The Spectrum of Sentiment: Optimistic, Pessimistic, and Neutral Voices in Online Depression Discourse

Ștefana-Arina Tăbușcă[1], **Ana-Maria Bucur**[2], **Liviu P. Dinu**[3,4]
[1]Interdisciplinary School of Doctoral Studies, [3]Faculty of Mathematics and Computer Science
[4]Human Language Technologies Research Center, University of Bucharest, Romania
[2]Università della Svizzera italiana, Switzerland
stefana.tabusca@s.unibuc.ro

Abstract

The relationship between depression and the concepts of optimism and pessimism has been extensively researched by psychologists. In this paper, we use computational approaches to study how optimism and pessimism are expressed in the online discourse of people with a depression diagnosis. Publicly available datasets are used for the development of an optimism/pessimism detection model, as well as for the analyses performed on social media posts of individuals with depression, as measured by BDI-II, a validated depression questionnaire. To analyze the optimistic and pessimistic posts by individuals with depression, we use LIWC features and perform topic modeling. We also investigate specific words driving mislabeling using SHAP. Our results show that while there may not be significant differences in the number of optimistic versus pessimistic posts between individuals in the depression and control groups, the content of the posts differs meaningfully, both in terms of linguistic features and approached topics.

Keywords: optimism and pessimism, social media analysis, mental disorders

1. Introduction

Depression is one of the most prevalent mental disorders and has been extensively researched (Lim et al., 2018; Xu et al., 2021). Many studies focus on understanding how depression manifests and its relationship with mood and emotions (Rottenberg, 2005). In addition, previous research has also investigated the connection between depression and the concepts of optimism and pessimism. Karhu et al. (2024) demonstrate a bidirectional relationship: optimism not only buffers against depressive symptoms but is also eroded by them, while pessimism both predicts and is intensified by depression. Complementary studies by Korn et al. (2014) and Hobbs et al. (2022) reveal that, unlike healthy individuals who display an optimistic bias when updating beliefs about the future, those with depression tend to weigh negative information more heavily. Furthermore, optimism is linked to improved psychological well-being and more effective coping (Scheier et al., 2001), as well as better treatment outcomes, including reduced rehospitalization rates (Tindle et al., 2012). Prior research also highlights a reduced risk of work disability and an enhanced likelihood of returning to work following a depression-related disability (Kronström et al., 2011). The literature shows that, while optimism is not itself a coping strategy, it influences the coping strategies people employ: optimistic individuals are more likely to use approach-based coping methods, such as problem-solving and seeking social support, while they tend to avoid avoidance-based strategies (Nes and Segerstrom, 2006).

Depression detection is a prominent topic in NLP; modern methods use attention, deep learning, and pre-trained models (De Santana Correia and Colombini, 2022), demonstrating significant performance improvements over traditional approaches. However, in addition to identifying mental health disorders, language can offer insights into broader psychological states, such as optimism and pessimism, which are often associated with conditions like depression (Herwig et al., 2009). Previous NLP research has explored the manifestations of emotions (Uban et al., 2021; Aragon et al., 2021) and even happy moments using social media data from individuals with depression (Bucur et al., 2024). Although research has focused on developing more effective models for detecting optimism and pessimism (Ruan et al., 2016; Caragea et al., 2018; Alshahrani et al., 2021), to our knowledge, no analysis of optimism and pessimism in the social media language used by individuals with depression has been conducted. We consider this task to have the potential to provide valuable insights into how users express affective states online, which may, in turn, contribute to a better understanding of linguistic markers associated with mental health conditions. Thus, we aim to answer the following research questions:

RQ1: In what proportions are optimism and pessimism manifested in the discourse of individuals with depression?

RQ2: How is optimism manifested in the so-

Proceedings of the Fifteenth Language Resources and Evaluation Conference (LREC 2026), pages 10970–10981
11–16 May 2026.

cial media language of individuals with depression?

Quantifying optimism and pessimism in depressive discourse (RQ1) challenges the traditional view that depression is solely characterized by negative affect. Analyzing social media language for manifestations of optimism (RQ2) reveals subtle linguistic cues that conventional assessments may overlook.

In this paper, we make the following contributions:

1. Fine-tune a RoBERTa-based model for the detection of optimistic, pessimistic, and neutral posts.
2. Create a manually annotated evaluation subset of the eRisk depression dataset specifically for optimism and pessimism.
3. Perform psycholinguistic, statistical, and topical quantitative and qualitative analyses with the use of a set of statistical significance tests, LIWC, and Topic Modeling via BERTopic.
4. Investigate and interpret the lexical features that incorrectly influence model predictions using SHAP.

2. Related Work

Although still in its early stages, research on detecting optimism and pessimism in social media is expanding. A deep-learning technique developed by Blanco and Lourenço (2022) was used to examine the expression of optimistic and pessimistic sentiments in COVID-19-related Twitter conversations. They examined several network configurations and found that bi-LSTM systems produced the most successful models. According to the study, optimistic interactions tended to stay positive, whereas conversations with strong pessimistic signals showed little emotional change.

To improve prediction accuracy for optimism and pessimism, Alshahrani et al. (2020) employed XLNet, a network that combines several autoregressive language models, to capture semantic relationships and negations. On the benchmark dataset OPT (Ruan et al., 2016), the method proposed by the authors achieved a 63.32% reduction in error, increasing the state-of-the-art accuracy (across two defined setups and thresholds) from 90.32% to 96.45%.

Cobeli et al. (2022) introduced a Multi-Task Knowledge Distillation architecture, achieving 86.60% accuracy on the OPT dataset. The research found that certain POS tags, such as nouns, are consistently prevalent throughout all optimism ranges. Other tags, like hashtags, are associated with higher optimism levels. The use of emoticons, punctuation, and user remarks also influenced optimism. As tweets became more positive, first-person singular pronouns were less frequent, suggesting a connection between pessimism and depression. The architecture outperformed earlier setups for the 1/-1 threshold definition of optimism.

The concept of computational analyses in the field of mental health detection correlations in social media speech has been investigated to an extent in the study by Bucur et al. (2021), which looks into the relationship between offensive language and depression by examining how people with depression use offensive speech in their social media posts. According to the authors' results, there is a greater prevalence of derogatory language in the online speech of individuals who have been diagnosed with depression.

