Analysing Emotions in Cancer Narratives: A Corpus-Driven Approach

Daisy Monika Lal[1], Paul Rayson[1], Sheila Payne[2], Yufeng Liu[3]
[1]UCREL Research Centre, Computing and Communications, Lancaster University, UK
[2]Health Research, Lancaster University, UK
[3]Linguistics and English Language, Lancaster University, UK
{d.m.lal, p.rayson, s.a.payne, yufengliu}@lancaster.ac.uk

Abstract

Cancer not only affects a patient's physical health, but it can also elicit a wide spectrum of intense emotions in patients, friends, and family members. People with cancer and their carers (family member, partner, or friend) are increasingly turning to the web for information and support. Despite the expansion of sentiment analysis in the context of social media and healthcare, there is relatively less research on patient narratives, which are longer, more complex texts, and difficult to assess. In this exploratory work, we examine how patients and carers express their feelings about various aspects of cancer (treatments and stages). The objective of this paper is to illustrate with examples the nature of language in the clinical domain, as well as the complexities of language when performing automatic sentiment and emotion analysis. We perform a linguistic analysis of a corpus of cancer narratives collected from Reddit. We examine the performance of five state-of-the-art models (T5, DistilBERT, Roberta, RobertaGo, and NRCLex) to see how well they match with human comparisons separated by linguistic and medical background. The corpus yielded several surprising results that could be useful to sentiment analysis NLP experts. The linguistic issues encountered were classified into four categories: statements expressing a variety of emotions, ambiguous or conflicting statements with contradictory emotions, statements requiring additional context, and statements in which sentiment and emotions can be inferred but are not explicitly mentioned.

Keywords: Clinical narratives, medical language processing, social media, cancer, sentiment analysis, emotion analysis

1. Introduction

Cancer is one of the most prevalent diseases impacting the lives of millions of individuals and families worldwide. According to cancer statistics, 14.1 million people worldwide are affected by the disease (Torre et al., 2015). A cancer diagnosis can be upsetting and cause challenging psychological reactions in patients, including anxiety, despair, isolation, and feelings of shame and self-blame. Some individuals may experience heightened emotions contemplating the emotional impact of this news on their loved ones (Muzzin et al., 1994; Ahn et al., 2009; Singer et al., 2010; Cho et al., 2013; Al-Azri et al., 2014a,b). According to estimates, up to one-third of cancer patients receiving hospital treatment also suffer from a prevalent mental health issue (Singer et al., 2010). To provide effective therapy for cancer patients it is important to monitor their emotional state and we also aim to support cancer patients, their families, and healthcare providers to better understand their options. Analysing emotions and sentiments is one part of the evidence base to support patients' treatment and care choices, at each stage of disease and treatment. Experts feel that focusing on patients' emotions can improve their health, self-efficacy (patient engagement or involvement in improving the quality of healthcare) (Lacy, 2016; Marzban et al., 2022), and well-being, hence assessing their mood is an important part of their treatment (Ryan et al., 2005; Harvey and Lawson, 2009).

Social media platforms have become more and more prevalent in providing a common place for patients and their loved ones to express their experiences with cancer (Bender et al., 2011, 2013; Kent et al., 2016; Domínguez and Sapiña, 2017). As a result, social media data can be used to examine the way patients and carers (family member, partner, or friend) talk about their journeys. To facilitate this large-scale analysis, we can use NLP approaches like sentiment analysis (SA), which have evolved over time from fundamental concepts to powerful deep learning (DL) algorithms that are becoming a valuable tool for a variety of NLP applications. Reddit has a huge collection of forums covering news, discussion, entertainment, and just about any topic. Through a network of discussion boards known as subreddits, hundreds of millions of active users regularly share their unfiltered opinions, experiences, ideas, and feelings on a wide range of topics.

In this exploratory study, we analyse the moods and attitudes among cancer patients and their loved ones by acquiring posts from various cancer-related forums on Reddit. We conducted both qualitative

Proceedings of CL4Health 2024 @LREC-COLING-2024, pages 73–83

and quantitative assessments to explore cancer-related attitudes and emotions for different cancer stages (I, II, III, IV) and cancer treatments (diagnosis, clinical trials, chemotherapy, radiation therapy, targeted therapy, and palliative care)[1], as well as to see how closely the automated techniques match manual annotation. Moreover, we dive deeper into challenges associated with medical or clinical data processing.

2. Related Work

A wide range of modern techniques, including rule-based, conventional machine learning, and more advanced deep learning approaches, have been effectively applied to the task of SA in health and well-being (Zunic et al., 2020). Prior research, however, reveals that relatively few attempts have been made to use NLP to conduct a large scale examination of cancer patients' views during their journey, specifically the course of their treatment. The majority of research in this area focuses on patient experiences (positive, negative, and neutral) to enhance patient satisfaction (how satisfied patients are with a specific type of treatment as opposed to how they feel during that treatment).

Aspect-based Sentiment Analysis has previously been applied to the healthcare industry, most notably for SA of social media data on drug reviews (Gräßer et al., 2018; Sweidan et al., 2021), the COVID-19 disease and its vaccination (Aygün et al., 2021; Jang et al., 2021; Chaudhary et al., 2023), as well as psychological clinical records concerning suicide (George et al., 2021). Gräßer et al. (2018) performed several experiments related to drug reviews using data scraped from online drug review websites. The study involved looking at people's attitudes toward their overall experiences, side effects, and the usefulness of certain treatments. It also addressed the problem of the absence of annotated data and looked into the transferability of learned classification models across domains. (Sweidan et al., 2021) aimed to create a hybrid ontology-XLNet transfer learning strategy for identifying Adverse Drug Reactions (ADRs) from social data using sentence-level ABSA.

3. Methodology

This section provides an elaborate discussion of the strategy adopted for sentiment and emotion classification of the cancer-related Reddits.

