

Enhancing Public Healthcare Through VADER Sentiment Analysis: A Case Study on Patient Complaints

Journal of Patient Experience
Volume 13: 1-9
© The Author(s) 2026
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/23743735251413861
journals.sagepub.com/home/jpx



João Vasco Coelho, PhD¹ and Liliana da Costa Barbosa, BSc²

Abstract

This study investigates the application of Valence Aware Dictionary and Sentiment Reasoner (VADER), a rule-based sentiment analysis tool, for rapidly analyzing patient complaints in healthcare environments. By leveraging VADER's efficiency in processing free-text data, the research examines how sentiment analysis can support digital transformation and enhance patient experience management. The study assesses VADER's effectiveness in detecting emotional tone within patient feedback, enabling early identification of systemic issues and informing data-driven decision-making. The dataset comprises 63 written complaints collected throughout 2024 from a key surgery service in a 10 000-employee public hospital in Portugal. Sentiment analysis was conducted using Orange text mining tools integrated with VADER. Three core findings highlight VADER's value: (a) its balance of efficiency and accuracy in sentiment classification, (b) its contribution to patient-centered care strategies, and (c) its support for process optimization, particularly in triage and prioritization workflows. These insights demonstrate how sentiment analysis can help build smarter, more responsive healthcare systems through enhanced digital capabilities.

Keywords

natural language processing, sentiment analysis, text classification, VADER sentiment analysis, healthcare, patient experience, patient complaints

MSC classification: 68T50, 92C50

JEL classification: I18, O33

Introduction

Complaints submitted by patients and their families are increasingly recognized not merely as expressions of dissatisfaction, but as pivotal insights into care quality, communication breakdowns, and unmet needs within healthcare delivery systems. As unstructured textual data, these narratives can serve as vital inputs for service redesign and quality improvement strategies. However, their full potential remains untapped due to challenges in manual analysis. This study addresses that gap by exploring the application of Valence Aware Dictionary and Sentiment Reasoner (VADER) sentiment analysis to swiftly and reliably interpret patient complaint data, thereby advancing digital capabilities that support responsive, *patient-centered* care.

As healthcare shifts toward patient-centered models, free-text data offers richer insights than traditional closed-ended surveys, helping capture nuanced patient experiences.¹⁻³

To support this intent and develop patient-oriented healthcare services, systems and processes need to be improved to collect, organize, analyze, and quickly address patient and family complaints, and establish an accurate *patient's voice*, which is proposed to be one of the core building blocks for

¹ Centro de Investigação e Estudos em Sociologia (CIES), Instituto Universitário de Lisboa (ISCTE-IUL), Lisbon, Portugal

² Research Centre in Asset Management and System Engineering (RCM2+), Instituto Superior de Engenharia de Coimbra (ISEC), Coimbra, Portugal

Corresponding Author:

João Vasco Coelho, Centro de Investigação e Estudos em Sociologia (CIES), Instituto Universitário de Lisboa (ISCTE-IUL), Edifício Sedas Nunes, Av. das Forças Armadas, 1649-026 Lisboa, Portugal.

Email: vasco.jcoelho@gmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

system building in contemporary complaint management in the healthcare sector.¹⁻³

To extract value from rich free-text data regarding patient experience, establishing trust by devising quicker case resolution loops is considered to be of key relevance.⁴⁻⁹ Regarding this intent, it is appropriate to say that the scale and diversity of free-text feedback make it difficult for human agents to swiftly extract and digest actionable information. Sentiment analysis, a pivotal technique in natural language processing (NLP) that involves identifying and extracting subjective information from textual data, can help automate this process by classifying textual feedback into emotional categories such as positive, neutral, or negative.² Beyond basic polarity detection, advanced sentiment models can also detect intensity and emotional cues, offering a nuanced understanding of customer grievances.⁵⁻⁹

VADER, a rule-based sentiment analysis tool, is considered to be a reliable solution in this domain.² It was initially designed to perform on social media text, but is increasingly being tested and valued as effective in other domains.¹⁰⁻¹³ VADER combines a robust lexicon with heuristic rules to capture contextual nuances in sentiment, making it easy to use and highly accurate.¹³⁻¹⁶

In a context (healthcare) where the use of sentiment analysis, namely in terms of patient experience management support, is considered to be still significantly underexplored,¹³ this paper aims to explore the extent to which rule-based sentiment analysis tools such as VADER can contribute to more efficient, swift, and fair patient complaints resolution by quickly establishing an accurate patient's voice,¹⁻³ in complex and critical sectors such as healthcare and large-sized organizations aiming to develop their technological capabilities such as public hospitals.