In our research, we use computational methods to analyze the online discourse of individuals with depression. We aim to explore the impact of optimism and pessimism, motivated by existing psychological research and advancements in NLP models designed to detect these two mental attitudes.

3. Data

We use two datasets in our experiments: the OPT dataset (Ruan et al., 2016), which includes annotations for optimism and pessimism, and the eRisk 2021 dataset (Parapar et al., 2021), which features social media users with depression.

The most popular dataset for optimism/pessimism identification was introduced by Ruan et al. (2016). It contains 7,475 randomly selected tweets from 500 individuals who were considered pessimistic and 500 who were considered optimists. To select the texts, tweets containing keywords related to optimism or pessimism were identified, highlighting both optimistic and pessimistic users. Each tweet was evaluated and classified by five human annotators using Amazon Mechanical Turk. To ensure accuracy, quality control procedures were implemented. Human annotators rated tweets on a disposition scale ranging from 3 (extremely optimistic) to -3 (very pessimistic); this scale enabled fine-grained distinctions among tweets, allowing different levels of optimism and pessimism to be identified within the text. The average of all the evaluations for the acquired annotations is the final score. The procedure resulted in a moderate final inter-annotator agreement (Krippendorff's alpha of 0.731).

In our experiments, we consider three classes: posts with an average annotation below -1 are labeled as pessimistic, those with a score of -1 to 1 are labeled as neutral, and the remaining posts are labeled as optimistic. This three-class setting

provides greater granularity and intuitiveness.

Our approach is different from the direction taken in the studies mentioned in Section 2. Both of those studies analyze posts with average scores ranging from -1 to 1 separately, as these scores are the most ambiguous in the given context, even for human interpretation. In one of their approaches, Cobeli et al. (2022) choose to eliminate this specific group of posts, and consider only two classes, optimistic and pessimistic. Alshahrani et al. (2020) employed the same method of ignoring the respective posts to address the ambiguity, calling it the -1/1 threshold. In both studies, this approach significantly improved model performance; however, for our work, we chose to keep the ambiguous data and create an additional class for it, for two main reasons:

1. We believe retaining this data helps preserve the complexity and authenticity of real-world social media posts, as realistically, not all posts are, and should not be, classified as either optimistic or pessimistic.

2. Eliminating the respective posts would mean reducing the data to almost half of the original size (3,847).

The eRisk 2021 dataset related to depression (Losada and Crestani, 2016; Parapar et al., 2021) contains social media users who were asked to fill in the BDI-II questionnaire (Beck et al., 1996) for the assessment of their depression status. Following this, their Reddit social media data was collected with their consent. The BDI-II questionnaire contains 21 questions related to depression symptoms, and the answers are used to calculate an overall score that indicates the level of depression. The training dataset comprises 90 users with ground-truth BDI-II scores and 46,502 posts from Reddit. The test dataset contains 80 users with a total of 32,237 posts. In our experiments, we use the data from all 170 users in the eRisk dataset. Because BDI-II is used by mental health professionals to diagnose depression, we consider users with a score above the established cut-off of 19 (Subica et al., 2014; von Glischinski et al., 2019) as having depression, while those with scores below this threshold are considered control users.

4. Methodology

4.1. Detection of optimism and pessimism

Following established methodologies in NLP(Liu et al., 2019; Guo et al., 2022; Amin et al., 2023), we use a RoBERTa-based model fine-tuned on the OPT dataset in our experiments, rather than focusing on incremental architectural improvements in this work. The fine-tuned RoBERTa-based model is used to predict optimism, pessimism, and neutral labels within the eRisk depression dataset, thereby addressing our research questions.

The model, which we will refer to as RoBERTa-OPT-3Labels from now on, was trained using the HuggingFace platform, with twitter-roberta-base-sentiment-latest serving as the base model (Camacho-Collados et al., 2022). The base model was refined for sentiment analysis using the TweetEval benchmark (Barbieri et al., 2020) after being trained on about 124 million tweets. In our fine-tuning, we set a learning rate of 5e-5, three epochs, a maximum sequence length of 128 tokens, an 8-batch size, and a warmup ratio of 0.1. To reduce overfitting, the AdamW optimizer was employed. The learning rate was decreased using the "linear" learning rate scheduler. To avoid the exploding gradient problem, the maximum gradient norm was set to 1. To guarantee consistency of outcomes, the seed was set to 42. If, after five successive evaluations, there was no progress in the validation metric, early stopping was employed with patience set to 5 and the threshold set at 0.01. We split the available data into 70% for training, 15% for validation, and 15% for testing.

4.2. Manually Annotated Evaluation Subset

To ensure the model reliably detects optimism and pessimism in Reddit posts, we are including a performance evaluation using a manually labeled sample. In this sense, we have randomly selected a total of 150 posts, divided equally among all possible subgroups: optimistic/pessimistic/neutral posts from individuals with depression, as well as optimistic/pessimistic/neutral posts from the control group. The texts were then manually rated by three human annotators, following the procedure outlined in Ruan et al. (2016), resulting in an average score within the -3/3 interval. The obtained score is then mapped according to the three available classes, the same as the original OPT data. These labels are used to validate the results obtained by our optimism/pessimism detection model, which are presented in a later section (Section 5.1).

We have also calculated inter-annotator agreement using Krippendorf's alpha, as was the case for the annotation process in the original OPT dataset. We obtained a value of 0.818, indicating high agreement among raters, which confirms the reliability of the process.

In terms of annotator characteristics, our collaborators are three PhD students, two female and one male, all with a minimum C1 English language proficiency level. They were instructed to assess the expressed outlook or expectation conveyed in

Optimism		Pessimism		Neutral	
Control	**Depression**	**Control**	**Depression**	**Control**	**Depression**
I'm happy that everything turned out rather well for you in the end, and that gives me a lot of hope for my future.	I graduated [...] and got my driver's license! [...] I know what the next goal to work for is. [...] I honestly value my friendships more.	It is sad to think that the life that we will live in is set for imminent destruction.	Something must always [...] remind me how painful life is and that it will never GENUINELY get better. [...] Everyone would be better off without me [...] I will never be good enough.	Beagles are usually listed as a breed that tends to get along well with cats [...]	I only consume great, but lesser-known media. Are you familiar with Steins;Gate and Morrowind? Thought so.