Table 1: Cancer aspect-specific post-sets. Here, Min and Max represent the minimum and maximum word length of the posts.

Aspect Category	**Total Count**	**Min (words)**	**Max (words)**
Diagnosis	224	22	2396
Clinical Trials	232	30	3829
Surgery	230	11	2396
Chemotherapy	227	22	2396
Palliative Care	227	43	3554
Radiation Therapy	228	27	3829
Targeted Therapy	121	22	3829
Stage I	53	45	2396
Stage II	15	62	3829
Stage III	20	62	691
Stage IV	142	37	2396

3.1. Data Collection

We collected a sample of cancer-related English Reddit posts using PRAW[2](Python Reddit API Wrapper) from several subreddits including 'r/cancer', 'r/cancersurvivors', 'r/cancerfamilysupport', and 'r/cancercaregivers'. We acquired around 1,500 public posts using cancer-specific (treatment and stage) aspects. For each aspect, we created a separate post-set comprised of all posts containing the aspect term as shown in Tables 1. As people discussed their journeys in these narratives, several posts featured overlapping aspects, such as diagnosis, which was nearly always reported, multiple stages and treatments, chemotherapy and surgery occurring concurrently in many posts, and so on. Because most posts contain several terms, selecting those covering a single aspect, for example, "stage 1" or "palliative care" makes it challenging to generate an independent set with adequate data samples. This yields significantly fewer data points for each set. To resolve this conflict, we just used the aspect-term to search the subreddits, and posts that had multiple aspects were added to all sets.

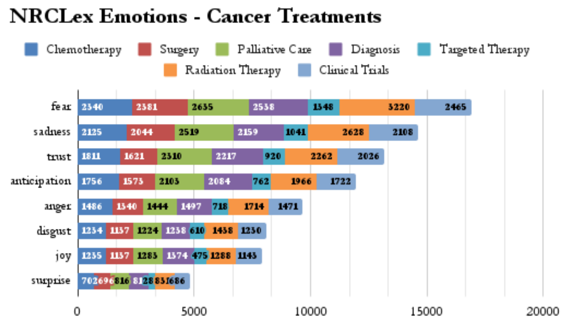
3.2. Sentiment and Emotion Classification Models

The idea is to analyse the sentiments or emotions keeping in mind the entire narrative (post). Since nearly all of the state-of-the-art algorithms have word count restrictions and because the entire narrative cannot be adequately tagged using a single sentiment or emotion, sentence-level classification was used to assess the attitudes and emotions in the post-sets. To examine the relative sentiments (positive, negative, or neutral) and emotions (sad-

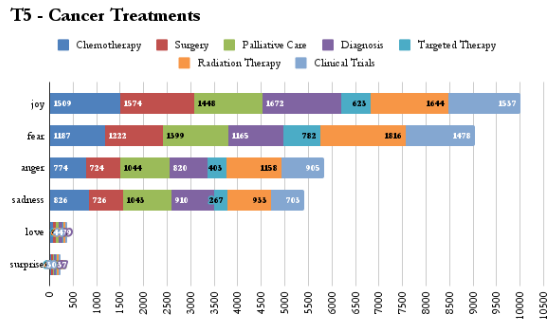
[1] *Data and Code* are available at `https://github.com/4dpicture/Emotion-Analysis`.

[2] *PRAW* available at `https://praw.readthedocs.io/en/latest/`.

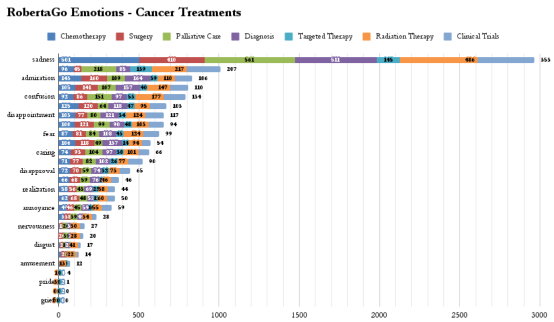
Figure 1: Cancer treatment-specific sentence-level emotion classification.



(a) Emotion classification using NRCLex. Emotion classes include fear, sadness, trust, anger, disgust, joy, surprise, and anticipation.



(b) Emotion classification using T5. Emotion categories include joy, fear, anger, sadness, love, and surprise.



(c) Emotion classification using RobertaGo Emotions with 27 emotion classes.

ness, anger, joy, surprise, etc.) from the gathered post-sets, five different models (T5, DistilBERT, Roberta, RobertaGo, and NRCLex) were utilized. The models are characterized as follows: ***T5*** [3]***:*** Google's T5 (Raffel et al., 2020) is an emotion detection model fine-tuned using the emotion recognition dataset introduced in (Saravia et al., 2018). It provides six emotion classes: sadness, joy, love, rage, fear, or surprise. ***DistilBERT*** [4] ***:*** DistilBert (Sanh et al., 2019) was built using the Stanford Sentiment Treebank (SST) (Socher et al., 2013) composed of 11,855 single sentences collected from movie reviews. It classifies text into two categories: positive and negative emotions. ***Roberta*** [5] ***:*** Roberta was fine-tuned for the SA task using TweetEval (Rosenthal et al., 2017), which contains roughly 124M tweets from January 2018 to December 2021. This provides three sentiment classes: positive, negative, and neutral. ***RobertaGo*** [6] ***:*** RobertaGo (Liu et al., 2019) is a multi-label classification model tweaked on the largest manually annotated dataset, Go-Emotions (Demszky et al., 2020) consisting of 58k English Reddit comments, labeled for 27 emotion categories. ***NRCLex*** [7] ***:*** NRCLexicon is a PyPI project designed to gauge ten emotion categories. It is created using the NRC emotion lexicon (Mohammad and Turney, 2013) and the WordNet synonym sets from the NLTK library. It provides eight basic emotions (anger, anticipation, disgust, fear, joy, sadness, surprise, trust) and two sentiments (negative, positive).