The study's dataset consists of the total number and content of patient complaints ($n = 63$) collected throughout 2024, in one of the key surgery services of a 10 000-employee public hospital located in Portugal. In terms of research methods, the analytical apparatus was based in intertwining the usage of Orange text mining capabilities with its VADER sentiment analysis widget, and in comparing outputs, for accuracy control, with the ones derived a grouped sentiment analysis using Bidirectional Encoder Representations from Transformers (BERT), a transformer-based model. The paper is organized as follows: the second section presents a review of the current state of the art, the third section presents the dataset and methods that were used, the fourth and fifth sections present the research results and respective discussion, including observed challenges and limitations. A conclusion section presenting a vision for future work ends the paper.

State of the Art

Over the last decade, sentiment analysis, conversational artificial intelligence (AI), and opinion mining have been emerging as critical areas of research, with significant applications across industries such as e-commerce, healthcare, social

media, and customer service.^{2,13-16} During this timeframe, sentiment analysis methods, in concrete, have developed significantly, driven by advances in NLP and machine learning techniques. Current primary methodologies in the sentiment analysis field include¹⁰⁻¹⁶:

1. *Rule-based, lexicon-based* approaches that rely on predefined linguistic rules to identify and extract sentiment from text; these approaches typically involve using predefined dictionaries of sentiment-bearing words to determine the sentiment orientation of text; each word in the lexicon is assigned a sentiment score, and the overall sentiment of a text is calculated based on the presence and scores of these words. While these methods are simple and easy to implement, they often struggle with context, sarcasm, and nuanced expressions. Despite the limitations, these methods are still widely used due to their interpretability and low computational requirements.
2. *Machine learning* models trained on labeled datasets, learning patterns in the data to predict sentiments in new, unseen text; these techniques offer greater flexibility and accuracy compared to rule-based approaches.
3. *Deep learning techniques*, such as long short-term memory networks (LSTM), convolutional neural networks (CNNs), and transformer-based models (eg, BERT) have demonstrated superior performance in sentiment analysis, by processing complex representations of text that enhance sentiment classification accuracy.

Closely intertwined with the aforementioned developments, sentiment analysis methods have become increasingly explored and applied in industries and domains where understanding customer and user sentiments can drive more effective managerial decision-making and improve overall experiences.⁴⁻⁹ For instance, businesses have been leveraging sentiment analysis to gain insights into customer feedback, enhancing product offerings and service quality.¹³ In healthcare, sentiment analysis has been explored to aid in monitoring patient opinions and emotions, contributing to better patient care and satisfaction.^{9,16}

As healthcare shifts from a *disease-centered* to a patient-centered approach, free-text data is expected to play an increasingly vital role in capturing and promptly handling patient experiences.^{2,3} To derive meaningful insights from rich free-text data on patient experiences, build trust and establish faster resolution loops, it is essential to be able to systematize complaints and complaint data related to health services.⁴⁻⁹ To advance this goal and foster patient-oriented healthcare services, systems and processes must be enhanced to efficiently organize and address patient and family complaints. Establishing an accurate *patient voice* is proposed as a fundamental building block in modern complaint management within the healthcare sector. As illustrated by previous studies,^{3,4,9,13,16} it is frequent to find ambiguity

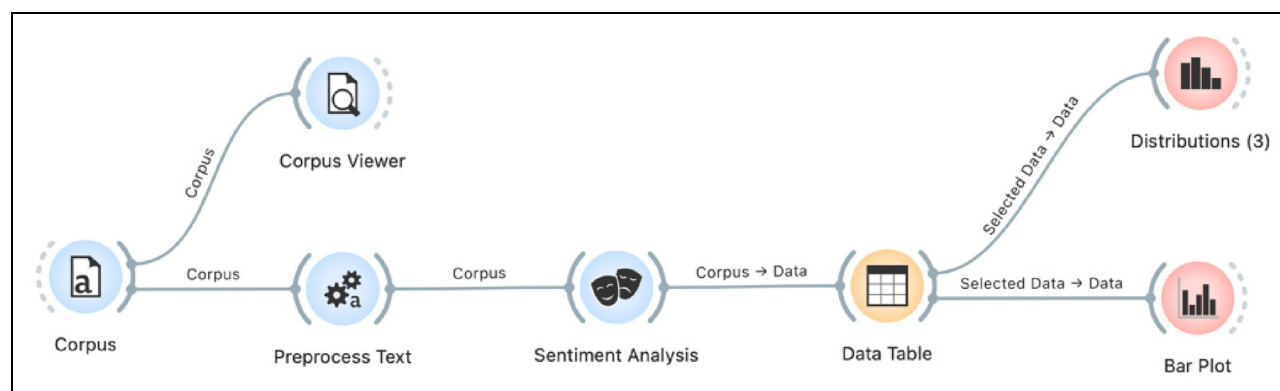


Figure 1. Text mining model (using Orange).