Table 1: Selected examples that were predicted as optimistic, pessimistic, or neutral from the depression and control groups.

each post, rather than the author's general mood or writing style. The task was to rate every post with a single integer score, disregarding linguistic form, unless it significantly altered the sentiment. Ambiguous or mixed cases were judged based on the annotator's overall impression, defaulting to 0 (neutral) when sentiment could not be reliably inferred. The identities of the annotators remained anonymous, as only the ratings were of interest in the study.

4.3. LIWC

LIWC 22 (Boyd et al., 2022) is an advanced text analysis tool that categorizes language into different dimensions, including psychologically meaningful ones, enabling the detection of cognitive, emotional, and social cues within the written content. In our study, we focus on the most context-significant LIWC-derived features to analyze optimistic and pessimistic posts by individuals with depression and control users. We quantify these differences using z-scores derived from the Mann–Whitney U test, a nonparametric statistical method that assesses whether one group systematically ranks higher or lower than another on a given variable, being particularly suited for analyzing linguistic features that may not follow a normal distribution. Specifically, we use the test to compare how the linguistic features (as categorized by LIWC) differ between the optimistic and pessimistic posts within the depression and control groups. The z-scores reflect the magnitude of these differences, enabling us to quantify the strength of association between specific language patterns (such as references to future focus, negative emotions, or social behavior) and either optimistic or pessimistic contexts within each group. We also apply the Benjamini–Hochberg procedure at a nominal alpha=0.05, which orders the p-values and computes adaptive thresholds to control the expected proportion of false discoveries. Features with FDR-adjusted p-values below 0.05 were deemed significant, ensuring that our inferences maintain high sensitivity to true effects while limiting the rate of false positives across all tests.

4.4. Topic Modeling

We implemented a topic modeling framework using BERTopic (Grootendorst, 2022) to uncover themes within social media posts, and to explore their associations with sentiment and mental health indicators. Our approach leveraged a BERTopic pipeline, which integrates text representation, dimensionality reduction, and clustering techniques.

First, we generated dense text embeddings with SentenceTransformer ('all-MiniLM-L6-v2') and reduced dimensionality using UMAP, opting for a reduced n_neighbors from 15 to 10, while preserving intrinsic data structure. Clustering was achieved with HDBSCAN, with min_cluster_size increased from 10 to 80 (for more robust topic clusters), following text preprocessing with a CountVectorizer configured for bi-grams that included the standard English stopwords, extended with common internet noise words: 'http', 'https', 'amp', 'com', 'www', 'r/'.

To enhance interpretability, topics were refined using a custom representation that leverages KeyBERT (Grootendorst, 2020), combined with Part-of-Speech filtering (via SpaCy's "en_core_web_sm") and Maximal Marginal Relevance (MMR), yielding high-quality, contextually relevant keywords. The final model assigned topics to each post, which were aggregated by sentiment (optimism, neutral, pessimism) and by depression status (depression vs. control groups). Chi-squared tests of independence were then employed to statistically assess differences in topic distributions across the target groups.

4.5. SHAP Error Analysis

As an error analysis of the mislabeled posts from the gold validation set, we used SHapley Additive Explanations (SHAP) (Lundberg and Lee, 2017), to investigate the lexical features erroneously driving model predictions for optimism, pessimism, and

Model	Val. Acc.	Test Acc.
Naïve Bayes	63.42	62.12
SVM	67.52	65.24
CNN	64.67	65.59
xlnet-base-cased	70.56	70.23
distilbert-base-uncased	71.27	71.03
RoBERTaOPT3Labels	71.54	71.65

Table 2: Baselines and comparison with RoBERTa-OPT-3Labels.

neutral sentiment. We applied SHAP to posts that were misclassified by RoBERTaOPT3Labels on the created gold validation set to examine patterns of confusion and contextual ambiguity. The approach quantified each token's contribution to the model's output, allowing us to identify words that increased confidence in a given class. The resulting token-level explanations allowed for aggregated visualizations (word clouds), which provided interpretable insight into how classification was influenced.

5. Results and Discussion

5.1. Model Performance

We have performed experiments with various other models, including Naïve Bayes, SVM, CNN, DistilBERT, and XLNet. We used GridSearchCV for hyperparameter tuning for both the Naïve Bayes model and the SVM classifier. The CNN model employs a sequential architecture with embedding, convolutional, global max pooling, and dense layers, along with dropout for regularization. It is trained over ten epochs to prevent overfitting, using early stopping. The additional two Transformer-based models, DistilBERT (Sanh et al., 2019), and XLNet (Yang et al., 2019), more specifically distilbert-base-uncased and xlnet-base-cased, respectively, were trained with the same setup as our selected model, RoBERTaOPT3Labels, described in Section 4.1. We chose the latter, as it was the best model based on our experiments. The reported results can be seen in Table 2.

The RoBERTa-OPT-3Labels model shows consistent and competitive performance, with an accuracy of 71.65%, and a weighted F1 score of 71.23%. The weighted AUC of 84.52% further underlines its ability to effectively distinguish among the three classes. As this represents the first study to adopt a three-class approach, to the best of our knowledge, it would be interesting to see the results of the state-of-the-art models that interpreted the 1/-1 scenario by eliminating the neutral/ambiguous posts (Caragea et al., 2018; Alshahrani et al., 2020, 2021; Cobeli et al., 2022). We present selected predicted samples in Table 1.

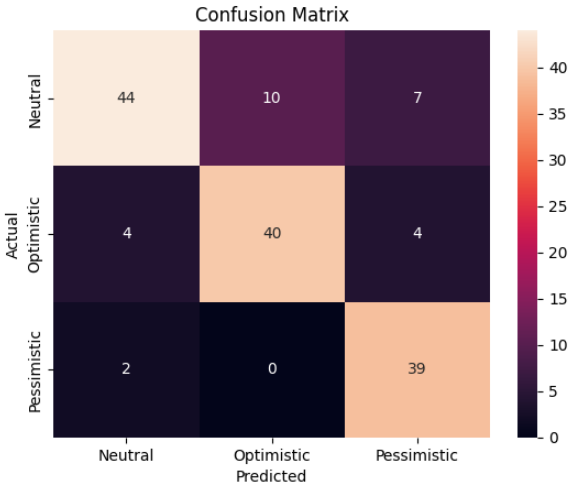

Figure 1: Confusion Matrix for results predicted by RoBERTa-OPT-3Labels on the gold validation data.