4. Key Findings

This section reports the findings of our experiments. All percentages for a specific post-set were calcu-

[3]"*t5-base-finetuned-emotion*" available at https://huggingface.co/mrm8488/t5-base-finetuned-emotion.

[4]"*distilbert-base-uncased-finetuned-sst-2-english*" available at https://huggingface.co/distilbert-base-uncased-finetuned-sst-2-english.

[5]"*twitter-roberta-base-sentiment-latest*" available at https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest.

[6]"*roberta-base-go-emotions*" available at https://huggingface.co/SamLowe/roberta-base-go_emotions.

[7]"*Lexicon source 2016 National Research Council Canada (NRC)*" available at http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm.

Table 2: Examples of statements in the cancer narratives tagged with counter-intuitive sentiment (positive) and emotions (joy, happiness, etc.) by the models.

"Going into town I will enjoy the delightful offerings from del taco and a Xanex. We dine like kings!" *"T5":"<pad>joy" | "RobertaGo":"{'label': 'joy', 'score': 0.8974}" | "Roberta":"positive = 0.9764" | "DistilBERT":"POSITIVE" | "NRCLex":"trust = 0.25, positive = 0.25, joy = 0.25, anticipation = 0.25"*
"I made him breakfast, and dinner and tried to keep a positive mindset and just do the happy stuff we always loved doing, jamming, joking, eating waffles." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'joy', 'score': 0.6846}" | "Roberta":"positive = 0.9417" | "DistilBERT":"POSITIVE" | "NRCLex":"positive = 0.5714"*
"I told her I loved her and held her hand while they ended life support." *"T5":"<pad>love" | "RobertaGo":"{'label': 'love', 'score': 0.9029}" | "Roberta":"positive = 0.7132" | "DistilBERT":"POSITIVE" | "NRCLex":" ""*
"I'm still battling side effects but I'm so relieved to be done! I get to come home to my beautiful wife and loving dog every time." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'joy', 'score': 0.7250}" | "Roberta":"positive = 0.9543" | "DistilBERT":"POSITIVE" | "NRCLex":"positive = 0.3333, joy = 0.3333"*
"But hearing the phrase, 'you no longer have cancer' makes me feel invincible again." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'joy', 'score': 0.3020}" | "Roberta":"positive = 0.7231" | "DistilBERT":"POSITIVE" | "NRCLex":"fear = 0.2857, negative = 0.2857"*

lated using the formula:

$$S^i_j/E^i_j = \left(\frac{N_{S^i_j/E^i_j}}{T_j}\right) \times 100 \tag{1}$$

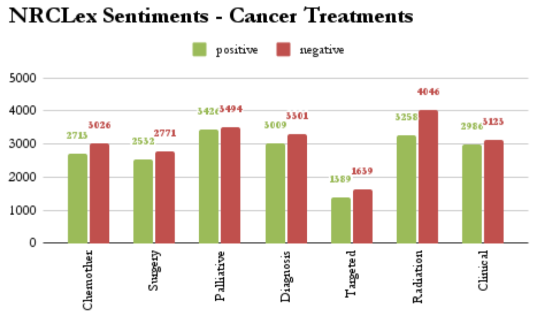
where, i refers to the i^{th} sentiment or emotion $\in$ {positive, negative, neutral, sadness, anger, joy, surprise, fear, etc.}, j refers to the j^{th} aspect $\in$ {stage 1, stage 2, ..., targeted therapy, surgery, and palliative care}, $N_{S^i_j/E^i_j}$ refers to the number of sentences tagged with sentiment or emotion i for aspect j, and T_j refers to the total number of sentences containing aspect j.

4.1. Cancer Treatments

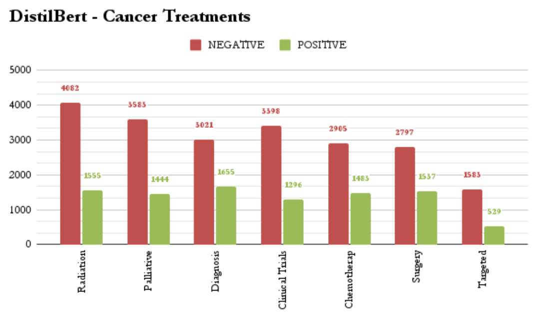
Sentiment Analysis: According to DistilBert (Figure 3b), negative sentiment is most typically relayed across all phases of treatment, in the order, diagnosis (64.6%), surgery (64.5%), chemotherapy (66.2%), palliative care (71.3%), clinical trials (72.4%), radiation therapy (72.4%), and targeted therapy (75.0%), indicating that patients have the greatest quantity of negative emotions during radiation therapy, followed by clinical trials and palliative care. Roberta (Figure 3c) detected more neutral feelings across multiple treatment aspects, clinical trials (45.0%), radiation therapy (44.2%), targeted therapy (51.8%), and palliative care (43.3%) while negative for the others. Across all phases, NRCLex (Figure 3a) detected more negative sentiments, diagnosis (52.3%), clinical trials (51.1%), chemotherapy (52.7%), surgery (52.3%), radiation therapy (55.4%), targeted therapy (54.1%), and palliative care (50.5%), with a narrow gap between the categories. Roberta, like DistilBert, has a wide disparity between the sentiment categories, with radiation therapy eliciting the most negative sentiment, followed by palliative care, diagnosis, clinical trials, and chemotherapy. In comparison to other aspects, Roberta indicated greater positive sentiments about diagnosis and surgery (see Table 2).