and nuances in free-text patient complaints, making it difficult to identify appropriate improvement actions. Complaints caused by hospital staff, as opposed to processes or environment (eg, doctors, nurses, and other staff) are often more frequent and more reported as complaints triggered by a range of events or attributes (eg, communication, attitude, nursing skill), with direct comparisons of complaint classification between hospitals and countries being difficult to do.^{3,4,9}

Despite the recent observed advancements in the healthcare space, sentiment analysis, particularly in patient experience management through free-text data mining, remains significantly underexplored.¹³ However, it continues to gain traction as healthcare organizations strive to enhance and optimize the services provided to patients and their families, in particular through process digitalization and automation. The present study aims to contribute to this progress.

Methodology

Materials

The dataset that was used in the study consisted of the total number and content of patient complaints ($n = 63$) collected during 2024, from one of the key healthcare services of a 10 000-employee public hospital located in Portugal. Complaints were collected between January and December 2024, submission channels included email, postal mail, and in-person forms, typically completed postdischarge. Given its size, nature (general surgery), and exposure to the subject considered in the research (patient experience management in healthcare organizations with developing digital capabilities), this healthcare service was positioned as an *exemplary*, pregnant case to address, in empirical terms, the research proposed intents. The reliance of the service on free-text survey data to handle patient complaints was a key factor supporting its suitability to address the research objectives in empirical terms.

Methods

Data Preprocessing. Data preprocessing included the following procedures: (i) Translation: complaints data was translated from Portuguese to English using DeepL Pro; (ii) stopword removal was ensured using Natural Language Toolkit (NLTK)'s English stopwords list; (iii) lemmatization was performed with spaCy's English model; and (iv) noise reduction was ensured by manual review which removed nontextual artifacts and informal language. Figure S1 in the Supplemental materials presents the dataset labeling per feedback type.

Data Analysis. Analysis was made by intertwining the usage of Orange text mining capabilities with its VADER sentiment analysis widget. In Orange, VADER is used to analyze English-language text, providing positive, negative, neutral, and compound sentiment scores. These scores help users understand the overall sentiment of the established *corpus*. For secondary comparative purposes (in terms of compound sentiment analysis scores accuracy), a grouped sentiment analysis using BERT (used for nuanced and complex sentence structures analysis)¹⁷ was made. Model evaluation was conducted including performance metrics, based on interrater validation of a complaints data sample. The text mining model that was used is presented in Figure 1. Sentiment analysis was conducted using Orange Data Mining (v3.34). Comparative analysis was performed using Python 3.10 and the Transformers library (v4.35) for BERT-based classification.

Composite scores were obtained by summing the valence scores of each word in the lexicon, including adjustments based on predefined rules, and normalizing the score to a range between -1 (extremely negative) and $+1$ (extremely positive). In the classification process, the following criteria was used to determine the sentiment (positive, negative, or neutral) of each comment¹⁴: *positive sentiment*: compound score ≥ 0.5 ; *neutral sentiment*: (compound score > -0.5) and (compound score < 0.5); *negative sentiment*: compound score ≤ -0.5 .

	Text	Label	positive	negative	neutral	compound	✓
63	**Good morn...	Praise	0.264	0.046	0.69	0.9994	
19	**Subject:** ...	Praise	0.173	0.027	0.8	0.9981	
54	Good evenin...	Praise	0.289	0	0.711	0.9961	
35	Dear Sirs,I a...	Praise	0.282	0.012	0.705	0.9955	
56	**Dear Sirs,*...	Complaint	0.115	0.074	0.811	0.9938	
6	Subject: Con...	Praise	0.237	0	0.763	0.9924	
30	Luís Joaquim...	Praise	0.311	0	0.689	0.9869	
53	Dear Sir,Hea...	Praise	0.423	0	0.577	0.9863	
34	**Subject:** ...	Dispute	0.106	0.04	0.854	0.9858	
55	Dear Profess...	Praise	0.28	0	0.72	0.9773	
42	I would like t...	Praise	0.416	0	0.584	0.976	
58	**To the Citiz...	Complaint	0.092	0.014	0.893	0.9698	
52	Dear Dr. Alex...	Praise	0.217	0.022	0.761	0.9595	
61	To the entire ...	Praise	0.231	0.04	0.728	0.9504	
62	Dear Profess...	Praise	0.394	0	0.606	0.9423	
13	Good aftern...	Complaint	0.125	0	0.875	0.936	
5	Good mornin...	Complaint	0.229	0.037	0.734	0.9235	
20	**Subject:** ...	Complaint	0.106	0.013	0.881	0.9197	
21	**Subject:** ...	Complaint	0.161	0	0.839	0.9125	
28	Request for ...	Complaint	0.141	0	0.859	0.8916	
48	Exmo. Sirs,I ...	Dispute	0.068	0.054	0.878	0.8799	
60	I would like t...	Praise	0.234	0	0.766	0.8658	
27	Complaint R...	Complaint	0.138	0.085	0.777	0.7707	
18	Urgent Com...	Complaint	0.126	0.1	0.775	0.7523	
46	Dear Sir/Mad...	Complaint	0.061	0.044	0.895	0.743	
11	Whom It May...	Complaint	0.103	0.081	0.816	0.7261	
50	Good mornin...	Complaint	0.095	0.064	0.841	0.6871	
14	**To the Dire...	Complaint	0.121	0.099	0.781	0.5759	
24	**Subject:** ...	Complaint	0.046	0.028	0.926	0.5565	
41	To the Estee...	Complaint	0.052	0.04	0.908	0.4776	
47	Good mornin...	Complaint	0.064	0.051	0.885	0.2732	
29	They called ...	Complaint	0	0	1	0	