Our classifier was tested on the constructed gold validation set described in Section 3, achieving an overall accuracy of 82%, which demonstrates reliable alignment with human annotations. Class-specific F1-scores were uniformly high - 0.79 for neutral, 0.82 for optimistic, and 0.86 for pessimistic - indicating balanced performance across categories. These results can also be seen in the confusion matrix shown in Figure 1. The outcomes support RoBERTa-OPT-3Labels's suitability for automated sentiment analysis.

5.2. General Statistics Interpretation

Examining the predictions from the RoBERTa-OPT-3Labels model, we observe that users in the depression group have, on average, fewer optimistic posts than the control group, but a similar number of pessimistic posts. In addition, users in the control group have more posts labeled as neutral.

To test for statistical significance, we compare the number of optimistic, pessimistic, and neutral posts between the two groups using the Mann-Whitney U test, Cohen's d, and Pearson correlation (Table 3). The Mann–Whitney U test yields non-significant z-scores and p-values for both optimistic (-1.23, p = 0.22) and pessimistic (-0.20, p = 0.84) posts, suggesting that both groups produce similar amounts of content in these categories. In addition, the small effect sizes (Cohen's d = -0.18 for optimism, 0.06 for pessimism) and weak Pearson correlations further support this lack of meaningful distinction.

However, a more significant difference can be seen in the number of neutral posts for the performed tests, with a small to moderate effect size (d = -0.35). This suggests that individuals with depres-

	Mann–Whitney U test (z, p)	Cohen's d	Pearson Corr. (r, p)	Spearman Corr. (ρ, p)
Opt.	(-1.23, 0.22)	-0.18	(-0.09, 0.26)	(-0.10, 0.17)
Pess.	(-0.20, 0.84)	0.06	(0.03, 0.68)	(0.03, 0.70)
Neutral	(-2.21, 0.03)	-0.35	(-0.17, 0.03)	(-0.26, 0.0006)

Table 3: Statistical Test Results for Optimism and Pessimism.

sion post significantly fewer neutral statements than people not diagnosed with depression, potentially reflecting a tendency to engage more with emotionally valenced (optimistic or pessimistic) language rather than neutral discourse (Broome et al., 2015). An additional test that points towards the same conclusion is the analysis of the correlation between users' BDI-II scores and their respective proportions of optimistic/pessimistic/neutral posts. For this experiment, we used Spearman's rank correlations. A significant negative correlation emerged between BDI-II scores and the proportion of neutral posts (ρ = -0.26, p < 0.001), indicating that individuals with higher depressive symptoms produced fewer neutral statements. Correlations with optimistic and pessimistic content were not significant.

While the statistical tests indicate no significant differences in the number of optimistic or pessimistic posts between depression and control individuals, our subsequent analyses will demonstrate that the content of these posts may vary substantially. We will proceed to show that the way optimism and pessimism are expressed in language differs between users in the depression and control groups in a meaningful way.

5.3. LIWC Analysis Results

Figure 2 presents a side-by-side comparison of statistically significant (p<0.05, as measured by the Mann-Whitney U test) LIWC feature usage across optimistic (left panel) and pessimistic (right panel) posts by individuals with and without depression, measured via z-scores. The categories marked by (*) are significant according to the FDR-adjusted p-values. By analyzing these scores, we have outlined several key patterns:

In optimistic posts made by individuals with depression, the increased use of assent and impersonal pronouns suggests a more detached or externally directed expression of optimism. This observation aligns with research showing that individuals with depression often display reduced self-focus in positive contexts. For instance, they tend to use fewer first-person pronouns when recalling positive memories, which suggests that they have difficulty integrating positive experiences into the self-concept (Himmelstein et al., 2018). In contrast, optimistic posts from control individuals are characterized by greater use of the "family" category, suggesting that their expressions of optimism are more socially anchored and relational. This may reflect an integration of social connectedness and support into positive emotional experiences.

When looking at pessimistic posts, the gap between users in the depression and control groups is pronounced. Individuals with depression exhibit a significant increase in words related to general negative emotion and tone, suggesting a tendency toward more negative and critical thought processes (Mor and Winquist, 2002). The elevated authenticity score suggests that these expressions of pessimism are likely perceived as more honest and self-revealing. Additionally, greater use of cognitive processing and general cognition words may reflect an effort to make sense of negative experiences, a pattern common in depressive perception. Individuals with depression also exhibit a higher frequency of adverbs and a generally higher linguistic score, indicating increased verbal complexity, which may signal ruminative thought patterns. Control users who also express negative content in pessimistic posts tend to do so with fewer markers of pervasive distress, and the scope of their pessimism appears to focus more on external, leisure-related topics.

The statistical differences measured with the Mann-Whitney U test and supported by the FDR validation reveal how individuals with depression use language differently, first compared with the control group and secondly depending on whether the content is optimistic or pessimistic.

5.4. Topic Modeling Results

The chi-squared results across the target groups (depression versus control) reveal significant differences in how individuals communicate optimism, pessimism, and neutrality.

We will present results for six distinct subgroups, categorized by the depression label and the associated optimism/pessimism/neutral attitudes, with visualizations provided in Figure 3 as a heatmap. We present in Table 4 the most overrepresented and underrepresented topics for each target group: pessimism in the depression group concentrates around self-referential or distressing themes (mental health, politics, school), while optimism in the control group revolves around cognitively or socially detached topics (language, fiction, gaming).