Emotion Analysis: According to the NRCLex model (Figure 2a), the two emotions that happen to surface most consistently throughout all treatment-specific aspects are *fear (18 - 22%)* and *sadness (15 - 17%)*. Other emotions include *trust (13 - 16%), anticipation (12 - 14%)*, and *anger (10 - 11%)*. Besides pointing out how emotions like *curiosity (2 - 6%), admiration (2 - 3%), approval (2 - 3%)*, and *confusion (2 - 3%)* are conveyed in the posts, RobertaGo (Figure 2c) ranks *sadness (6 - 11%)* as the most frequently relayed emotion throughout the phases. T5 (Figure 2b), in contrast to the two models, demonstrates that for all treatment elements, *joy (28 - 36%)* is the most prominent emotion, followed by *sadness (24 - 37%)*, and *anger (16 - 21%)*. Almost all emotion classifiers agree that *fear* and *sadness* are the most prevalent emotions. Every model also highlights *joy* and *happiness* as frequently seen emotions in the posts (see Table 2). Among the top five most expressed emotions, RobertaGo and T5 identify *joy* as one of the major feelings reported in the narratives. Apart from *joy*, RobertaGo mentions *gratitude*, *approval*, *caring*, and *admiration* as recurring emotions over the stages. T5 also reports *love* as an uncommon emotion encountered mostly during diagnosis and palliative care. NRCLex identifies *trust* and *anticipation* as significant, along with *fear*, *sadness*, and

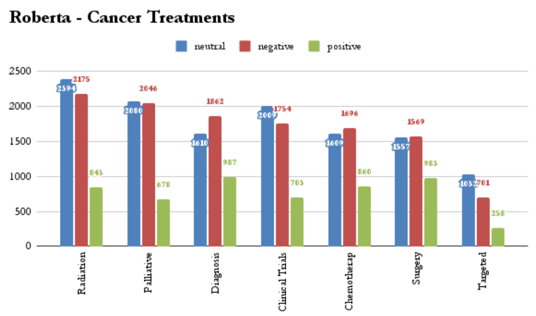
Figure 2: Cancer treatment-specific sentence-level sentiment classification.



(a) Sentiment classification using NRCLex. Sentiment categories include positive and negative.



(b) Sentiment classification using DistilBert. Sentiment categories include positive and negative.



(c) Sentiment classification using Roberta. Sentiment categories include negative, positive, and neutral.

anger.

4.2. Cancer Stages

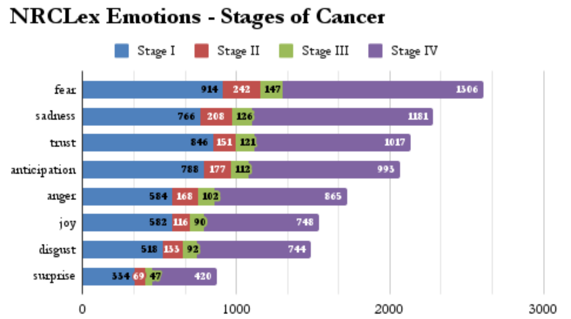
Sentiment Analysis: Negative sentiment is most commonly communicated across all cancer stages, particularly Stage IV cancer, as determined by all the classifiers. According to the DistilBert (Figure 5b), negative sentiment is expressed prominently across all cancer stages, Stage I (62.7%), Stage II (60.9%), Stage III (64.7%), and Stage IV (64.6%) where there exists a substantial difference between the sentiment categories. NRCLex (Figure 5a), on the other hand, exhibits a small gap between the sentiment categories throughout all stages and relays more positive attitudes for Stages I (52.2%) and III (53.3%) as compared to Stages II (46.3%) and IV (48.1%). At Stages III (37.7%) and IV (41.0%), Roberta (Figure 5c), expresses more negative sentiment, while at Stages I and II, it showcases a more neutral attitude.

Emotion Analysis: The two emotions that appear most consistently throughout all cancer stages, according to the NRCLex model (Figure 4a) are *fear (17 - 19%)* and *sadness (14 - 16%)*. RobertaGo (Figure 4c) highlights *sadness (15 - 24 %)* as the most frequently relayed emotion. In addition to that, it also emphasizes emotions like *admiration (8 - 10%)*, *optimism (5 - 9%)*, and *joy (5 - 6%)* being conveyed. According to NRCex and RobertaGo, sadness is mostly expressed in Stage IV cancer posts as compared to other stages. T5 (Figure 4b) contrasts the two models and reveals that for all cancer stages, *joy (33 - 39%)* is the emotion that is most displayed, followed by *fear (23 - 29%)*. The most prevalent emotions extracted by all classifiers are *fear* and *sadness*. Every model also points out that the feelings *joy* and *happiness* can frequently be seen in posts (see Table 2). *Joy* is one of the top five most frequently reported emotions by patients and carers, according to RobertaGo and T5. Apart from *joy*, RobertaGo mentions *gratitude*, *approval*, *caring*, and *admiration* as persistent emotions throughout the stages. NRCLex identifies *trust* and *anticipation* as significant, along with *fear*, *sadness*, and *anger*.

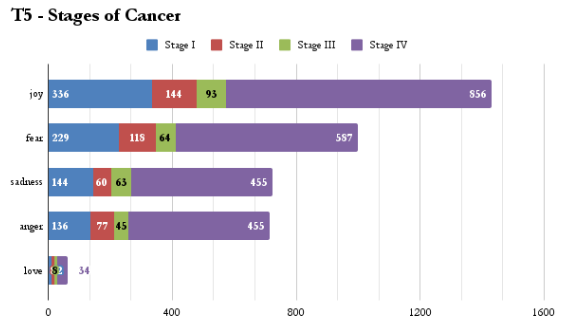
5. Human Evaluation

We reviewed a random subset of the collected posts having 50 sentences that were manually examined with the assistance of two distinct groups of annotators, three NLP/Linguistic researchers, and two medical domain specialists, to determine the impact of domain knowledge on the tasks for cancer-related data. The annotators evaluated the ***Sentiment Polarity***: To determine whether the language conveys a positive, negative, or neutral attitude, and ***Emotion Class***: To identify the emotion conveyed in the statement by using one of the following categories: sadness, anger, fear, joy, love, surprise, and neutral. Cohen's Kappa coefficient was used to determine the inter-rater agreement. Human evaluation was performed using three sentiment categories (positive, negative, and neutral) and seven emotion categories (sadness, anger, fear, joy, love, surprise, and neutral). The annotators were asked to choose one sentiment and emotion category they felt best suited in either scenario. Although some sentences contained