Figure 2. VADER output data table with negative emotional tone (compound score).

Abbreviation: VADER, Valence Aware Dictionary and Sentiment Reasoner.

In terms of analytical results presentation, given the research goal of exploring the potential usefulness or appropriateness of using a sentiment analysis technique and tool as means to digest patient complaints in healthcare settings and support quick resolution actions and/or decisions by key stakeholders, it was decided to split analytical outputs per emotional tone—*negative* sentiments versus *neutral* or *positive*, regardless the nature of the feedback being presented by the patients (eg, a complaint, a praise, or a dispute/contestation).

Results

Key Findings

Overall, 3 main research results should be highlighted: (1) the existence of a significant number of neutral and positive

scores in complaints free-text suggesting that, although labeled as complaints, the tone is not necessarily (strongly) emotional; (2) where an overall positive emotional tone was observed (compound score), text labeling was mixed (complaint, praise, and/or disputes/contestations); and (3) to be able to identify and differentiate these nuances in a timely manner, can be particularly useful for users and decision-makers to quickly trigger (or not) follow-up case resolution actions.

In a more qualitative perspective, complementing VADER's sentiment analysis output, it was possible to summarize the most frequent content feedback and link it with negative toned feedback (eg, misdiagnosis and incorrect surgery scheduling, postsurgery pain and neglect, poor communication/updates), neutral feedback (eg, constructive

	Text	Label	positive	negative	neutral	compound ^
15	**COMPLAI...	Complaint	0.038	0.125	0.838	-0.9984
23	**Subject:** ...	Complaint	0.058	0.134	0.808	-0.996
33	**Subject:** ...	Dispute	0.076	0.167	0.757	-0.9948
7	Good Mornin...	Complaint	0.075	0.129	0.796	-0.9921
49	Formal Com...	Complaint	0.056	0.162	0.783	-0.9912
16	To Whom It ...	Complaint	0.05	0.189	0.761	-0.9904
59	**Outeiro Ma...	Complaint	0.064	0.117	0.819	-0.9863
10	**NCC. 998...	Complaint	0.049	0.16	0.791	-0.9861
4	Suspected M...	Complaint	0.063	0.102	0.835	-0.9857
12	**To Whom I...	Complaint	0.098	0.129	0.773	-0.9842
17	Urgent Com...	Complaint	0.061	0.11	0.83	-0.9784
31	**Subject:** ...	Complaint	0.071	0.082	0.847	-0.97
40	Dear Sir/Mad...	Complaint	0.122	0.137	0.741	-0.9692
45	Hello, good a...	Complaint	0.033	0.067	0.901	-0.9682
32	**Subject:** ...	Complaint	0.079	0.119	0.802	-0.9584
2	I underwent ...	Complaint	0.042	0.158	0.8	-0.9555
36	Good aftern...	Complaint	0.062	0.093	0.845	-0.9555
1	I, Maria, a pa...	Complaint	0.055	0.116	0.829	-0.9544
39	**Case Num...	Complaint	0.016	0.049	0.935	-0.9523
26	**Subject:** ...	Complaint	0.049	0.087	0.864	-0.9458
43	Dear Sir/Mad...	Complaint	0.051	0.071	0.879	-0.8373
57	Dr. José Guil...	Complaint	0.084	0.097	0.819	-0.8305
37	I've been wai...	Complaint	0	0.169	0.831	-0.7579
51	Dear Sirs,I a...	Complaint	0.045	0.06	0.895	-0.7556
3	My complain...	Complaint	0	0.067	0.933	-0.7227
8	Good aftern...	Complaint	0.113	0.127	0.759	-0.712
25	The complai...	Complaint	0	0.113	0.887	-0.6705
38	I went to the ...	Complaint	0.034	0.092	0.874	-0.6483
44	I've been wai...	Complaint	0	0.065	0.935	-0.3506
22	**Subject:** ...	Complaint	0.075	0.076	0.849	-0.0451
9	Dear HUC, ...	Complaint	0.069	0.086	0.846	-0.0323

Figure 3. VADER output data table with a neutral or positive emotional tone (compound score).
Abbreviation: VADER, Valence Aware Dictionary and Sentiment Reasoner.

suggestions of service improvements), and positive feedback (eg, gratitude for the professionalism and care).

