LDA)	Precision up to 88%; SVM accuracy 72%)
[13]	N-gram, Bigrams, Part-of-Speech (POS) Tagging, Word Frequency, Word Clouds	Artificial Neural Networks (ANN - Keras-Sequential model with dense and dropout layers)	Precision-0.83, Recall-0.82, F1-0.82, Support-103 sample
[20]	Word level TF-IDF, N-gram level TF-IDF(n=2)	SVM, Multifactor Logistic Regression (LR), Multinomial NB	Accuracy (up to 0.91), F1-Score, Precision, Recall, AUC (up to 0.94)
[19]	TF-IDF, N-gram	Finetuned Multilingual Bert, NMF for topic modeling	F1-Score (Positive: 0.97, Negative: 0.63), Machine-Human Topic Match: 90%, Topic Representativeness: 80.9
[21]	LIWC-22, Meaning Extraction Method (MEM), Principal Component Analysis (PCA)	Multivariable Linear Regression	Not given
[17]	Bag of words, tri-gram analysis.	Decision Tree (DT), Random Forest (RF), SVM, K-Nearest Neighbour (KNN), NB and Gradient Boosted Trees (GB)	SVM F1-score 94%
[42]	TF-IDF, Bag of Words, Name entity recognition, Word embedding	Transformer models (RoBERTa) and CNNs	RoBERTa F1-Score: Neurology (1.0), Combined datasets (0.995). CNN: 0.760.
[43]	Name Entity Recognition, TF-IDF, BERT	RF, GB models	85–90%
[16]	TF-IDF, Topic modeling	Topic classification, LDA.	87%
[14]	BERT, Bag of Words	RF, LR, DT and Social Network Analysis	RF: 87.6% (courtesy), 81.9% (clarity, listening).
[44]	Tokenization, lemmatization, Domain-specific lexicons	SVM, NB, DT	F1-score: 60%
[45]	TF-IDF, POS Tagging, BERT	Machine learning models for sentiment categorization	78.2–87%
[46]	Bag of words, TF-IDF	Sentistrength (for sentiment analysis), LDA	89.3% (general), 92.6% (healthcare), 90.8% (life
[47]	Word count, TF-IDF, Boolean features	NB, Multinomial NB, SVM, LR, RF	81% (cleanliness), 84% (dignity), 89% (recommendation).
[48]	N-grams, SNOMED CT, BERT	Rule-based NLP, SVM	AUC: 0.997; Sensitivity: 88%; Specificity: 96%.
[21]	Topic modeling	Topic modeling to identify themes (e.g., communication, logistics).	78.5%–87% across different aspects of care
[49]	TF-IDF, Sentiment lexicons, bag of words	LR, t-test/ANOVA	78.5%–87% across different aspects of care
[13]	TF-IDF from lemmatized, synonym-standardized text	Sequential Deep Neural Network (Keras); 3 dense layers with dropout	Accuracy peaked at epoch 35; ReLU + Softmax
[15]	TF-IDF, Bi-grams, Lexicon-based (Bing) SMOTE	Artificial Neural Network (ANN), SVM, Logistic Regression	SVM: Acc. = 0.720, AUC = 0.725
[50]	TF-IDF vectorization, 1–4 grams, Harvard emotional dictionary	N-gram Deep Learning model; also compared with RF, NB, Linear Regression	N-Gram model: Acc. = 89.4%
[51]	UMLS mapping, Symptom dictionaries, Term frequency, Lexicon usage, Clustering, Patient-authored symptom terms	Rule-based NLP, Machine Learning (SVM, RNN, Logistic Regression), Text Mining	F1-scores up to 90%, Precision/Recall/AUC (e.g., AUC = 0.899); task-dependent metrics like Jaccard Index for symptom clusters
[52]	Concept extraction, Topic modeling (LDA), Word embeddings; NLP pipelines using MetaMap, cTAKES,	Hybrid of SVM, CRFs, Deep Neural Networks; MetaMap, cTAKES	Accuracy: up to 92.68%; F1-scores: 0.54–0.83; AUC: up to 0.94; Task-specific benchmarks like SemEval

feedback, a few studies are discussed in greater detail here. These were chosen, because they use new or advanced methods, apply powerful AI models, like LLMs, work well on large-scale real-world data or combine human insight with AI tools. These examples will help us better understand the latest trends to use sentiment analysis in healthcare.