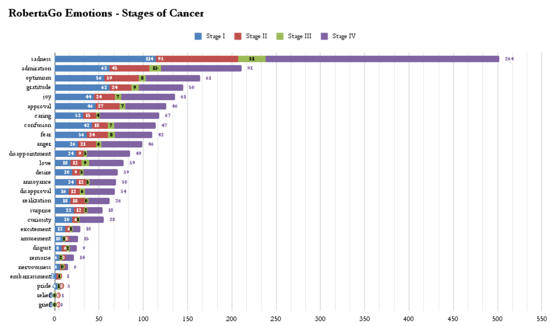
Figure 3: Cancer stage-specific sentence-level emotion classification.



(a) Emotion classification using NRCLex. Emotion classes include fear, sadness, trust, anger, disgust, joy, surprise, and anticipation.



(b) Emotion classification using T5. Emotion categories include joy, fear, anger, sadness, love, and surprise.

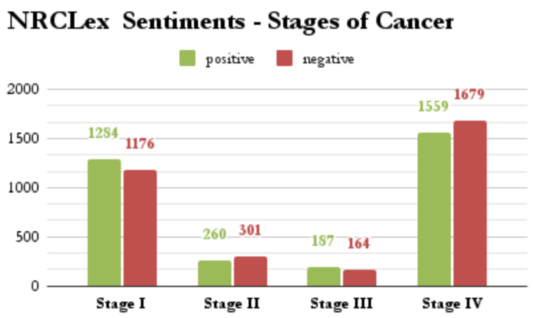


(c) Emotion classification using RobertaGo Emotions with 27 emotion classes.

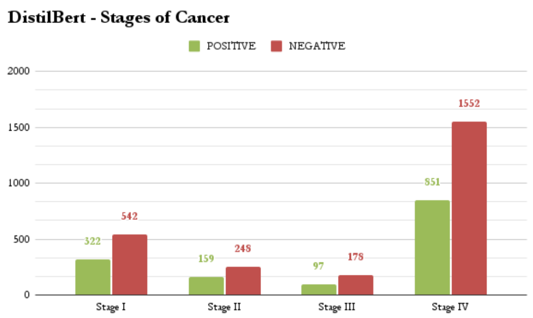
Table 3: Examples of statements in the clinical narratives tagged with contradicting sentiments and emotions by the SA models.

"I have been told I will never be cancer free, I have learnt to accept that." *| "T5":"<pad>joy" | "RobertaGo":"{'label': 'approval', 'score': 0.4907}" | "Roberta":"negative = 0.6521" | "DistilBERT":"POSITIVE" | "NRCLex":"fear = 0.2, anger = 0.2, negative = 0.2, sadness = 0.2, disgust = 0.2"*
"Finished a year of treatment and continued on maintenance chemotherapy for another year and was cancer free for a while until I relapsed at 22." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'neutral', 'score': 0.5965}" | "Roberta":"neutral = 0.6418" | "DistilBERT":"NEGATIVE" | "NRCLex":"fear = 0.17, anger = 0.17, trust = 0.17, negative = 0.17, sadness = 0.17, disgust = 0.17"*
"I'm going to die, and I'm going to do it with as much dignity as possible, and have the best last few months I can possibly have." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'optimism', 'score': 0.6297}" | "Roberta":"positive 0.8141" | "DistilBERT":"POSITIVE" | "NRCLex":"fear = 0.2, trust = 0.2, positive = 0.2, negative = 0.2, sadness = 0.2"*
"They were able to cut out the tumor but weren't successful in getting clean margins on the first pass." *"T5":"<pad>joy" | "RobertaGo":"{'label': 'neutral', 'score': 0.6801}" | "Roberta":"neutral = 0.5605" | "DistilBERT":"NEGATIVE" | "NRCLex":"trust = 0.2222, positive = 0.2222, joy = 0.2222"*
"This morning was my last day of radiation!" *"T5":"<pad>sadness" | "RobertaGo":"{'label': 'excitement', 'score': 0.4861}" | "Roberta":"positive = 0.5139" | "DistilBERT":"NEGATIVE" | "NRCLex":"fear = 0.5, negative = 0.5"*

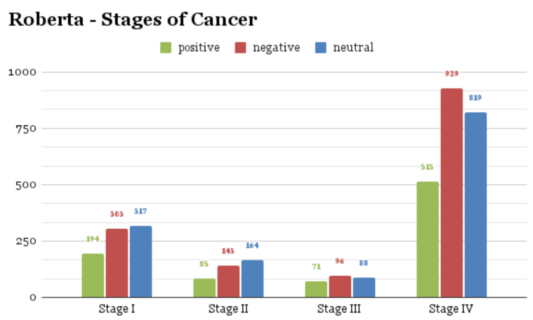
Figure 4: Cancer stage-specific sentence-level sentiment classification.



(a) Sentiment classification using NRCLex. Sentiment categories include positive and negative.



(b) Sentiment classification using DistilBert. Sentiment categories include positive and negative.



(c) Sentiment classification using Roberta. Sentiment categories include negative, positive, and neutral.

numerous emotion categories, the evaluation only permitted one choice for each category.