The disparities suggest that psychological states

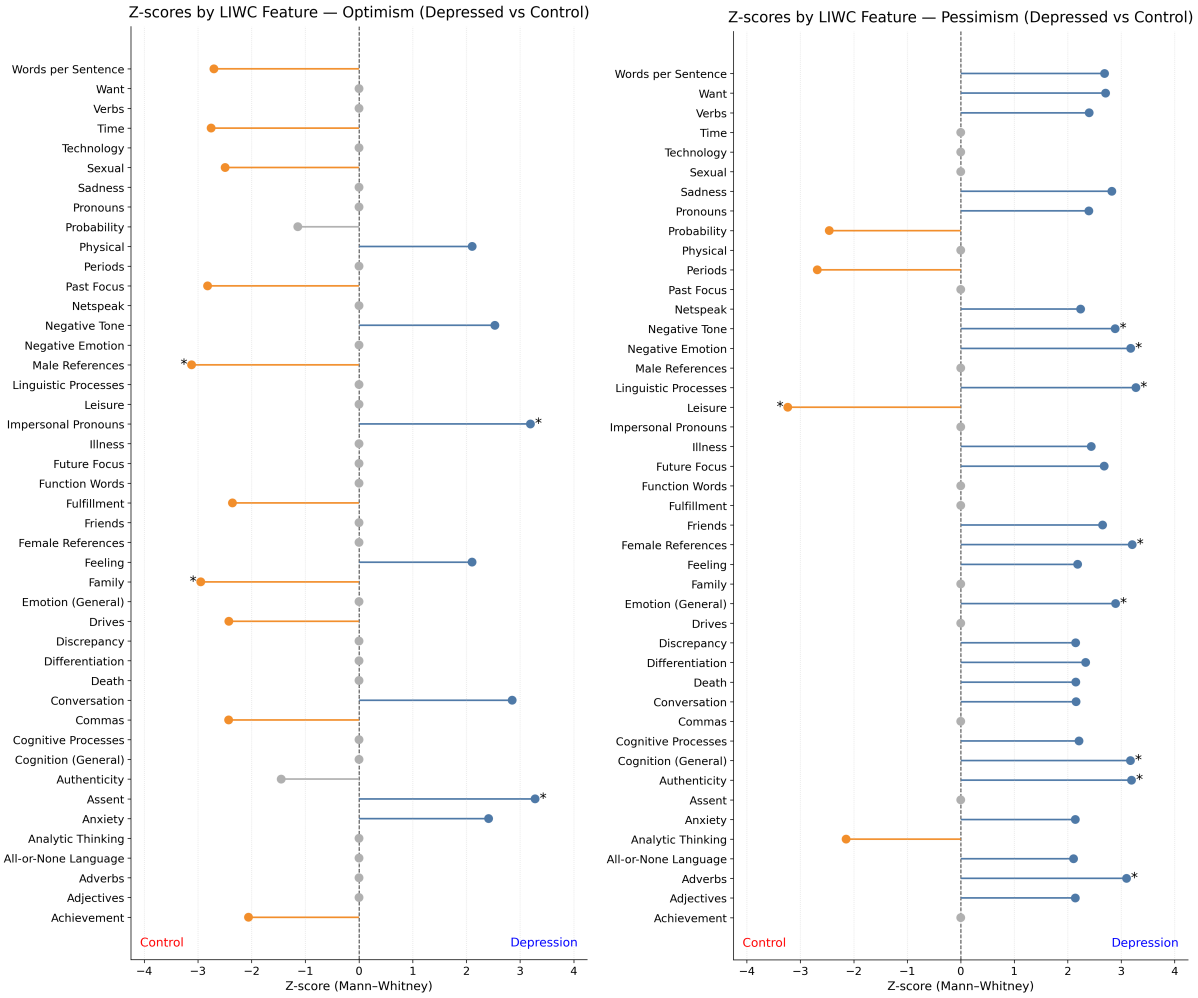

Figure 2: Statistically significant ($p<0.05$, Mann-Whitney U test) z-scores for the differences between depression and control groups for optimistic and pessimistic posts. Positive z-scores (blue, right) indicate greater feature use among participants in the depression group, and negative scores (orange, left) indicate higher use among controls. Gray dots mark non-significant features in that context; asterisks (*) denote those significant after FDR correction. Features are ordered alphabetically to enable easier direct comparison.

Group	Category	Overrepresented Topics	Underrepresented Topics
Depression	**Neutral**	Medical, AI, Online Debate	Fiction, Language, E-sports
	Optimism	Mental Health, Medical, AI	Language, E-sports, Fiction
	Pessimism	Mental Health, School, Politics	Online Debate, AI, Pets
Control	**Neutral**	Fiction, Language, E-sports	Online Debate, AI, Medical
	Optimism	Language, E-sports, Fiction	Mental Health, AI, Medical
	Pessimism	E-sports, Weight Loss, Food	Mental Health, School, Politics

Table 4: Top three overrepresented and underrepresented topics across the depression and control groups.

influence topic preferences in online discourse. The pronounced engagement of Depression-Neutral posts in online debate and artificial intelligence contrasts with the avoidance of these topics in Control-Neutral posts. On the other hand, individuals with depression seem overall more comfortable engaging in mental health-related discourse in all of the sentiment settings, even in optimistic posts. The significant engagement with e-sports in Control-Pessimism posts may indicate a preference for structured, competitive digital interactions in this category, perhaps as an outlet for engagement that