Table 4. Studies utilizing large language models (LLMs) for patient-experience analysis.

Ref.	Architecture	Embedding / Features	Performance Metrics
[12]	ChatNorT5 (T5-based, 808M), NorMistral (Mistral 7B-based)	Transformer embeddings; instruction-tuned LLMs	F1: ChatNorT5 = 42.4% (4-class), 89.3% (2-class); NorMistral = 39.9% (4-class), 89.1% (2-class)
[53]	Llama2-70B, Mistral-7B, GPT- 3.5; Chatbot + Dialogue Management System	LLM embeddings, Prompt Engineering, User Profile memory, SVD, Reddit/Chatbot transcripts	Llama2 > GPT-3.5 in 40–44% of summarization tasks; GPT-4 used as evaluator; promising pilot results for chatbot system
[18]	DeBERTa, BERT, Bi-LSTM, LSTM, ChatGPT-3.5 (few- shot)	Word embeddings, Transformer-based ABSA (DeBERTa)	ChatGPT-3.5: F1 =90%; ABSA-BERT: F1 = 73.2%; BiLSTM: Acc.= 85%; Manual eval.: Cohen's Kappa = 0.87

4.1 Studies Employing ML/DL for Analyzing Sentiment in Patient Feedback

Several studies applied traditional ML methods to classify patient feedback into positive, negative and neutral sentiment categories. Feature engineering techniques, like TF-IDF, n-grams, POS tagging, have been applied followed by supervised classification algorithms, such as SVM, Naive Bayes or Logistic Regression.

The authors of [20] collected 1817 Chinese complaint cases from two hospitals from 2015 to 2019 and divided them into four categories. First, the Chinese text was translated to English using ChatGPT-3.5 and tokenization was carried out using jieba (Chinese NLP library). The features were then extracted, followed by balancing the dataset using Synthetic Minority Over-sampling Technique (SMOTE). ML techniques were then employed for classification purposes, out of which SVM gave the best accuracy value. Another study, [17], worked on patient feedback collected through the Friends and Family Test (FFT) system in the UK's National Health Service (NHS). Nearly 10% of the responses (6,900 comments) were manually labeled by an annotation team to create a training dataset for model training and themes and sentiments were derived for each comment. The study used 10 core themes adapted from the NHS Patient Experience Framework. Six ML models were then trained using the annotated dataset to automatically classify the remaining 90% of the responses, with SVM achieving the best performance. In 2021, the authors of [15] demonstrated sentiment analysis, topic modeling and text classification on the publicly available drug-review dataset. Relying on the Bing sentiment lexicon where each word is tagged as either positive or negative, sentiment analysis was performed on reviews for four specific drugs (two of which had higher positive sentiments). Further, they grouped the text data by topic (topic modeling) and manually labeled each topic by looking at the most frequent words associated with it. They also assigned good and bad labels to the reviews based on star ratings, handled data imbalance through SMOTE and utilized ML models to classify the reviews.

In 2023, the authors of [16] combined design thinking with ML to make the process of understanding and analyzing patient experience in a more accurate, detailed and useful manner. In the first study, the authors used supervised ML to analyze 14,391 cancer forum posts. They also applied association rule mining to uncover relationships between topics, which helped in refining an initial journey map. In the second study, they used unsupervised learning to analyze 30,037 online patient stories, to identify hidden themes and map them to different stages of care. This was followed by designers looking at the most common topics found and labeling them to show what patients need and how they feel at different points in their care. This mix of computer analysis and human insight helped create detailed maps of the patient journey.

A few studies also worked on developing recommendation systems and automated analysis tools. The authors of [50] analyzed patient-written drug reviews obtained from Kaggle, to recommend the most suitable medicine for a health condition. After pre-processing the dataset with TF-IDF and N-Gram models, the reviews were classified as positive or negative using ML models. The sentiment analysis was carried out by using 1-gram to 4-gram models, with the 4-gram model achieving best results.