Corresponding to sentiment polarity (Figure 5), a substantial agreement (0.6383 - 0.6691) was discovered between the NLP/Linguistics researchers, and a substantial agreement (0.7534) exists among the domain experts. Moderate to substantial agreement (0.5064 - 0.6299) exists between the two groups indicating that scientific background and understanding somewhat influenced the sentiment polarity communicated in the sentences. Also, it was observed that the NLP/Linguistics researchers tended to classify a slightly larger share of sentences as neutral compared to the domain experts. Corresponding to the emotion analysis (Figure 6b), a fair to moderate agreement (0.2946 - 0.5037) was found among the NLP/Linguistics researchers, while a substantial agreement (0.7258) was observed among the domain experts. Between the two groups, there is fair to moderate agreement (0.2273 - 0.5811), indicating that scientific background and understanding influence the emotions communicated in the sentences.

6. Comparative Analysis

To assess how well the SA models performed on data related to cancer, we compared the human and model-assigned sentiment tags using precision, recall, and accuracy (Table 4).

Roberta: The model achieves performance scores ranging from 0.6 to 0.8, indicating its ability to capture a significant portion of sentiment tags as assigned by annotators. With identical recall, precision, and accuracy scores of 0.7, it demonstrates balanced performance in identifying instances of the target emotion class while minimizing false identifications. Consistency in scores suggests that the model consistently captures the same instances as human-assigned labels, indicating comparable performance to human annotators. This implies effective training of the machine learning model, which remains accurate and reliable in assigning labels. A low trade-off between recall and precision indicates efficient identification of relevant instances while minimizing false positives, reflecting a well-tuned model. This reduces the need for manual verification or correction of labels, leading to significant time and cost savings in large-scale labeling tasks. Furthermore, the performace metrics indicate that Roberta surpasses both NRCLex and DistilBert on the SA task.

NRCLex and DistilBert: Both models achieve average recall and accuracy scores of 0.52, indicating a moderate performance. The precision scores are notably high, with NRCLex averaging 0.76 and DistilBert scoring 0.83. This suggests that the models tend to be conservative in their predictions, preferring to refrain from labeling instances as belonging to the emotion class unless they are highly confident in their prediction. However, the low recall indicates that a significant portion of instances belonging to the emotion class is being missed. This selective behavior, characterized by high precision and low recall, suggests that the models prioritize precision over recall, opting to make fewer predictions but ensuring their

Table 4: Precision, recall, and accuracy between human annotations and machine-generated sentiment tags for the sampled evaluation set. Here, **A**, **P**, **R** denote accuracy, precision, and recall, respectively. **a1**, **a2**, and **a3** denote experts in NLP/Linguistics, while a3 and a4 represent domain experts. **avg** denotes the average scores w.r.t all annotators.

	Roberta			NRCLex			DistilBert		
	A	**P**	**R**	**A**	**P**	**R**	**A**	**P**	**R**
a1	0.78	0.79	0.78	0.46	0.75	0.46	0.48	0.84	0.48
a2	0.74	0.73	0.74	0.62	0.83	0.62	0.58	0.85	0.58
a3	0.72	0.72	0.72	0.51	0.76	0.51	0.6	0.88	0.6
a4	0.70	0.71	0.70	0.54	0.78	0.54	0.5	0.83	0.5
a5	0.64	0.64	0.64	0.46	0.70	0.46	0.46	0.72	0.46
avg	**0.72**	**0.72**	**0.72**	**0.52**	**0.76**	**0.52**	**0.52**	**0.83**	**0.52**

correctness. Consequently, despite the moderate performance indicated by the accuracy score of 0.52, there remains substantial potential for enhancing the models' ability to capture more relevant instances.

All models exhibit relatively higher recall, precision, and accuracy when assessed by NLP/Linguistics researchers (a1, a2, a3) compared to domain experts (a4, a5). The low performance with domain expert annotators suggests that the model may not effectively capture the nuances or specific characteristics of the domain it was trained on. This could be due to limitations in the training data, inadequate representation of domain-specific features, or biases in the model architecture.

7. Linguistic and Semantic Challenges

Challenges that surfaced during the classification of cancer narratives are stated below. In all the examples, the color red represents a negative sentiment, blue represents a positive sentiment, and green represents a neutral or a sometimes ambiguous expression. The sentiment and emotion analysis is based on the tags provided by model predictions and human annotation.

1. Statements expressing a range of emotions: When examining the posts, it appeared that people expressed a wide range of emotions in the same statement. For example, in the statement, "I have been told I will never be cancer free, I have learnt to accept that," the expression "never be cancer free" expresses a negative sentiment and a variety of emotions, sadness, fear, disappointment, etc. while "learnt to accept" displays a positive attitude and emotions like love, approval, optimism, admiration, etc. In another statement, "I'm going to die, and I'm going to do it with as much dignity as possible and have the best last few months I can possibly have," the patient expresses a negative sentiment (though, given the circumstances, they may see it positively) in "going to die" while expressing a positive attitude in "do it with as much dignity as possible." The phrase "best last few months" conveys approval, adoration, optimism, sadness, and grief all at once. It is challenging to gauge the overall feeling relayed in such texts. Another example includes, "The hospital I live in right now had given me less than 2 months, and I outlived it." One of the major challenges for automatic sentiment or emotion classification is the ability to identify the overall attitude and pick the most likely emotion when the text is capable of multiple interpretations based on the context.

2. Statements with contradicting emotions: When analysing the posts it was also observed that individuals often express their feelings using contradicting emotions (see Table 3). For example, in the statement, "This may sound like hell, but it's actually pretty peaceful," the patient expresses a negative attitude towards something comparable to an experience in hell when they use "may sound like hell," while contradicting the assumption and concluding that the experience is positive when they use "it's actually pretty peaceful". The statement exemplifies opposing feelings of disgust, fear, admiration, and optimism. This type of uncertainty is tough to capture and resolve not only for humans but also for machines. Furthermore, "I won't say that I hope my long sleep comes soon, but I don't fear it, it's almost time for me to sleep forever". In the preceding statement, the patient expresses a positive acceptance of a sad and undesirable circumstance. From the sentence "They were able to cut out the tumor but weren't successful in getting clean margins on the first pass," it is difficult to discern the sentiment because "able to cut out the tumor" indicates a positive sentiment or a sense of excitement, yet "weren't successful in getting clean margins" expresses a negative sentiment or sense of disappointment.

