Modeling Depressive Patterns in Italian Discourse: Insights from Natural Language Processing

by

Annapia Borraccino

A thesis submitted to the Graduate Faculty in Linguistics in partial fulfillment of the requirements for the degree of Master of Arts, The City University of New York

2025

APPROVAL

Modeling Depressive Patterns in Italian Discourse: Insights from Natural Language Processing

by

Annapia Borraccino

This manuscript has been read and accepted by the Graduate Faculty in Linguistics in satisfaction of the thesis requirement for the degree of Master of Arts.

Approved: April 2025

Rivka Levitan, Advisor

Sarah Ita Levitan, Second Reader

Jason Kandybowicz, Executive Officer

The City University of New York

Abstract

MODELING DEPRESSIVE PATTERNS IN ITALIAN DISCOURSE:

INSIGHTS FROM NATURAL LANGUAGE PROCESSING

by

ANNAPIA BORRACCINO

Advisor: Rivka Levitan

Depression does not always speak in obvious ways. It often hides in silence, in brief pauses, and in the quiet texture of language. While the prosodic and linguistic characteristics of depression have been extensively studied in English and other languages, they remain underexplored in Italian, with this being a first focused analysis. This study investigates how depression manifests in the speech of Italian speakers, using transcriptions from a clinical corpus. Drawing on emotional, structural, and cognitive-linguistic features, including first person verb use, sentence complexity, verb tense, and negation, the study evaluates how these markers differ between individuals with and without depression. Even small changes, such as favoring the past tense over the present, appeared to carry psychological meaning. A baseline model trained on these features showed promising results, indicating that minor linguistic variations can reflect deeper emotional states. The study also trained a deep learning model known as BERTino, which achieved an accuracy rate of 92%. Unlike traditional models, this transformer was able to pick up on more subtle expressions of emotional distress, signals that may escape both listener and speaker. This study sheds light on how depression surfaces in the language of Italian speakers and highlights the potential of combining psychologically grounded features with deep learning to better understand what is often left unsaid.

Acknowledgments

First and foremost, I would like to express my deepest gratitude to my family. Your unwavering love, patience, and sacrifice made it possible for me to pursue my studies here in the United States. Thank you for believing in me every step of the way and for providing the support that sustained me through this journey.

I am especially grateful to my advisor, Professor Rivka Levitan. Your feedback, insightful guidance, and constant encouragement helped shape the ideas that underpin this project. Thank you for challenging me to refine my thinking and for generously sharing your expertise.

Finally, I would like to acknowledge the many friends, colleagues, and faculty members who offered advice and shared in both the challenges and triumphs of this process. Although I cannot name you all here, your contributions have been invaluable, and I thank you for your support and camaraderie.

Contents

List of Tables

Chapter 1

Introduction

Depression is a widespread and debilitating mental health condition that affects millions of individuals globally, circa 5% of the adult population according to the World Health Organization (WHO) as of 2023. A condition which interferes with emotional well-being, cognitive functioning, and in some cases with simple daily life tasks of those who are affected. Despite its prevalence, depression remains frequently misunderstood and stigmatized, particularly in cultural contexts where mental illness is not openly discussed nor fully accepted. In Italy, for instance, the public understanding of depression is shaped by longstanding cultural beliefs and social attitudes that often hinder open dialogue and timely intervention. A national survey conducted by Munizza et al. (2013) highlights the extent of this issue. In fact, while nearly all Italians surveyed (98%) were aware of depression, a substantial portion, representing the 55%, still viewed it as a sign of personal weakness, and 75% believed it should not be openly discussed. Moreover, many respondents believed depression could be resolved without professional help, and 27% considered those with depression to be dangerous, reinforcing harmful stereotypes. These beliefs contribute to a pervasive climate of stigma that not only discourages individuals from seeking appropriate care, but also leads many to avoid acknowledging their symptoms or pursuing assessment, even when they suspect the early onset of depression. For example, while primary care physicians are often the first point of contact in the healthcare system, the survey found that many Italians felt embarrassed to speak with their doctor about depression, believing that physicians were either too busy or inadequately prepared to address mental health concerns. Pharmacological treatments were also viewed with skepticism, with many respondents expressing fear of addiction and side effects. These findings reveal a clear disconnect between public beliefs and clinical knowledge, as well as a critical need for tools that can bridge this gap. Although the data from Munizza et al. (2013) may be somewhat dated, familiarity with contemporary Italian cultural attitudes suggests that little has changed in public discourse around depression. While younger generations are gradually more open to discussing mental health, older generations still tend to approach the subject with skepticism and extreme discretion. Therefore, in contexts where stigma impedes people to seek support, Natural Language Processing (NLP) offers a promising avenue for the early detection of depression, particularly through the analysis of subtle linguistic markers in everyday text. Depression frequently manifests in language which makes automated language analysis a powerful, nonintrusive method for identifying emotional distress.