They further ranked the drugs by average sentiment score and built a drug-recommendation system based on it. However, the original dataset did not have a dedicated sentiment column and how the sentiments were computed for model training was not mentioned by the authors in the study. Further, the authors of [19] developed a new tool called AI-PREM, which combined an open-ended patient-experience questionnaire, an NLP pipeline to automatically analyze responses and a visual interface for easily understanding the results. Patients' responses were pre-processed and sentiment analysis was conducted using a fine-tuned multi-lingual BERT model to classify the feedback. For topic modeling, the authors used Non-negative Matrix Factorization (NMF) to group similar responses based on themes, with separate models created for each question and sentiment. An interactive three-layer dashboard was developed to visualize and interpret the results.

Researchers have also integrated Social Network Analysis (SNA) and DL techniques along with ML to enhance the analysis of patient feedback. In [13], the authors analyzed unstructured patient feedback using NLP and DL. First, free-text comments were pre-processed followed by exploratory data analysis using word clouds, frequency distributions and part-of-speech tagging to identify common themes and key concerns. The authors utilized a neural network model with a sequential architecture with dense and dropout layers to classify sentiments as positive, negative or neutral. This model was especially used to separate and label comments that had both positive and negative parts, by looking at each sentence one by one. This helped get a more detailed understanding of the feedback. Another study, [14], combined ML and Social Network Analysis (SNA) to develop a system that can both predict negative patient experiences and identify key doctors who have a direct impact on those experiences. The authors classified the responses into two classes - best response and all other responses. They utilized a variety of ML classifiers to predict negative patient experiences. Further, they utilized SNA (degree, betweenness and closeness centralities) to identify influential doctors who can help improve the overall patient experience.

4.2 Studies Employing LLMs for Analyzing Sentiment in Patient Feedback

A piece of research [12] in 2024 focused on Norwegian-language feedback from patients and developed a sentiment-labeled dataset from free-text patient-survey comments. The authors used two LLM architectures with zero and few-shot learning (to guide the model with no or minimal training examples) and achieved good classification results for binary labels - positive and negative. They used 48 custom prompts based on English datasets, translated into Norwegian. However, the models failed in the case of 4-class classification achieving less than 50% accuracy values. The study highlighted the importance of manual annotation to achieve good results. Another research, [18], collected patient posts from a health forum and identified aspects that patients talk about and checked whether people spoke positively, negatively or in a neutral way using DeBERTa neural network and ChatGPT-3.5. It was found that ChatGPT performed the best in understanding detailed feedback with few-shot learning (where a few examples are provided to the model in the prompt).

5. CHALLENGES

This section addresses RQ3 by discussing the key challenges related to the collection and analysis of patient feedback. Collecting and analyzing patient feedback is essential for improving healthcare quality. However, it comes with several practical and systemic challenges that must be addressed for these systems to be effective. First, the terms “patient satisfaction” and “patient experience” create confusion, since they are used interchangeably [54]. While satisfaction is subjective and based on expectations of an individual, experience is more objective and measures what actually happened during care. Hence, satisfaction may not accurately capture the quality of care. For example, two patients undergo the same surgery with identical medical outcomes. Patient A expected a painful recovery, but found it manageable leading to high satisfaction. Patient B expected a quick, painless recovery, but experienced discomfort leading to low satisfaction.

There can be many reasons for patients not giving feedback - low literacy in health, socio-economic inequalities, fear of being treated unfairly because of giving negative feedback and lack of trust in healthcare systems. In low-income and middle-income countries, many patients are unaware that feedback mechanisms even exist [55]. Moreover, there is an absence of clear guidelines and health workers also take feedback mechanisms as a threat rather than a scope to improve. They are reluctant to receive patient feedback fearing that negative feedback may harm their professional repute. Some institutions do not even integrate patient feedback into strategic planning effectively, since negative feedback over-shadows positive comments. Bias and reliability issues also arise while feedback is being collected, since it is influenced by the emotions and health conditions of the patients. Further, patients, being both a care recipient and a feedback provider, feel conflicted [56]. Also, healthcare professionals, being both experts and learners, are hesitant to invite feedback. Hence, there is an imbalance of power where patients may hesitate to provide negative feedback and professionals may feel vulnerable when receiving criticism. There is a lack of structured methods for engaging in feedback dialogues. Patients prefer verbal feedback for positive experiences, but written feedback when dissatisfied. Even after the feedback is collected, there are hardly any mechanisms for following it up and even if actions are taken, patients are hardly informed about them. Hence, participation is decreased over time.