Figure 2 presents a sample of the output that was obtained with VADER that had a negative tone; Figure 3 presents an output sample where a neutral or a positive tone was observed. In Figure 3, that aggregates negative tone cases, most texts were labeled as complaint, with 1 case of dispute/contestation. Compound scores range from -0.9984 (extremely negative) to -0.0323 (slightly negative), indicating VADER's ability to differentiate the intensity of the sentiment expressed in a complaint. Most complaints have negative compound scores (see Figure S2 in the Supplemental materials), reflecting predominantly negative sentiments. The coexistence of neutral and positive scores in complaint texts suggests that, although they are classified as complaints, the tone is not necessarily (strongly) emotional.

In Figure 3, despite overall positively toned answers, text labeling is mixed—complaint, praise, or disputes/contestations were observed. Compound scores range from 0

(neutral) to 0.9994 (extremely positive), indicating the intensity of the sentiment. The negative scores are generally low, with most values close to 0, indicating minimal negative sentiment. Neutral scores are high (ranging from 0.577 to 1), suggesting that many texts have a neutral tone. Some of these cases likely do not require the immediate direct intervention of a healthcare service member.

Comparative Analysis With BERT

For analytical control and output comparison purposes, BERT was used to double-check the accuracy of the results achieved with VADER. BERT grouped results are presented in Figure 4.

Overall, as illustrated in the literature,² the overall orientation of VADER and BERT results can be considered to be similar, namely the existence of a significant number of neutral and positive scores in complaints free-text. VADER