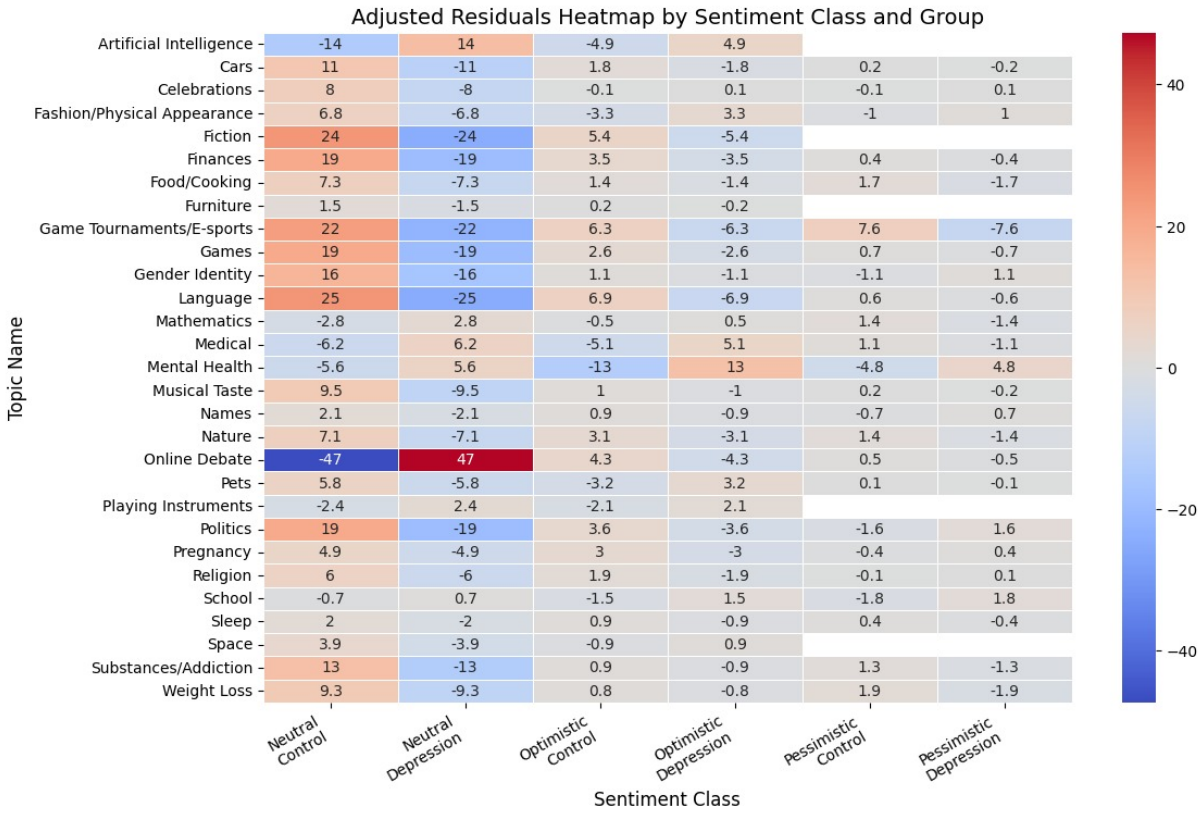

Figure 3: Heatmap of standardized residuals. The colors indicate which topics are significantly overrepresented (red) or underrepresented (blue) in each group.

does not imply personal disclosure. The control group also seems to be engaged in talks about fictional works and general leisure/lifestyle topics, which don't seem as prevalent in the depression group, a theory also supported by literature that suggests reduced engagement in such activities by people diagnosed with depression (Eisemann, 1984). This result is consistent with the observations from the LIWC feature analysis. It is worth noting that the missing values observed in the Pessimistic category (for both the depression and control groups) were intentionally excluded, as there were no posts on the respective topics belonging to that specific subgroup, making them statistically insignificant.

Coherence scores were also calculated for the generated topics, utilizing gensim (Rehurek and Sojka, 2011), both with the c_v and u_mass formulas. The obtained scores were 0.795 using the c_v formula (where the range is from 0 to 1, values closer to 1 being considered better) and -0.583, respectively, when we used the u_mass score (values around 0 being considered good). Based on the c_v score, our model generates coherent and interpretable topics, with the u_mass score supporting this view as a secondary interpretation.

5.5. Error Analysis with SHAP

The SHAP-based lexical analysis reveals some nuances in linguistic mechanisms behind the model's misclassifications. Posts falsely identified as optimistic were characterized by overtly positive terms (e.g., "good", "lovely", "favorite"), suggesting that the model might over-rely on lexical sentiment while omitting cues such as irony or contextual negation, features that often tone down positivity in natural discourse. Pessimistic misclassifications were predominantly driven by profanity and emotionally charged terms (e.g., "f*ck", "b*stards", "bloody", "motherf*cker")[1], showing that the model tends to treat emotional intensity as automatically negative. Finally, posts incorrectly labeled as neutral frequently contained mild or ambiguous emotional terms (e.g., "offensive", "sad", "sigh", "sh*t"), reflecting the model's difficulty in detecting subtle or context-dependent emotional cues.

Examining the results, the SHAP error analysis revealed systematic patterns in the model's sentiment judgments, providing insights into how various linguistic cues influence its interpretation of optimism and pessimism.

[1]We obfuscate offensive terms, according to the guidelines provided by Nozza and Hovy (2023).

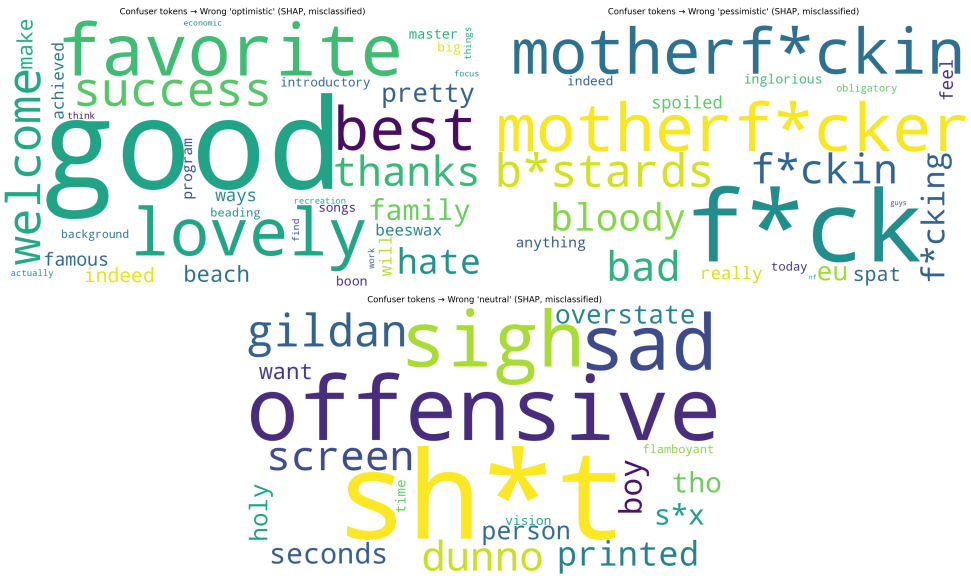

Figure 4: Wordclouds of "confusers" for each class: driving words that influence the classification towards a certain class (incorrectly), in the order of Optimism, Pessimism and Neutral classes.

5.6. Revisiting Research Questions

Addressing **RQ1**, our analyses reveal that the overall proportions of optimistic and pessimistic posts among individuals with depression are statistically similar to those of the control group. This indicates that, in terms of frequency, individuals in the depression group do not necessarily exhibit a reduced tendency to express optimism compared to control users, though the control group moderately engages more in neutral content. However, while the quantity of such expressions appears consistent, the qualitative content differs markedly.