Analyzing the collected feedback comments to get useful insights for decision-making can be expensive and time-consuming if carried out manually. Utilizing ML and DL techniques to process and analyze such unstructured data also requires careful intervention. These models should be carefully selected and validated, especially in healthcare contexts, where misclassification can have serious consequences. Further, LLMs like LLaMA and GPT are also very expensive to train and require significant resources.

6. Conclusion and Future Directions

This study has provided a thorough review of current methods for collecting and analyzing patient feed- back in healthcare. It examined both traditional tools, such as open-ended questionnaires and interviews and emerging digital platforms that support scalable and timely feedback collection. A particular emphasis was placed on sentiment analysis techniques, showcasing the application of machine learning (ML), deep learning (DL) and large language models (LLMs) to interpret unstructured patient responses. The review synthesized findings from recent studies, detailing the datasets used, feature-extraction strategies, classification approaches and performance outcomes. Furthermore, challenges and limitations associated with data collection, processing and analysis were discussed. By aligning sentiment analysis techniques with real-world feedback systems, this review supports the development of automated and patient-centered solutions that can enhance service quality and enable continuous healthcare improvement.

In future work, feedback systems should be designed to function across multiple platforms, such as mobile apps, websites, SMS, in-person interviews and voice input, to increase participation from diverse patient populations. Also, family members should be allowed to submit feedback on behalf of elderly or critically ill patients, to expand the scope of feedback collection. The process of feedback collection and analysis should be automated using NLP and AI tools to reduce manual efforts and analyze large amounts of data. Moreover, there is a lack of publicly available patient feedback datasets. Future work should focus on curating and sharing large-scale, representative datasets to improve the generalizability and robustness of sentiment-analysis models, across different demographics, languages and care settings. Lastly, feedback gathered must be fed directly into quality-improvement programs, performance evaluations and strategic planning.

References

[1] J. A. Wolf et al., "Reexamining “Defining Patient Experience”: The Human Experience in Healthcare," Patient Experience Journal, vol. 8, no. 1, pp. 16– 29, 2021.

[2] R. Kalaja, "Determinants of Patient Satisfaction with Health Care: A Literature Review," European Journal of Natural Sciences and Medicine, vol. 6, pp. 43–54, May 2023.

[3] B. Brown et al., "Clinical Performance Feedback Intervention Theory (CP-FIT): A New Theory for Designing, Implementing and Evaluating Feedback in Health Care Based on a Systematic Review and Meta-synthesis of Qualitative Research," Implementation Science, vol. 14, no. 1, pp. 1–20, 2019.

[4] M. Khanbhai et al., "Applying Natural Language Processing and Machine Learning Techniques to Patient Experience Feedback: A Systematic Review," BMJ Health Care Informatics, vol. 28, p. e100262, March 2021.

[5] M. Nasori, M. Mak-van der Vossen, M. Holtrop and J. Bont, "Exploring Effective Patient Feedback Methods for e-Health in General Practice," BMC Primary Care, vol. 26, p. 40, 2025.