This thesis explores how emotional and cognitive language features manifest in transcribed interviews conducted in Italian, with the aim of supporting depression detection. Importantly, it is the first to analyze depressive language in Italian purely through text and raises the state-of-the-art on this corpus beyond the work of Tao et al. (2020).
The motivation for this work is both academic and societal: to address a clear research gap in the computational study of depression in Italian, and to contribute to a broader conversation

about recognizing depression as a legitimate and urgent health issue.

1.1 Research Questions

Building on this motivation, the present study is guided by two research questions aimed at exploring how language reflects depressive symptomatology. The first question is: what role do emotional and cognitive language features play in distinguishing depressed individuals from non-depressed individuals? Drawing on established psychological theories and previous NLP studies, mostly based on English corpora, this project aims at investigating whether features such as negative emotional language or self focused pronouns, appear with greater frequency also in the case of depressed Italian speakers. The second research question asks: what linguistic features do deep learning models, in this case transformer models, rely on for depression detection, and how do these align with psychological theories of depression? To address this, a transformer model is fine-tuned on the binary classification task of distinguishing depressed from non-depressed participants. Beyond classification performance, this study also interrogates model interpretability with the goal of visualizing attention patterns and highlighting which words or structures the model deems most salient. The aim is to examine whether these attention patterns correspond to psychologically meaningful features, such as increased self referential language or expressions of hopelessness.[1]

[1] `https://github.com/annapiab/Depressive-Italian-Discourse-nlp`

Chapter 2

Literature Review

2.1 Psychological Foundations

Before diving into what previous studies on depression detection in the NLP field discovered, it is essential to explore the origins of the intersection between psychology and linguistics. Beck's (1967) Cognitive Theory of Depression is relevant to understanding how cognitive and emotional processes manifest linguistically in individuals with depression. According to the author, negative cognitive schemas predispose individuals to depressive thought patterns, often leading to automatic negative interpretations of events involving the self, the external world, and future possibilities. When activated by stressors like loss or rejection, these schemas shape not only internal thought processes but also outward linguistic expressions. The negative cognitive triad manifests itself through an increased use of negative emotion words and a reduced positive emotional expressions, reflecting the dominance of negative appraisal systems. Further studies (Tølbøll, 2019) consistently find that depressed individuals rely more heavily on first person singular pronouns, indicating a heightened state of self focus that further perpetuates negative self evaluation. These linguistic markers serve as a lens through which underlying psychological processes can be observed, offering empirical support for previous theories. These psychological insights provide a theoretical backdrop for understanding how depression manifests linguistically, an idea that has been empirically explored across various languages using computational tools.

2.2 Language Depression Markers: Evidence Across Languages

Research into the linguistic manifestations of depression has consistently demonstrated that depressed individuals exhibit distinct speech and writing patterns that reflect their cognitive and emotional states. Automated text analysis tools such as Linguistic Inquiry and Word Count (LIWC) have been instrumental in identifying these markers across various studies. A recent study by Trifu et al. (2024) applied the RO-LIWC 2015, the Romanian translation of LIWC, to analyze the speech of 62 patients diagnosed with Major Depressive Disorder (MDD) and 43 control participants. Their findings revealed that depressed individuals tend to produce shorter sentences, resulting in increased use of punctuation marks, specifically periods. This may suggest a more direct and restricted communication style. Differently from previous English based studies that identified a strong correlation between depression and heightened first person singular pronoun use ("I"), Trifu et al. (2024) found that MDD patients favored impersonal pronouns and first person plural pronouns ("we"), suggesting a shift away from individualistic expression. This discrepancy may be attributed to linguistic and cultural differences, particularly the pronoun-drop feature in Romanian, which makes self referential speech more implicit. The study also stresses the importance of cultural and linguistic adaptations when applying LIWC-based models across different languages, as linguistic structures like pronoun-drop in Romance languages may obscure markers commonly observed in English corpora. The Romanian study not only reinforces the presence of depression-related linguistic patterns but also reveals key limitations in using English-based tools across languages, highlighting the need for culturally adapted alternatives.