3. Statements that require additional context:

We discovered that sentiment and emotion analysis of cancer-related texts frequently necessitates prior knowledge and awareness of the healthcare domain, and that sentences cannot be successfully categorised without additional context. The sentence, "So I had my results today from my first Ct scan since treatment," is neutral in attitude and emotion because we do not know whether the results were favorable or not at this point. Some models selected joy as an acceptable emotion tag for this text, and two annotators suggested surprise. The second example, "I had an aggressive cancer in my left lung that spread to my lymph nodes quickly," lacks information regarding whether or not the situation was later improved. The NLP/Linguistic annotators identified a neutral sentiment and emotion for the sentence which might appear to be negative as identified by the domain experts who also chose fear and surprise as the appropriate emotion tag. Similarly the statement, “I'm having really complicated feelings about this” is ambiguous since we do not know what the individual is having "complicated feelings" about and, as a result, what kind of sentiment or emotion should be associated with it. It clearly portrays a sense of confusion, disapproval, disappointment, and nervousness, all at once.

4. Statements where sentiment and emotions can be inferred but are not explicitly mentioned: It has also been observed that, while emotions are not always explicitly mentioned in the sentence, they can be deduced using domain knowledge. For example, in the statement, “Finished a year of treatment and continued on maintenance chemotherapy for another year and was cancer-free for a while until I relapsed at 22,” the expression "until I relapsed at 22" can be construed as conveying a negative sentiment, but lacks any explicit emotional words to indicate that the person is afraid, sad, surprised, disappointed, etc. In these cases, we might hypothesise that human annotators might annotate a sentence using an inferred emotion, and dictionary-based NLP approaches would be less able to capture such emotions, whereas DL-based methods might detect such subtle clues to annotate an implied or inferred emotion.

8. Conclusion

In this study, we conduct sentiment and emotion analysis of Reddit forum data on aspects (specifically stages and treatments) unique to cancer. We intend to analyse spontaneous clinical narratives to better understand the wide range of emotions that a patient or carer experiences throughout the various stages of cancer or treatments from diagnosis to palliative care. Through this study we discovered that: 1) Besides negative emotions (fear, anger, and sadness), there are many (potentially unexpected) examples of positive emotions (joy, happiness, admiration, approval, and optimism) in cancer-related posts. 2) Human evaluation results further indicates the dependency of both tasks on domain knowledge and comprehension. 3) The precision, recall, and accuracy scores suggest difficulties in accurately capturing the nuances of the target domain. Addressing these challenges may necessitate domain adaptation, careful examination of biases in the training data, and potentially utilizing transfer learning techniques to enhance model performance across all domains. 4) Various challenges encountered in annotation, both manual and automatic, include statements expressing diverse emotions, ambiguity or inconsistency in statements with conflicting emotions, and statements requiring additional context. We believe linguists can gain useful insights from this study when manually annotating such narratives. Additionally, we think that NLP researchers conducting comparable studies or developing new models would benefit from the analysis of the NLP models.

Figure 5: Human evaluation of 50 instances of cancer-related Reddit data. The table shows the kappa reliability scores between annotators. Here, ai denotes the i^{th} annotator.

	a2	a3	a4	a5
a1	0.5468	0.6691	0.5166	0.6285
a2		0.5791	0.7534	0.6299
a3			0.5791	0.6383
a4				0.5064

(a) Sentiment polarity.

	a2	a3	a4	a5
a1	0.3181	0.2946	0.2273	0.3724
a2		0.5718	0.7258	0.537
a3			0.4878	0.5037
a4				0.5811

(b) Emotion classification.

9. Ethics Statement

The large-scale analysis of sentiment and emotions expressed in open or closed online forums, particularly related to sensitive topics such as cancer requires ethical approval, and we have been granted approval for secondary data analysis of previously analysed datasets. The research presented in this paper is part of a larger multilingual multinational research project, and each partner will apply it in their organization or country to replicate our analysis. The overall aim of the research is to improve the cancer patient journey and ensure personal preferences are understood and respected during treatment discussions with medical professionals, thereby supporting treatment and care choices, at each stage of disease or treatment.

10. Acknowledgements

This publication presents research from the 4D PICTURE project,[8] which is a collaboration of research teams from Austria, Belgium, Denmark, Germany, the Netherlands, Slovenia, Spain, Sweden and the UK. The research leading to these results has received funding from EU research and innovation programme HORIZON Europe 2021 under grant agreement 101057332 and by the Innovate UK Horizon Europe Guarantee Programme, UKRI Reference Number 10041120.

11. Bibliographical References

Eunmi Ahn, Juhee Cho, Dong Wook Shin, Byeong Woo Park, Sei Hyun Ahn, Dong-Young Noh, Seok Jin Nam, Eun Sook Lee, and Young Ho Yun. 2009. Impact of breast cancer diagnosis and treatment on work-related life and factors affecting them. *Breast cancer research and treatment*, 116:609–616.

Mohammed Al-Azri, Huda Al-Awisi, Samira Al-Rasbi, Kawther El-Shafie, Mustafa Al-Hinai, Hamdan Al-Habsi, and Mansour Al-Moundhri. 2014a. Psychosocial impact of breast cancer diagnosis among omani women. *Oman medical journal*, 29(6):437.

Mohammed H Al-Azri, Huda Al-Awisi, Samira Al-Rasbi, and Mansour Al-Moundhri. 2014b. Coping with a diagnosis of breast cancer among omani women. *Journal of health psychology*, 19(7):836–846.