[6] R. Gualandi et al., "What Does the Patient Have to Say? Valuing the Patient Experience to Improve the Patient Journey," BMC Health Services Research, vol. 21, pp. 1–12, 2021.
[7] S. M. Bez, I. Georgescu and M. S. Farazi, "TripAdvisor of Healthcare: Opportunities for Value Creation through Patient Feedback Platforms," Technovation, vol. 121, p. 102625, 2023.
[8] J. Kaipio et al., "Improving Hospital Services Based on Patient Experience Data: Current Feedback Practices and Future Opportunities," Studies in Health Technology and Informatics, vol. 247, pp. 266–270, 2018.
[9] S. Berger, A. M. Saut and F. T. Berssaneti, "Using Patient Feedback to Drive Quality Improvement in Hospitals: A Qualitative Study," BMJ Open, vol. 10, no. 10, p. e037641, 2020.
[10] C. M. Bastemeijer et al., "Patient Experience Monitor (PEM): The Development of New Short-form Picker Experience Questionnaires for Hospital Patients with a Wide Range of Literacy Levels," Patient Related Outcome Measures, vol. 11, pp. 221–230, 2020.
[11] E. Wong, F. Mavondo and J. Fisher, "Patient Feedback to Improve Quality of Patient-centred Care in Public Hospitals: A Systematic Review of the Evidence," BMC Health Services Research, vol. 20, no. 1, p. 530, 2020.
[12] P. Mæhlum et al., "It‘s Difficult to be Neutral–human and LLM-based Sentiment Annotation of Patient Comments," Proc. of the 1st Workshop on Patient-oriented Language Processing (CL4Health), pp. 8–19, Torino, Italia, May 2024.
[13] K. Nawab, G. Ramsey and R. Schreiber, "Natural Language Processing to Extract Meaningful Information from Patient Experience Feedback," Applied Clinical Informatics, vol. 11, pp. 242–252, March 2020.
[14] V. Bari et al., "An Approach to Predicting Patient Experience through Machine Learning and Social Network Analysis," J. of the American Medical Informatics Association, vol. 27, pp. 1834–1843, 2020.
[15] C. J. Harrison and C. J. Sidey-Gibbons, "Machine Learning in Medicine: A Practical Introduction to Natural Language Processing," BMC Medical Research Methodology, vol. 21, no. 1, p. 158, 2021.
[16] J. Jung, K.-H. Kim, T. Peters, D. Snelders and M. Kleinsmann, "Advancing Design Approaches through Data-driven Techniques: Patient Community Journey Mapping Using Online Stories and Machine Learning," Int. Journal of Design, vol. 17, no. 2, 2023.
[17] M. Khanbhai et al., "Using Natural Language Processing to Understand, Facilitate and Maintain Continuity in Patient Experience across Transitions of Care," Int. Journal of Medical Informatics, vol. 157, p. 104642, 2022.
[18] O. S. Alkhnbashi, R. Mohammad and M. Hammoudeh, "Aspect-based Sentiment Analysis of Patient Feedback Using Large Language Models," Big Data and Cognitive Computing, vol. 8, no. 12, 2024.
[19] M. M. van Buchem et al., "Analyzing Patient Experiences Using Natural Language Processing: Development and Validation of the Artificial Intelligence Patient Reported Experience Measure (AI-PREM)," BMC Medical Informatics and Decision Making, vol. 22, no. 1, p. 183, 2022.
[20] X. Li, Q. Shu, C. Kong, J. Wang, G. Li, X. Fang, X. Lou and G. Yu, "An Intelligent System for Classifying Patient Complaints Using Machine Learning and Natural Language Processing: Development and Validation Study," Journal of Medical Internet Research, vol. 27, p. e55721, 2025.
[21] A. Azarpey et al., "Natural Language Processing of Sentiments Identified in Patient Comments Associated with Less than Top-rated Care," J. of Patient Experi., vol. 12, p. 23743735251323677, 2025.
[22] A. Esuli and F. Sebastiani, "SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining," Proc. of the 5th Int. Conf. on Language Resources and Evaluation (LREC), pp. 417–422, Genoa, Italy, 2006.
[23] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," Proc. of the 8th Int. AAAI Conf. on Weblogs and Social Media, vol. 8, no. 1, Association for the Advancement of Artificial Intelligence, 2014.
[24] F. Nielsen, "Afinn: A New Word List for Sentiment Analysis," [Online], Available at: https://github.com/fnielsen/afinn, 2011.
[25] M. Wankhade, A. C. S. Rao and C. Kulkarni, "A Survey on Sentiment Analysis Methods, Applications and Challenges," Artificial Intelligence Review, vol. 55, no. 8, pp. 5731–5780, 2022.
[26] T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," Proc. of the European Conf. on Machine Learning (ECML), pp. 137–142, Springer, 1998.
[27] G. Salton and C. Buckley, "Term-weighting Approaches in Automatic Text Retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
[28] T. Mikolov, K. Chen, G. Corrado and J. Dean, "Efficient Estimation of Word Representations in Vector Space," arXiv preprint, arXiv: 1301.3781, 2013.
[29] J. Pennington et al., "GloVe: Global Vectors for Word Representation," Proc. of the 2014 Conf. on Empirical Methods in Natural Language Processing (EMNLP), pp. 1532–1543, ACL, 2014.
[30] G. A. Seber and A. J. Lee, Linear Regression Analysis, John Wiley & Sons, 2nd Edn., 2012.
[31] J. R. Quinlan, Induction of Decision Trees, vol. 1, Springer, 1986.