2.3 Limitations of English-Centric Tools and the Case for ELIta

The above findings not only validate prior research but also emphasize the need for more cross-linguistic investigations to refine automated depression detection models and tailor them for morphologically rich languages. In fact, this growing recognition of linguistic and cultural variability in the expression of depression has prompted researchers to seek out more language specific tools capable of capturing affective nuances that may be obscured by direct translations of English based resources. While the LIWC dictionary has been widely adopted in psychological text analysis, its Italian version remains limited in both scope and granularity. In response to these limitations, recent efforts have produced culturally grounded alternatives, such as ELIta - the Emotion Lexicon for Italian. Developed by Di Palma (2024), ELIta represents a significant advancement in Italian affective lexical resources, while also being open source. As opposed to earlier lexicons, ELIta integrates both categorical emotion annotations and dimensional evaluations (valence, arousal, and dominance), allowing for more flexible and theoretically grounded analyses. In contrast to LIWC's fixed and aggregated format, ELIta is notable for offering a non aggregated version of its data, enabling researchers to examine variation in emotional interpretation across annotators. Additionally, it includes a broader range of lexical items, encompassing not only verbs, nouns, adjectives, and adverbs, but also idiomatic expressions and emojis, making it well suited to both formal and informal text types in Italian. In this project, ELIta is employed as a customizable tool, with selected features tailored to the analytical aims of the study. Rather than applying the lexicon wholesale, this thesis will adopt a targeted approach, drawing on specific emotional categories and dimensions from ELIta that are most relevant to identifying depression related language patterns. While lexical resources like ELIta enhance our ability to identify emotional content, they are only part of the picture. A deeper look at syntactic and grammatical structures also reveals distinct linguistic patterns associated with depression.

2.4 Syntactic and Lexical Patterns of Depressive Speech

Further supporting prior research demonstrated that depressed individuals' speech patterns reflected dichotomous thinking, characterized by a limited use of tentative words, a linguistic feature associated with cognitive rigidity in depression (Al-Mosaiwi & Johnstone, 2018). Additionally, the study uncovered grammatical patterns unique to depression, including fewer prepositions but increased use of auxiliary verbs, conjunctions, and negations, indicating a structured yet restrictive linguistic style. These patients also relied more heavily on past tense verbs, while their usage of present and future tense verbs was significantly lower, suggesting diminished prospective focus and goal formulation skills. Beyond syntax and lexicon, the study found that individuals with MDD express heightened concern with topics related to health, time, and money, pointing to preoccupations with personal struggles and existential concerns. Morales and Levitan (2016) add that while textual features such as n-grams and LIWC-derived scores yielded strong predictive performance, prosodic cues in speech also contributed valuable information, highlighting the potential benefits of integrating multimodal approaches to depression detection. These patterns form the basis for feature extraction in machine learning models, which increasingly power automated depression detection systems.