Irfan Aygün, Buket Kaya, and Mehmet Kaya. 2021. Aspect based twitter sentiment analysis on vaccination and vaccine types in covid-19 pandemic with deep learning. *IEEE Journal of Biomedical and Health Informatics*, 26(5):2360–2369.

Jacqueline L Bender, M Carolina Jimenez-Marroquin, Lorraine E Ferris, Joel Katz, and Alejandro R Jadad. 2013. Online communities for breast cancer survivors: a review and analysis of their characteristics and levels of use. *Supportive Care in Cancer*, 21:1253–1263.

Jacqueline L Bender, Maria-Carolina Jimenez-Marroquin, and Alejandro R Jadad. 2011. Seeking support on facebook: a content analysis of breast cancer groups. *Journal of medical Internet research*, 13(1):e1560.

Meghna Chaudhary, Kristin Kosyluk, Sylvia Thomas, and Tempestt Neal. 2023. On the use of aspect-based sentiment analysis of twitter data to explore the experiences of african americans during covid-19. *Scientific Reports*, 13(1):10694.

Juhee Cho, Katherine Smith, Eun-Kyung Choi, Im-Ryung Kim, Yoon-Jung Chang, Hyun-Young Park, Eliseo Guallar, and Young Mog Shim. 2013. Public attitudes toward cancer and cancer patients: a national survey in korea. *Psycho-Oncology*, 22(3):605–613.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A Dataset of Fine-Grained Emotions. In *58th Annual Meeting of the Association for Computational Linguistics (ACL)*.

Martí Domínguez and Lucía Sapiña. 2017. “others like me”. an approach to the use of the internet and social networks in adolescents and young adults diagnosed with cancer. *Journal of Cancer Education*, 32:885–891.

Amy George, David Johnson, Giuseppe Carenini, Ali Eslami, Raymond Ng, and Elodie Portales-Casamar. 2021. Applications of aspect-based sentiment analysis on psychiatric clinical notes to study suicide in youth. *AMIA Summits on Translational Science Proceedings*, 2021:229.

Felix Gräßer, Surya Kallumadi, Hagen Malberg, and Sebastian Zaunseder. 2018. Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning. In *Proceedings of the 2018 international conference on digital health*, pages 121–125.

JN Harvey and VL Lawson. 2009. The importance of health belief models in determining self-care behaviour in diabetes. *Diabetic Medicine*, 26(1):5–13.

[8] https://4dpicture.eu/

Hyeju Jang, Emily Rempel, David Roth, Giuseppe Carenini, and Naveed Zafar Janjua. 2021. Tracking covid-19 discourse on twitter in north america: Infodemiology study using topic modeling and aspect-based sentiment analysis. *Journal of medical Internet research*, 23(2):e25431.

Erin E Kent, Julia H Rowland, Laurel Northouse, Kristin Litzelman, Wen-Ying Sylvia Chou, Nonniekaye Shelburne, Catherine Timura, Ann O'Mara, and Karen Huss. 2016. Caring for caregivers and patients: research and clinical priorities for informal cancer caregiving. *Cancer*, 122(13):1987–1995.

Amanda Lacy. 2016. Patient emotions and patient education technology. In *Emotions, Technology, and Health*, pages 23–39. Elsevier.

Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Sima Marzban, Marziye Najafi, Arjola Agolli, and Ensieh Ashrafi. 2022. Impact of patient engagement on healthcare quality: A scoping review. *Journal of Patient Experience*, 9:23743735221125439.

Saif M Mohammad and Peter D Turney. 2013. Nrc emotion lexicon. *National Research Council, Canada*, 2:234.

Linda J Muzzin, NJ Anderson, Alvaro T Figueredo, and Susan O Gudelis. 1994. The experience of cancer. *Social science & medicine*, 38(9):1201–1208.

Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *The Journal of Machine Learning Research*, 21(1):5485–5551.

Sara Rosenthal, Noura Farra, and Preslav Nakov. 2017. SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 502–518, Vancouver, Canada. Association for Computational Linguistics.

H Ryan, Penelope Schofield, Jill Cockburn, Phyllis Butow, Martin Tattersall, Jane Turner, Afaf Girgis, D Bandaranayake, and Deborah Bowman. 2005. How to recognize and manage psychological distress in cancer patients. *European journal of cancer care*, 14(1):7–15.

Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. Distilbert, a distilled version of bert: smaller, faster, cheaper and lighter. *arXiv preprint arXiv:1910.01108*.

Elvis Saravia, Hsien-Chi Toby Liu, Yen-Hao Huang, Junlin Wu, and Yi-Shin Chen. 2018. CARER: Contextualized affect representations for emotion recognition. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 3687–3697, Brussels, Belgium. Association for Computational Linguistics.

S Singer, J Das-Munshi, and E Brähler. 2010. Prevalence of mental health conditions in cancer patients in acute care—a meta-analysis. *Annals of oncology*, 21(5):925–930.

Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. 2013. Recursive deep models for semantic compositionality over a sentiment treebank. In *Proceedings of the 2013 conference on empirical methods in natural language processing*, pages 1631–1642.

Asmaa Hashem Sweidan, Nashwa El-Bendary, and Haytham Al-Feel. 2021. Sentence-level aspect-based sentiment analysis for classifying adverse drug reactions (adrs) using hybrid ontology-xlnet transfer learning. *IEEE Access*, 9:90828–90846.

Lindsey A Torre, Freddie Bray, Rebecca L Siegel, Jacques Ferlay, Joannie Lortet-Tieulent, and Ahmedin Jemal. 2015. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2):87–108.

Anastazia Zunic, Padraig Corcoran, and Irena Spasic. 2020. Sentiment analysis in health and wellbeing: systematic review. *JMIR medical informatics*, 8(1):e16023.