In response to **RQ2**, our findings indicate that optimism in the social media language of individuals with depression is manifested in a more nuanced and complex manner. Although optimistic posts are present at comparable rates, the linguistic features and thematic content of these posts suggest a distinct expression of optimism that is intertwined with elements of resilience and coping. Specifically, while their optimistic posts are not characterized by a significant negative tone, they often reflect a more detached or externally directed expression of optimism. This may suggest a distancing strategy from personal agency. Notably, even within ostensibly positive contexts, individuals with depression demonstrate less engagement with cultural, lifestyle, and leisure topics, maintaining a great focus on mental health discussions.

6. Conclusions and Future Work

Our study investigated how optimism and pessimism are expressed in the social media discourse of individuals with depression using computational linguistic methods. Although no significant quantitative differences were found in the frequency of optimistic or pessimistic posts between the depression and control groups, the linguistic and thematic analyses revealed meaningful qualitative distinctions. Pessimistic posts from individuals with depression displayed a more intense and self-referential negative tone, while expressions of optimism, though often subtle, appeared to serve reflective or self-regulatory purposes, hinting at coping and resilience processes rather than unqualified positivity. These findings suggest that affective expression in depression-related discourse cannot be captured solely through sentiment polarity, and that examining the functions of optimism and pessimism may offer insight into adaptive communicative patterns in online mental health contexts. Subsequent experiments may benefit from a longitudinal approach to examine how expressions of optimism and pessimism evolve over time in relation to depressive symptoms. Additionally, integrating multimodal data, such as images, user interactions, and metadata, may provide a more comprehensive understanding of online expressions of optimism and pessimism.

Limitations

In our experiments, we used the OPT dataset collected from Twitter/X to train a transformer-based model for predicting optimism and pessimism labels in depression-related content sourced from Reddit. While there may be limitations when applying models across different platforms, previous research indicates that transformer-based models

are effective for transfer learning between platforms (Uban et al., 2022). In addition, we assessed the predicted labels on Reddit data by creating a manually annotated gold-standard subset. We chose to include the Twitter OPT dataset due to the limited availability of datasets from the same platform. The OPT dataset is the most commonly used dataset for optimism/pessimism identification (Caragea et al., 2018; Cobeli et al., 2022). Additionally, we selected the eRisk 2021 dataset because it includes social media users who have completed the validated BDI-II questionnaire, which provides a more reliable assessment than self-reported diagnoses, as is usually employed in social media depression datasets. While the BDI-II offers valuable insight into depressive symptom severity, its use as a diagnostic proxy must be interpreted with caution, given that it does not constitute a substitute for a professional clinical assessment. In this paper, all analyses are conducted using English data, and the applicability of our findings across different languages and cultures requires further investigation.

Ethical Considerations

This paper uses OPT, a publicly available dataset with annotations for optimism and pessimism. In addition, the eRisk 2021 dataset was made available to us after signing a data usage agreement form. We have adhered to the data agreement and have not attempted to contact users or de-anonymize the data. The sample of posts presented in this paper has been paraphrased to ensure the anonymity of the users. A small subset of the data was annotated, with the consent of the annotators, ensuring their anonymity. Our main focus was to quantify and analyze optimistic and pessimistic mental attitudes within the texts from the mental health dataset. We do not aim to predict mental health statuses or conditions based on this dataset.

Acknowledgements

This research was partially supported by the Ministry of Education and Research, CNCS-UEFISCDI, project SIROLA, number PN-IV-P1- PCE-2023-1701, within PNCDI IV, and by the project "Romanian Hub for Artificial Intelligence - HRIA", Smart Growth, Digitization, and Financial Instruments Program, 2021-2027, MySMIS no. 334906.

Bibliographical References

A. Alshahrani, M. Ghaffari, K. Amirizirtol, and X. Liu. 2020. Identifying optimism and pessimism in twitter messages using xlnet and deep consensus. In *International Joint Conference on Neural Networks (IJCNN)*.

A. Alshahrani, M. Ghaffari, K. Amirizirtol, and X. Liu. 2021. Optimism/pessimism prediction of twitter messages and users using bert with soft label assignment. In *International Joint Conference on Neural Networks (IJCNN)*.

Mostafa M. Amin, E. Cambria, Björn Schuller, and E. Cambria. 2023. Will affective computing emerge from foundation models and general artificial intelligence? a first evaluation of chatgpt. *IEEE Intelligent Systems*, 38:15–23.

Mario Ezra Aragon, Adrian Pastor Lopez-Monroy, Luis Carlos González-Gurrola, and Manuel Montes-y Gómez. 2021. Detecting mental disorders in social media through emotional patterns-the case of anorexia and depression. *IEEE transactions on affective computing*, 14(1):211–222.

Fabio Barbieri, José Camacho-Collados, Lisa Ellen Anke, and Leandro Neves. 2020. Tweeteval: Unified benchmark and comparative evaluation for tweet classification. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. Association for Computational Linguistics.

Aaron T Beck, Robert A Steer, Gregory K Brown, et al. 1996. Beck depression inventory.

G. Blanco and A. Lourenço. 2022. Optimism and pessimism analysis using deep learning on covid-19 related twitter conversations. *Information Processing and Management*, 59(3):102918.

Ryan L Boyd, Ashwini Ashokkumar, Sarah Seraj, and James W Pennebaker. 2022. The development and psychometric properties of liwc-22. *Austin, TX: University of Texas at Austin*, 10.

M. R. Broome, K. E. A. Saunders, P. J. Harrison, and S. Marwaha. 2015. Mood instability: Significance, definition and measurement. *British Journal of Psychiatry*, 207(4):283–285.

A. Bucur, M. Zampieri, and L. P. Dinu. 2021. An exploratory analysis of the relation between offensive language and mental health. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 3600–3606.

Ana-Maria Bucur, Berta Chulvi, Adrian Cosma, and Paolo Rosso. 2024. The expression of happiness in social media of individuals reporting depression. *IEEE Transactions on Affective Computing*.

J. Camacho-Collados, K. Rezaee, T. Riahi, A. Ushio, D. Loureiro, D. Antypas, J. Boisson, L. E. Anke, F. Liu, and E. M. Camara. 2022. Tweetnlp: Cutting-edge natural language processing for social media. In *Proceedings of the*

2022 Conference on Empirical Methods in Natural Language Processing: System Demonstrations.

C. Caragea, L. P. Dinu, and B. Dumitru. 2018. Exploring optimism and pessimism in twitter using deep learning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ş. Cobeli, I. Iordache, S. Yadav, C. Caragea, L.P. Dinu, and D. Iliescu. 2022. Detecting optimism in tweets using knowledge distillation and linguistic analysis of optimism. In *International Conference on Language Resources and Evaluation*.