2.5 Integrating NLP Techniques and Model Performance

Recent research in NLP-driven depression detection has increasingly turned to machine learning to enhance diagnostic accuracy by leveraging linguistic features. Trotzek et al. (2020), for instance, explored early depression detection using social media texts and compared a convolutional neural network (CNN) trained on word embeddings with a user-level classifier incorporating LIWC-derived metadata. Their hybrid approach achieved state-of-the-art performance and confirmed several depression markers, including increased use of first person singular pronouns, high frequency of negative emotion words, and reduced use of positive emotional expressions. Notably, their models also detected elevated use of past tense and absolutist terms (e.g., “always,” “never,” “nothing”). While these results highlight the predictive potential of integrating linguistic metadata with deep architectures, they also underscore ethical concerns such as dataset bias, privacy risks in social media monitoring, and the need for explainable AI. Building on these findings, cross-linguistic research has emphasized the importance of language-specific NLP adaptations. Spruit et al. (2022), analyzing Dutch psychiatric interviews, found that self referential language, repetitive phrasing, and reduced emotional range remained consistent indicators of depressive speech, even in morphologically rich languages. Their evaluation showed that traditional models using LIWC features and Random Forest classifiers outperformed deep learning approaches like fastText and RobBERT (a Dutch BERT variant). Similarly, Lorenzoni et al. (2024) conducted a comparative analysis of classifiers on the DAIC-WOZ corpus, confirming that depressed speakers displayed lower lexical diversity, more repetition, and increased use of negations and self focused language. Random Forest and XGBoost models outperformed SVMs in this study, achieving 84% accuracy. However, both studies highlighted challenges such as data imbalance and limited interpretability, reinforcing the need for transparent, linguistically grounded systems in clinical applications.

2.6 Advancements in Deep Learning for Depression Detection

Recent work has introduced hybrid deep learning models that combine linguistic insight with the sequential processing capabilities of neural networks. Tejaswini et al. (2022) proposed a FastText-CNN-LSTM (FCL) architecture trained on social media texts, successfully identifying depressive markers such as first person pronouns, negative emotion words, repetitive phrasing, and limited lexical variety. Their model outperformed standalone CNN and LSTM baselines, reaching an accuracy of 88%. Amanat et al. (2022) addressed the limitations of earlier frameworks by introducing an LSTM-RNN architecture enhanced with attention mechanisms. This model, trained on a Kaggle dataset of over 4,000 tweets, achieved 99% accuracy, substantially higher than traditional SVM and TF-IDF models. The authors attributed this performance to a robust preprocessing pipeline involving One-Hot encoding and Principal Component Analysis (PCA), combined with context-sensitive attention layers. While these advancements are promising, they also raise concerns about generalizability across domains. Models trained on social media may not transfer easily to structured clinical settings, and interpretability remains a significant hurdle. Nonetheless, these studies illustrate how combining linguistic metadata with deep learning techniques enables more accurate and flexible depression detection. The need for explainable AI and culturally sensitive model adaptation continues to be a central concern as the field evolves.

2.7 Summary

The linguistic manifestations of depression have been widely studied, with research consistently demonstrating that language serves as a valuable indicator of depressive symptoms. Across multiple studies, key linguistic markers such as increased self focus (first person singular pronouns), heightened use of negative emotion words, reduced lexical diversity, have been identified as distinguishing characteristics of depressive language. Automated text analysis tools like LIWC have played a central role in identifying these patterns, while cross-linguistic research has highlighted the need for language specific adaptations to improve detection accuracy across different grammatical and morphological structures. With the advancement of NLP and machine learning, depression detection models have increasingly incorporated hybrid approaches, integrating linguistic metadata with deep learning architectures to enhance predictive accuracy. Studies employing machine learning classifiers and deep learning have demonstrated promising results in detecting depression from text inputs, particularly in social media data and clinical transcripts. However, challenges remain in ensuring interpretability, mitigating bias in training data, and adapting models for different linguistic and cultural contexts. While research in this area continues to evolve, it is evident that computational linguistic analysis offers significant potential for improving depression detection and monitoring.

Chapter 3

Methodology

3.1 Data Acquisition

The data used for this project was obtained from a recently published study by Tao et al. (2023), titled *"The Androids Corpus: A New Publicly Available Benchmark for Speech-Based Depression Detection"*. The dataset comprises audio samples from 118 native Italian speakers, including 64 participants diagnosed with depression and 54 control participants with no known history of mental health issues. Diagnosis was conducted by psychiatrists following the criteria outlined in the Diagnostic and Statistical Manual of Mental Disorders, or, DSM-5 (2022). Among the 64 diagnosed individuals, the distribution includes 22 cases of Major Depressive Disorder, 15 of bipolar disorder in a depressive phase, 8 of reactive depression, 7 of endo-reactive depression, 5 with anxious-depressive disorder, and 1 with persistent depressive disorder. For the remaining six, no specific pathology was reported, though they were clinically classified as depressed. This diversity enables a rich representation of depressive symptomatology within the dataset. However, for the purposes of this study, patients will not be differentiated by severity or subtype of depression, they will all be treated uniformly as depressed.