[32] I. Rish, "An Empirical Study of the Naive Bayes Classifier," IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence, vol. 3, no. 22, pp. 41–46, 2001.
[33] C. Cortes and V. Vapnik, "Support-vector Networks," Machine Learning, vol. 20, no. 3, pp. 273–297, 1995.
[34] D. W. Hosmer, S. Lemeshow and R. X. Sturdivant, Applied Logistic Regression, John Wiley & Sons, 3rd Edn., 2013.
[35] T. Cover and P. Hart, "Nearest Neighbor Pattern Classification," IEEE Transactions on Information Theory, vol. 13, no. 1, pp. 21–27, 1967.
[36] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
[37] T. B. Brown et al., "Language Models Are Few-shot Learners," Advances in Neural Information Processing Systems, vol. 33, pp. 1877–1901, 2020.
[38] A. Vaswani et al., "Attention Is All You Need," Advances in Neural Information Processing Systems, vol. 30, Curran Associates, Inc., 2017.
[39] Y. Wu and G. Hu, "Exploring Prompt Engineering with GPT Language Models for Document-level Machine Translation: Insights and Findings," Proc. of the 8th Conf. on Machine Translation, pp. 166–169, Association for Computational Linguistics, 2023.
[40] J. Devlin, M.-W. Chang, K. Lee and K. Toutanova, "Bert: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. of the 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, pp. 4171–4186, 2019.
[41] Y. Liu et al., "RoBERTa: A Robustly Optimized BERT Pretraining Approach," arXiv preprint, arXiv:1907.11692, 2019.
[42] G. Lysandrou et al., "Classifying Patient Voice in Social Media Data Using Neural Networks: A Comparison of AI Models on Different Data Sources and Therapeutic Domains," arXiv preprint, arXiv:2312.03747, 2023.
[43] M. Amoei and D. Poenaru, "Patient-centered Data Science: An Integrative Framework for Evaluating and Predicting Clinical Outcomes in the Digital Health Era," arXiv preprint, arXiv:2408.02677, 2024.
[44] S. Jadhav et al., "A Systematic Review on the Role of Sentiment Analysis in Healthcare," Authorea, DOI: 10.22541/au.172516418.81416769/v1, September 2024.
[45] D. Panchal, M. Shelke, S. Kawathekar and S. Deshmukh, "Prediction of Healthcare Quality Using Sentiment Analysis," Indian Journal of Science and Technology, vol. 16, no. 21, pp. 1603–1613, 2023.
[46] A. Yazdani et al., "Use of Sentiment Analysis for Capturing Hospitalized Cancer Patients' Experience from Free-text Comments in the Persian Language," BMC Medical Informatics and Decision Making, vol. 23, no. 1, p. 275, 2023.
[47] P. Smith, Sentiment Analysis of Patient Feedback, PhD Thesis, University of Birmingham, 2017.
[48] J. Sim et al., "Natural Language Processing with Machine Learning Methods to Analyze Unstructured Patient-reported Outcomes Derived from Electronic Health Records: A Systematic Review," Artificial Intelligence in Medicine, vol. 146, p. 102701, 2023.
[49] L. Cho et al., "Sentiment Analysis of Online Patient-written Reviews of Vascular Surgeons," Annals of Vascular Surgery, vol. 88, pp. 249– 255, 2023.
[50] T. Shahid, S. Singh, S. Gupta and S. Sharma, "Analyzing Patient Reviews for Recommending Treatment Using NLP and Deep Learning-based Approaches," Proc. of the Int. Conf. on Advancements in Interdisciplinary Research, pp. 179–190, Springer, 2022.
[51] D. C, K. TA, B. PE and B. S, "A Systematic Review of Natural Language Processing and Text Mining of Symptoms," Int. Journal of Medical Informatics, vol. 125, pp. 37–46, 2019.
[52] G. Gonzalez-Hernandez et al., "Capturing the Patient's Perspective: A Review of Advances in Natural Language Processing of Health-related Text," IMIA Yearbook of Medical Inform., pp. 214–227, 2017.
[53] B. Wen, R. Norel, J. Liu, T. Stappenbeck, F. Zulkernine and H. Chen, "Leveraging Large Language Models for Patient Engagement: The Power of Conversational AI in Digital Health," Proc. of the 2024 IEEE Int. Conf. on Digital Health (ICDH), pp. 104–113, July 2024.
[54] B. Berkowitz, "The Patient Experience and Patient Satisfaction: Measurement of a Complex Dynamic," Online Journal of Issues in Nursing, vol. 21, no. 1, 2016.
[55] T. Mirzoev, S. Kane, Z. Al Azdi, B. Ebenso, A. A. Chowdhury and R. Huque, "How Do Patient Feedback Systems Work in Low-income and Middle-income Countries? Insights from a Realist Evaluation in Bangladesh," BMJ Global Health, vol. 6, no. 2, p. e004357, 2021.
[56] C. Sehlbach, M. H. Bosveld, S. Romme, M. A. Nijhuis, M. J. Govaerts and F. W. Smeenk, "Challenges in Engaging Patients in Feedback Conversations for Health Care Professionals' Workplace Learning," Medical Education, vol. 58, no. 8, pp. 970–979, 2024.
[57] S. S. Ibrahiem, S. S. Ismail, K. A. Bahnasy and M. M. Aref, "Convolutional Neural Network Multi-Emotion Classifiers," Jordanian Journal of Computers and Information Technology (JJCIT), vol. 05, no. 02, pp. 97-108, August 2019.