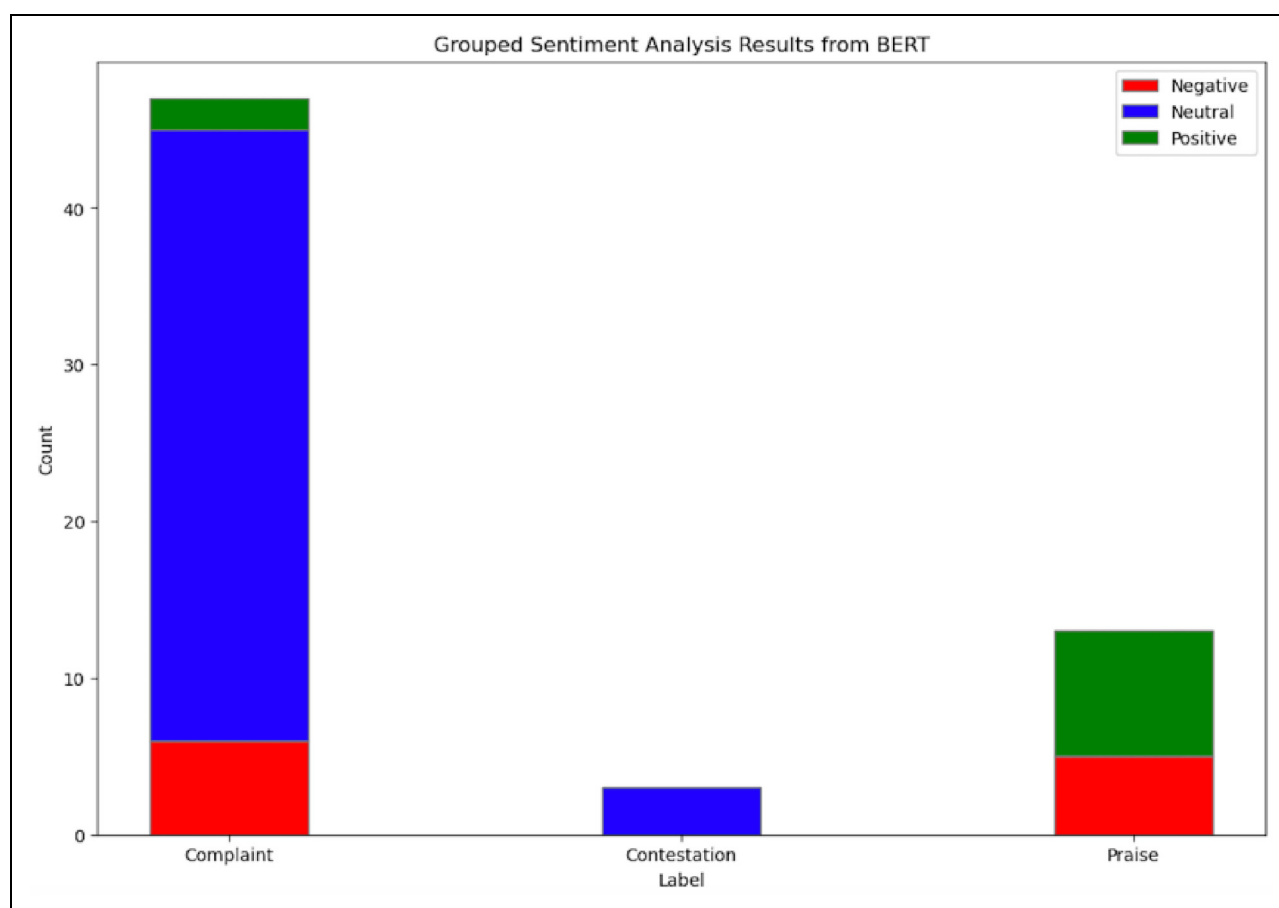


Figure 4. BERT output data table—grouped sentiment analysis.
Abbreviation: BERT, Bidirectional Encoder Representations from Transformer.

analysis, given its rule-based nature, is more straightforward in terms of legibility and call-to-action, but the chance of missing subtle sentiments or contextual-dependent details exists (eg, a sarcastic comment can be classified as positive; overestimation of neutral or positive sentiments in ambiguous cases can be observed). BERT results are likely to be more accurate, especially for nuanced or sarcastic text. A sarcastic comment classified as neutral by VADER is likely to be classified as negative by BERT, thus providing approach to analytical accuracy.¹⁸ A subset of 20 complaints was manually annotated by 2 independent reviewers as means to support model evaluation. Interrater agreement was high (Cohen's $\kappa=0.82$). Performance metrics are presented in Figure 5.

Model	Accuracy	Precision	Recall	F1 Score
VADER	0.76	0.78	0.74	0.76
BERT	0.84	0.86	0.82	0.84

Figure 5. VADER versus BERT model evaluation.
Abbreviations: VADER, Valence Aware Dictionary and Sentiment Reasoner; BERT, Bidirectional Encoder Representations from Transformer.

Benchmarks for VADER in general-purpose sentiment tasks often hover around 0.70 to 0.75 accuracy, depending on the dataset.¹⁵ Research results (accuracy: 0.76, $F1$: 0.76) are therefore slightly above average, which is significant given the complexity of healthcare complaints.^{5–9} BERT-based models often achieve 0.80 to 0.90 accuracy when fine-tuned.¹⁷ Research results (accuracy: 0.84, $F1$: 0.84) are right in line with solid performance.

Discussion

Practical Contributions of VADER

The research demonstrates VADER's potential to support patient experience management in digitally evolving healthcare settings, by transforming informal patient feedback into actionable insights, especially in healthcare settings with emerging digital capabilities. Its efficiency in processing free-text enables quick, meaningful overviews that support responsive decision-making.

In terms of research results, the research provides an illustration of how VADER sentiment analysis can support the extraction of meaningful insights from rich free-text data on patient experiences, offering a view of customer grievances. As illustrated by past studies,^{4–9} VADER human-centric design

enhances sentiment understanding, making it a potentially valuable tool in healthcare-related analytics, going beyond basic polarity detection, by detecting the intensity of emotional cues. This makes it particularly useful for applications and domains where interpretability and speed are crucial, as is the case of patient experience management and the growing need to craft an accurate *patient's voice* in healthcare settings.¹⁻³

Beyond reinforcing previous studies findings,^{2,3,13,15,16} it is worthwhile highlighting, in particular, three (3) concrete aspects provided by the present study that suggest the appropriateness of deepening and expanding the usage of VADER into the patient experience management space:

1. *Efficiency–effectiveness (accuracy) balance*: VADER's rule-based approach offers a quick and computationally efficient method for sentiment analysis without sacrificing accuracy. This is particularly beneficial in healthcare settings and patient experience handling where timely, easy-to-understand insights are of paramount importance.
2. *Support to patient-centric healthcare services*: by analyzing patient complaints, healthcare providers can gain a deeper understanding of patient experiences and identify areas needing improvement. This aligns with the broader trend of patient-centered care, which emphasizes the importance of considering patient feedback in healthcare service delivery.
3. *Process optimization support (eg, case prioritization)*: the sentiment analysis made in the study revealed that most complaints had a negative tone, pinpointing concrete areas where patients frequently encounter problems; this can help healthcare providers prioritize these issues and allocate resources more effectively.