Participants were asked to complete two distinct tasks: a Reading Task (RT) and an Interview Task (IT). The RT involved reading Aesop's fable *"The North Wind and the Sun"*, a simple and culturally neutral text chosen to minimize the influence of educational background on reading fluency. The IT consisted of answering open-ended questions related to everyday life (e.g. *'What were you up to last week?', 'Tell me about your family', 'What did you do yesterday?'*), and producing spontaneous speech. These two modalities enable the collection of both read and natural speech samples, offering insights into how depression may manifest across different speaking contexts. 110 of the 118 participants completed both tasks, while the remaining few contributed to only one. Crucially, participants in the control group were matched with the depression group in terms of age, gender, and educational background, minimizing potential confounding variables. The dataset reflects a higher proportion of female participants (approximately 2.5 times more than males), in line with epidemiological evidence indicating that depression is more commonly diagnosed in women.

Table 3.1: Demographics of the IT group

Task	Age (Mean ± SD)	M	F	L	H
Control PT	47.3 ± 12.7	11	41	19	33
Depressed PT	47.5 ± 11.6	21	43	29	33
Total	47.4 ± 12.1	32	84	48	66

* Acronyms F and M stand for Female and Male.
† Acronyms L and H stand for Low (8 years of study at most) and High (at least 13 years of study) education level, respectively.

All audio recordings were collected while the depressed participants were actively experiencing the disorder, supporting the assumption that depressive symptoms would be detectable in their speech patterns. For the purposes of this study, only the Interview Task responses were transcribed, totaling 115 interviews, using Whisper, an automatic speech recognition model developed by OpenAI. Since this research focuses on the application of NLP to text, speech analysis was excluded, and the dataset was limited to textual transcriptions of spontaneous speech. Although speech patterns offer valuable insights, they fall outside the scope of this study and may be explored in future work. While this decision reduces the overall size of the dataset, it aligns with the primary objective of examining linguistic and affective markers of depression in written transcriptions. In addition to the interview transcriptions, this study also required an external resource for emotional language annotation. Since the Italian version of LIWC was not accessible, the ELIta lexicon (2024) was chosen as an alternative. The lexicon was retrieved from its official GitHub repository, and the ELIta Golden version was selected, which provides a curated set of annotations based on a non-aggregated format. In fact, for each word, the five most consistent annotations were retained, and a gold standard label was generated using a majority vote mechanism. This approach ensures both annotation reliability and lexical richness, with the dataset achieving an inter-annotator agreement of 0.87. For the purpose of this study, the Valence-Arousal-Dominance (VAD) dimensions were excluded during pre-processing, as they were not directly relevant to the project's focus on discrete emotional categories and their relationship to depressive language patterns.[1]

[1] `https://github.com/LaurentFossati/Androids-Corpus`

3.2 Data Pre-processing

Following data acquisition, the IT audio samples were transcribed using the Whisper speech recognition system. To ensure the reliability of the textual data, each transcription was subsequently manually reviewed and corrected for transcription errors, misrecognitions, or formatting inconsistencies. During this process, one interview was excluded from the dataset as the audio was spoken entirely in Neapolitan dialect, and Whisper was unable to generate an intelligible transcription. A manual attempt at transcription was also unsuccessful due to limited intelligibility and lack of standardized orthography. Therefore, the final dataset comprises 115 manually verified transcriptions of spontaneous Italian speech, which serve as the foundation for all subsequent linguistic and computational analyses. Although the Androids Corpus (Tao et al., 2023) uses filename conventions to indicate speaker groups, "P" for clinically diagnosed patients and "C" for control participants, each transcript was manually reviewed and relabeled to ensure consistency and clarity in the dataset. Labels were standardized to binary format, with 1 indicating depressed participants and 0 indicating non-depressed controls, resulting in 63 transcripts from clinically depressed individuals and 52 from non-depressed individuals. While modest in size, the dataset is relatively balanced, providing a suitable basis for exploratory analysis and classification tasks.