ملخص البحث:

تلعب التّغذية الرّاجعة من المرضى دوراً حاسماً في تحسين الجودة للرّعاية الطّبية المتمركزة حول المريض فيما يتعلّق بخدمات الرّعاية. تقدم هذه الورقة مراجعةً شاملةّ للطُّرق التّقليدية والطُّرق الرّقمية المستخدمة في جمع التّغذية الرّاجعة من المرضى، مع التّركيز على ما لتلك الطُّرق من قيمةٍ في تحسين تقديم الرّعاية الطّبية للمرضى، كما تفحص الأدوات والقنوات المستخدمة في ذلك، بما فيها المسوحات والمقابلات والمنصّات الرّقمية متعدّدة القنوات.

من ناحيةٍ أخرى، يناقش هذا البحث تقنيات تحليل المشاعر المطبّقة على بيانات التّغذية الرّاجعة من المرضى، مع التّركيز على الكيفية الّتي تعمل بها تقنيات التّعلُّم الآلي والتّعلُّم العميق والنّماذج اللّغوية الضّخمة على تفسير البيانات وتبويبها في التّغذية الرّاجعة من المرضى. يتم تحليل الأدبيات ذات العلاقة بطريقةٍ منظّمة، إلى جانب جداول مقارنة تسلّط الضّوء على طُرق استخلاص السِّمات، وخوارزميات التّصنيف، ومؤشّرات الأداء المستخدمة في الدّراسات السابقة المتعلقة بالموضوع.

كذلك تتناول هذه الورقة التّحدّيات الأساسية الّتي تنطوي عليها عملية جمع بيانات التّغذية الرّاجعة من المرضى وعملية تحليل المشاعر. ويمكن للبحوث المستقبلية أن تبحث في أتمتة أنظمة جمع وتحليل التّغذية الرّاجعة من المرضى وتضمين وجهة نظر المرضى أنفسهم في أُطر العمل الخاصّة بتحسين جودة الرّعاية الطّبية، إلى جانب الاستفادة من الذّكاء الاصطناعي في ذلك.