Evaluation and Benchmarking

To further validate these findings, a subset of complaints was manually annotated by 2 independent reviewers. Interrater agreement was high (Cohen's $\kappa = 0.82$), indicating strong consistency. VADER achieved an *F1* score of 0.76, while BERT reached 0.84. These results align with published benchmarks in healthcare sentiment analysis, where rule-based models typically perform in the 0.70 to 0.75 range and transformer-based models often exceed 0.80 when fine-tuned. This reinforces the practical viability of sentiment analysis tools in patient experience management. Although VADER offers interpretability and speed, its rule-based nature may miss subtle emotional cues. BERT, by contrast, demonstrated greater sensitivity to context, particularly in ambiguous or sarcastic expressions. These findings suggest that hybrid approaches, combining rule-based and deep learning models, may offer improved accuracy and nuance in future applications.

Sentiment Complexity and Limitations

While VADER-based sentiment analysis may offer valuable, actionable insights to patient feedback handling, three (3) challenges and limitations need to be considered:

1. *Dealing with nuances and ambiguity*: Sentiment analysis models tend to struggle with ambiguous expressions where the sentiment is not clear-cut. This can lead to misclassification and reduced accuracy. In the present study, the existence of a significant number of neutral and positive scores in complaints free-text signals possible advantages in using a deeper integration of different models in future work, hybridizing traditional techniques with deep learning models. Developing model capabilities to understand nuanced opinions and subtle sentiments remains a significant challenge. Future research should explore multidimensional sentiment filters and toxicity classifiers to better detect emotional ambiguity and implicit negativity.
2. *Unstructured data and domain adaptation*: Models trained on specific datasets often perform poorly when applied to different domains. To this regard, in the present study, a significant challenge was dealing with unstructured data, often containing noise, slang, abbreviations, and inconsistent use of language. Effective preprocessing methods, such as tokenization, stemming, lemmatization, and noise removal, are essential to enhance the quality of the data and improve analysis outcomes.
3. *Contextual dependency (for interpretation)*: Understanding the context in which a sentiment is expressed is crucial for accurate analysis. In the present study, following the existing literature,^{17,18} the use of BERT with comparative purposes, highlighted the relevance of considering the surrounding context to correctly interpret sentiments, especially in complex sentences.

Additionally, the scope of this study is constrained by its dataset: 63 complaints from a single-surgery service within one (1) public hospital. While this focused sample allowed for detailed analysis and methodological control, it limits the generalizability of the findings. Patient sentiment may vary significantly across departments, institutions, and cultural contexts. Future research should expand to multisite, cross-departmental datasets to validate these insights and explore how sentiment patterns differ across healthcare environments. This broader approach would help ensure that sentiment-informed triage and analytics tools are scalable and adaptable to diverse healthcare systems.

Future Directions

Although direct metrics on downstream impact were not captured in this study, we propose a framework for sentiment-informed triaging. By flagging high-intensity

negative feedback, healthcare teams could prioritize cases more effectively and allocate resources with greater precision. Future pilot testing will evaluate whether this approach improves resolution times, satisfaction scores, and overall responsiveness. As healthcare systems evolve toward digital-first models, sentiment analysis offers scalable ways to amplify the patient's voice. Future work will explore hybrid models that combine rule-based methods for initial sentiment scoring with deep learning techniques for refinement. Integrating syntactic and semantic features with learned embeddings may further improve the detection of subtle, context-dependent sentiment cues in patient narratives.

For instance, sentiment-informed triage can be operationalized through automated dashboards that flag complaints with compound scores below -0.8 for immediate review. A complaint expressing severe dissatisfaction with postsurgical care, for example, could be routed directly to a patient experience officer or clinical lead. This prioritization mechanism can be embedded into digital workflows, enabling real-time alerts and escalation protocols. Over time, such systems can be calibrated to incorporate additional metadata (such as department, complaint type, or historical resolution time) to refine triage accuracy and improve service responsiveness. This approach aligns with broader efforts to integrate AI-driven insights into healthcare operations and enhance the timeliness of patient-centered interventions.

Conclusion

The present study illustrates the effectiveness and potential usefulness of using VADER sentiment analysis to analyze patient complaints in healthcare settings. Three key study findings suggest the advantages of using VADER to support healthcare patient experience management: (a) the efficiency–effectiveness (accuracy) balance offered by VADER, by delivering quick and computationally efficient sentiment analysis without sacrificing accuracy; (b) support given to patient-centered care; and (c) support given to process optimization intents (via, eg, sentiment data-infused case prioritization). Future work will focus on integrating advanced machine learning models to enhance the accuracy and depth of sentiment analysis, as demonstrated by the comparison with the BERT sentiment analysis model. Additionally, expanding the dataset to include cross-sectional feedback from multiple departments and hospitals will provide a broader and more comprehensive perspective on patient sentiments across healthcare systems, helping to identify recurring issues and prioritize areas for improvement.