Linguistic features related to emotion were extracted from the interview transcripts using a preprocessing pipeline implemented with the Stanza NLP library, which offers robust support for Italian language processing. Each transcript was first tokenized, lemmatized, and POS tagged, allowing for more accurate lexical matching against the ELIta lexicon, which is also lemmatized. Lemmatization was particularly crucial given the inflectional nature of Italian, where words can appear in many morphological variants; by reducing words to their base form, this step ensured that lexical comparisons with the ELIta entries were consistent and meaningful. The filtered ELIta lexicon was then converted into a dictionary for efficient lookup. For each transcript, the pipeline computed average emotion scores across all matched words, producing a vector of categorical emotion values per speaker. Transcripts with no lexical matches were flagged to ensure transparency in coverage. These emotion scores were then combined with participant labels and saved in a structured dataset for downstream analysis. The preprocessing steps were kept intentionally focused and interpretable, prioritizing linguistic fidelity in Italian and compatibility with emotion based feature extraction, rather than employing more abstract transformations.

3.3 Feature Extraction

The investigation of linguistic markers of depression focused on extracting two main categories of features from the corpus:

1. emotional features using the ELIta lexicon

2. cognitive and structural linguistic features using morphological analysis from Stanza.

3.3.1 Emotional Features

The emotional features were derived using ELIta, a manually annotated Italian emotion lexicon which assigns each word a score, ranging from 0 to 1, across several basic emotion dimensions, including *tristezza* (sadness), *rabbia* (anger), *gioia* (joy), and *fiducia* (trust). The lexicon was loaded as a dictionary for efficient lookup. Lemmatization was used to ensure alignment with ELIta's base form entries. Tokens were then matched against ELIta: for each matching lemma, the corresponding emotion scores were stored. Finally, the mean value for each emotion category was computed across all matched tokens, yielding one numerical feature per category. If a transcript contained no matching lemmas for a given emotion, a default score of 0 was assigned.

3.3.2 Linguistic Features

The linguistic features were extracted using Stanza's POS and morphological tagging pipeline. The analysis focused on structural and cognitive markers known to be linked with depressive language patterns, including pronoun usage, verb tense, negation, and sentence structure. Each transcript was processed and the following features were computed:

- First person pronouns: counted dynamically using morphological features (Person=1) and POS tag PRON.

- First person singular verbs: included to compensate for Italian being a pro-drop language, where subject pronouns are often omitted. These verbs were identified using morphological features marked as Person=1 and Number=Sing.

- First person total: the sum of first person pronouns and first person verbs, used as a composite indicator of self focus.

- Negations: count of negative markers such as non, mai, niente, and related terms.

- Verb tense: number of verbs in past and present tense, extracted via morphological features Tense=Past and Tense=Pres.

- Average sentence length: calculated as the number of tokens per sentence, used as a proxy for sentence complexity.

- Total word count and total sentence count.
- Additional normalized features were computed per sentence (e.g., first person per sentence, negation per sentence), but were not included in the final modeling.

All numerical features were rounded to four decimal places and stored for integration into downstream classification tasks. The emotional and linguistic feature sets were later merged into a final dataset used for training.

3.4 Baseline Modeling

Traditional machine learning models were implemented using structured feature sets derived from the linguistic and emotional dimensions of the transcriptions, providing an interpretable benchmark for the depression classification task. These models aimed to assess the predictive utility of psycholinguistic cues when used in isolation from deep contextual embeddings. Following feature extraction, two complementary feature sets were constructed. The first merged emotional scores from the ELIta lexicon and linguistically grounded features (such as pronoun usage, verb tense, sentence complexity), resulting in a structured feature set. The second combined this feature set with lexical content captured through Term Frequency–Inverse Document Frequency (TF-IDF) vectorization of the raw transcriptions. The final dataset included both these inputs and the binary classification label (0 = control, 1 = depressed), and was used for supervised training. Feature selection was grounded in prior literature. Seven features were retained for modeling based on theoretical relevance and empirical utility: sadness, anger, total use of first person markers, negation frequency, present and past tense usage per sentence, and average sentence length. These features were then standardized using z-score normalization, transforming each feature x into a standardized value z as follows:

$$z = \frac{x - \mu}{\sigma}$$

where μ and σ represent the mean and standard deviation of the feature across the dataset. This normalization ensured that all features contributed comparably during training. Lexical information was encoded using TF-IDF, a method that weighs terms by how frequently they appear in a document relative to their rarity across the corpus. For each term t in document d, its TF-IDF score is computed as:

$$\text{TF-IDF}(t, d) \; = \; \text{tf}(t, d) \; \cdot \; \log\!\Big(\frac{N}{\text{df}(t)}\Big)$$

where N is the total number of documents and *df(t)* is the number of documents in which t appears. This approach downweights common stopwords and emphasizes terms that are distinctive to individual speakers. The standardized linguistic and emotional features were concatenated with the TF-IDF vectors to form a single input matrix for classification. Two supervised classifiers were trained on this input: logistic regression and a linear kernel support vector machine (SVM). Logistic regression models the probability of a binary outcome using the sigmoid function:

$$P\big(y = 1 \mid \mathbf{x}\big) = \frac{1}{1 + \exp\!\big(-(\beta_0 + \beta_1 x_1 + \cdots + \beta_n x_n)\big)}$$

where x is the feature vector and β_i are the learned weights for each feature. This formulation enables not only classification, but also interpretation of feature importance based on weight magnitudes. The SVM classifier, by contrast, constructs a maximum margin hyperplane to separate the two classes and is particularly effective in high dimensional spaces. Model performance was evaluated using a stratified 80/20 train-test split and 5-fold stratified cross-validation. This provided both a point estimate of predictive performance and a more robust measure of generalizability across the relatively small dataset.

3.5 Evaluation of the Baseline Model

The performance of the baseline classification model was evaluated using both a stratified train-test split and a 5-fold cross-validation procedure. This stratification was necessary due to the moderate class imbalance present in the dataset, and it ensured that both classes were represented during training and evaluation. Model performance was assessed using standard classification metrics, including accuracy, precision, recall, and F1-score. Accuracy provided an overall measure of correct predictions, while precision and recall were especially relevant for the depressed class, which carried greater diagnostic importance. The F1-score, as the harmonic mean of precision and recall, was used to summarize the model's ability to correctly identify depressed individuals while avoiding false positives. These metrics were reported using scikit-learn's classification report, allowing for detailed breakdowns per class. To further validate the generalizability of the model, a cross-validation procedure was implemented. In this setup, the dataset was split into five equally sized folds, preserving class distribution in each fold. The model was trained on four folds and evaluated on the remaining fold, iteratively rotating through all combinations. The mean accuracy across the five folds, along with its standard deviation, was used as an estimate of the model's robustness and stability. This type of evaluation provided a more reliable performance estimate than a single split and mitigated the risk of overfitting to a particular subset of the data.

3.6 Deep Learning Model

A core question of this study is whether advanced neural language models rely on the same emotional and cognitive linguistic patterns that have been usually associated with depression in psychological research. To investigate this, a transformer architecture was fine-tuned using the BERTino model, a pre-trained transformer specifically optimized for Italian. Deep learning models learn to classify text based on contextual information and distributed representations of words. Testing such a model allows for a direct comparison between psychologically grounded features and the patterns implicitly prioritized by a state-of-the-art neural model. This approach not only provides a performance benchmark, but also serves as a means of examining whether these predictions align with known markers of depressive language. Each transcript was passed in its raw textual form and tokenized using the BERTino tokenizer, which applies WordPiece tokenization and automatically adds the special [CLS] token required by the BERT architecture. Although not handled manually in the code, this token is inserted at the beginning of every sequence and plays a critical role in classification. During inference, the final hidden state corresponding to the [CLS] token is extracted and passed to a linear classification layer. This layer outputs a pair of unnormalized logits, which are then transformed into probabilities using the softmax function. For the binary classification task, the predicted probability of the transcript being classified as depressed is given by:

$$P\big(y = 1 \mid \mathbf{x}\big) = \frac{e^{z_1}}{e^{z_0} + e^{z_1}}$$

where z_0 and z_1 represent the logits for the non-depressed and depressed classes, respectively.