A. De Santana Correia and E. L. Colombini. 2022. Attention, please! a survey of neural attention models in deep learning. *Artificial Intelligence Review*, 55(8):6037–6124.

M. Eisemann. 1984. Leisure activities of depressive patients. *Acta Psychiatrica Scandinavica*, 69(1):45–51.

Maarten Grootendorst. 2020. KeyBERT: Minimal keyword extraction with BERT. `https://doi.org/10.5281/zenodo.4461265`.

Maarten Grootendorst. 2022. Bertopic: Neural topic modeling with a class-based tf-idf procedure.

Yuting Guo, Yao Ge, Yuan-Chi Yang, Mohammed Ali Al-Garadi, and Abeed Sarker. 2022. Comparison of pretraining models and strategies for health-related social media text classification. In *Healthcare*, volume 10, page 1478. MDPI.

U. Herwig, A. B. Brühl, T. Kaffenberger, T. Baumgärtner, H. Boeker, and L. Jäncke. 2009. Neural correlates of 'pessimistic' attitude in depression. *Psychological Medicine*, 40(5):789–800.

P. Himmelstein, S. Barb, M. A. Finlayson, and K. D. Young. 2018. Linguistic analysis of the autobiographical memories of individuals with major depressive disorder. *PloS one*, 13(11):e0207814.

Catherine Hobbs, Petra Vozarova, Aarushi Sabharwal, Punit Shah, and Katherine Button. 2022. Is depression associated with reduced optimistic belief updating? *Royal Society Open Science*, 9(2):190814.

Jutta Karhu, Juha Veijola, and Mirka Hintsanen. 2024. The bidirectional relationships of optimism and pessimism with depressive symptoms in adulthood–a 15-year follow-up study from northern finland birth cohorts. *Journal of Affective Disorders*, 362:468–476.

Christoph W Korn, Tali Sharot, Hendrik Walter, Hauke R Heekeren, and Raymond J Dolan. 2014. Depression is related to an absence of optimistically biased belief updating about future life events. *Psychological medicine*, 44(3):579–592.

Kim Kronström, Hasse Karlsson, Hermann Nabi, Tuula Oksanen, Paula Salo, Noora Sjösten, Marianna Virtanen, Jaana Pentti, Mika Kivimäki, and Jussi Vahtera. 2011. Optimism and pessimism as predictors of work disability with a diagnosis of depression: a prospective cohort study of onset and recovery. *Journal of affective disorders*, 130(1-2):294–299.

Grace Y Lim, Wilson W Tam, Yanxia Lu, Cyrus S Ho, Melvyn W Zhang, and Roger C Ho. 2018. Prevalence of depression in the community from 30 countries between 1994 and 2014. *Scientific reports*, 8(1):2861.

Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *OpenReview*.

D. E. Losada and F. Crestani. 2016. A test collection for research on depression and language use. In *Lecture notes in computer science*, pages 28–39.

Scott Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *arXiv preprint arXiv:1705.07874*.

N. Mor and J. Winquist. 2002. Self-focused attention and negative affect: A meta-analysis. *Psychological Bulletin*, 128(4):638–662.

Lise Solberg Nes and Suzanne C. Segerstrom. 2006. Dispositional optimism and coping: A meta-analytic review. *Personality and Social Psychology Review*, 10(3):235–251.

Debora Nozza and Dirk Hovy. 2023. The state of profanity obfuscation in natural language processing scientific publications. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 3897–3909.

Javier Parapar, Patricia Martín-Rodilla, David E Losada, and Fabio Crestani. 2021. Overview of erisk at clef 2021: Early risk prediction on the internet (extended overview). *CLEF (Working Notes)*, 1:864–887.

Radim Rehurek and Petr Sojka. 2011. Gensim–python framework for vector space modelling. *NLP Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic*, 3(2).

Jonathan Rottenberg. 2005. Mood and emotion in major depression. *Current Directions in Psychological Science*, 14(3):167–170.

X. Ruan, S. R. Wilson, and R. Mihalcea. 2016. Finding optimists and pessimists on twitter. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*.

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*.

Michael F Scheier, Charles S Carver, and Michael W Bridges. 2001. Optimism, pessimism, and psychological well-being. *Optimism & pessimism: Implications for theory, research, and practice.*, pages 189–216.

Andrew M Subica, J Christopher Fowler, Jon D Elhai, B Christopher Frueh, Carla Sharp, Erin L Kelly, and Jon G Allen. 2014. Factor structure and diagnostic validity of the beck depression inventory–ii with adult clinical inpatients: Comparison to a gold-standard diagnostic interview. *Psychological assessment*, 26(4):1106.

Hilary Tindle, Bea Herbeck Belnap, Patricia R Houck, Sati Mazumdar, Michael F Scheier, Karen A Matthews, Fanyin He, and Bruce L Rollman. 2012. Optimism, response to treatment of depression, and rehospitalization after coronary artery bypass graft surgery. *Psychosomatic medicine*, 74(2):200–207.

Ana-Sabina Uban, Berta Chulvi, and Paolo Rosso. 2021. An emotion and cognitive based analysis of mental health disorders from social media data. *Future Generation Computer Systems*, 124:480–494.

Ana Sabina Uban, Berta Chulvi, and Paolo Rosso. 2022. Multi-aspect transfer learning for detecting low resource mental disorders on social media. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3202–3219.

Michael von Glischinski, Ruth von Brachel, and Gerrit Hirschfeld. 2019. How depressed is "depressed"? a systematic review and diagnostic meta-analysis of optimal cut points for the beck depression inventory revised (bdi-ii). *Quality of Life Research*, 28:1111–1118.

Dong Xu, Yi-Lun Wang, Kun-Tang Wang, Yue Wang, Xin-Ran Dong, Jie Tang, and Yuan-Lu Cui. 2021. A scientometrics analysis and visualization of depressive disorder. *Current neuropharmacology*, 19(6):766–786.

Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V. Le. 2019. Xlnet: Generalized autoregressive pretraining for language understanding. *arXiv preprint arXiv:1906.08237*.