Acknowledgments

The authors would like to acknowledge the healthcare organization included in the study for providing access to the data used in this research. No funding is associated with the research.

Author Contributions

João Vasco Coelho: conceptualization and writing—review and editing. Liliana da Costa Barbosa: writing—original draft, project management, research operationalization, and data curation.

Informed Consent

Patient consent was obtained but is not needed for the study accomplishment.

Data Availability

The data that has been used is confidential.

Declaration of Competing Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Ethical Approval

This study was approved by the Ethics Committee of healthcare organization included in the study on October, 2024. Email written consent by the Ethics Committee was obtained prior to the development of the study. This research was conducted ethically in accordance with the World Medical Association Declaration of Helsinki.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Supplemental Material

Supplemental material for this article is available online.

ORCID iDs

João Vasco Coelho  <https://orcid.org/0000-0002-7880-0679>

Liliana da Costa Barbosa  <https://orcid.org/0009-0008-1265-3056>

References

1. Borg A, Boldt M. Using VADER sentiment and SVM for predicting customer response sentiment. *Expert Syst Appl*. 2020;162. doi:10.1016/j.eswa.2020.113746
2. Hutto CJ, Gilbert E. VADER: a parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the 8th international conference on weblogs and social media (ICWSM), 2014. doi:10.1609/icwsm.v8i1.14550
3. Harrison R, Walton M, Healy J, Smith-Merry J, Hobbs C. Patient complaints about hospital services: applying a complaint taxonomy to analyse and respond to complaints. *Int J Qual Health Care*. 2016;28(2):240–45. doi:10.1093/intqhc/mzw003
4. Mann CD, Howes JA, Buchanan A, Bowrey DJ. One-year audit of complaints made against a university hospital surgical department. *ANZ J Surg*. 2012;82(10):671–74. doi: 10.1111/j.1445-2197.2012.06240.x
5. Donaldson L. An organisation with memory. *Clin Med (Lond)*. 2002;2(5):452–57. doi:10.7861/clinmedicine.2-5-452

6. Lloyd-Bostock S, Mulcahy L. The social psychology of making and responding to hospital complaints: an account model of complaint processes. *Law Policy*. 1994;16(2):123–47. doi:10.1111/j.1467-9930.1994.tb00120.x
7. Mulcahy L, Tritter JQ. Pathways, pyramids and icebergs? Mapping the links between dissatisfaction and complaints. *Sociol Health Illn*. 1998;20(6):825–47. doi:10.1111/1467-9566.00131
8. Carmel S. Patient complaint strategies in a general hospital. *Hosp Health Serv Adm*. 1990;35(2):277–88.
9. Reader TW, Gillespie A, Roberts J. Patient complaints in healthcare systems: a systematic review and coding taxonomy. 2014. doi:10.1136/bmjqs-2013-002437
10. Socher R, Perelygin A, Wu J, Chuang J, Manning C, Ng A, et al. Recursive deep models for semantic compositionality over a sentiment treebank; Conference on empirical methods in natural language processing; 2013 Accessed November 11, 2025. <https://api.semanticscholar.org/CorpusID:990233>.
11. Agarwal A, Xie B, Vovsha I, Rambow O, Passonneau R. Sentiment analysis of Twitter data. In: Proceedings of the workshop on languages in social media, 2011.
12. Wang H, Can D, Kazemzadeh A, Bar F, Narayanan S. A system for real-time Twitter sentiment analysis of 2012 US presidential election cycle. 2012.
13. Mao Y, Liu Q, Zhang Y. Sentiment analysis methods, applications, and challenges: a systematic literature review. *J King Saud Univ Comput Inf Sci*. 2024;36(4):102048. doi:10.1016/j.jksuci.2024.102048
14. Hoti M, Ajdari J. Sentiment analysis using the VADER model for assessing company services based on posts on social media. *SEEU Rev*. 2023;18(2):19–33. doi:10.2478/seur-2023-0043
15. Youvan D. Understanding sentiment analysis with VADER: a comprehensive overview and application. 2024. doi:10.13140/RG.2.2.33567.98726
16. Al Qarana G, Rianto L, Purnomo L. Analyzing perceptions of maternal and pediatric care in Jakarta: an integrated VADER and GloVe analysis of Google reviews in mother and child hospitals. Accessed November 11, 2025. www.joiv.org/index.php/joiv
17. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: NAACL HLT 2019—proceedings of the conference of the North American chapter of the association for computational linguistics: human language technologies, 2019.
18. Languré A, Zareei M. Improving text emotion detection through comprehensive dataset quality analysis. *IEEE Access*. 2024;12:166512–36. doi:10.1109/ACCESS.2024.3491856