This probabilistic output allows the model to make binary predictions while encoding classification confidence. The dataset was split into training and testing subsets using an 80/20 stratified split to maintain class balance. Fine-tuning was performed using the HuggingFace Trainer API, which abstracts much of the training loop. Although the loss function is not manually defined in the code, when labels are provided and the model is configured for classification, the Trainer automatically applies a standard cross-entropy loss:

$$\mathcal{L} = -\big[y\ \log(\hat{y}) + (1-y)\ \log\big(1-\hat{y}\big)\big]$$

This objective function penalizes confident incorrect predictions and is well suited for binary classification tasks such as this one. Training was conducted for up to eight epochs, with early stopping triggered if the validation loss did not improve over two consecutive epochs. A learning rate of 1×10^{-5} was used, with weight decay applied to prevent overfitting. Mixed precision training (fp16) was enabled to reduce memory usage and accelerate computation on GPU.

3.7 Evaluation of the Transformer Model

The fine-tuned BERTino model was evaluated on the held out 20% of the dataset using standard classification metrics, including accuracy, precision, recall, and F1-score. Predictions were obtained by passing the tokenized test set through the model in evaluation mode, extracting the logits, and selecting the most probable class using the argmax function. These predictions were then compared against the true labels to assess performance.

Evaluation metrics were computed using scikit-learn's classification report. Given the focus on depression detection, particular attention was paid to the recall of the depressed class, as false negatives are especially undesirable in clinical or screening settings. Precision and F1-score were also considered to assess the model's ability to make balanced and reliable predictions. Unlike the baseline model, no cross-validation was performed on the transformer due to its higher computational cost. However, the evaluation based on a stratified 80/20 split provided a representative estimate of model generalization. The resulting performance metrics served as a direct point of comparison with the feature based SVM baseline, allowing for analysis of whether the transformer captured more subtle or complex linguistic patterns associated with depression.

3.8 Additional Analysis Tools

To better understand what linguistic cues the transformer model relied on, a set of interpretability techniques was applied after training. The model was used to assign depression probabilities to all transcripts in the dataset. Transcripts were then ranked by their predicted scores, and the top ten most confidently classified as depressed were selected for further analysis. Attention weights from BERTino's final layer were extracted for these top transcripts to identify which tokens received the most focus. Token level attention scores were averaged and mapped back to words, filtering out common Italian stopwords while retaining potentially relevant terms such as negations and first person pronouns. These scores were then visualized using a word cloud, offering insight into the linguistic patterns the model prioritized when identifying depression.

Chapter 4

Results

To assess whether the observed differences in emotional and structural language features between depressed and non-depressed individuals were statistically meaningful, two-sample independent t-tests were conducted. This test compares the means of two independent groups and determines whether any observed difference is likely due to chance or reflects a real underlying effect. The t-statistic is calculated using the following simplified formula:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

where $\bar{X}_1$ and $\bar{X}_2$ are the sample means, s_1^2 and s_2^2 are the variances, and n_1 and n_2 are the sample sizes of the two groups. Degrees of freedom were calculated as $n_1 + n_2 = 114$. Differences with ρ-values less than 0.05 were considered statistically significant.

4.1 Emotional and Structural Patterns in Depressed Language

Emotional language patterns provide an initial window into how depression may surface linguistically. Based on an initial analysis, individuals labeled as depressed (1) demonstrated slightly higher average use of sadness (*tristezza*) 0.1448 compared to the 0.1194 resulting from non-depressed individuals. Depressed patients also manifested a slightly higher fear (*paura*), anger (*rabbia*), and disgust (*disgusto*) compared to non-depressed controls. In contrast, non-depressed individuals used more positively valenced words, showing higher average scores for joy (*gioia*), trust (*fiducia*), surprise (*sorpresa*), and anticipation (*aspettativa*), where anticipation is intended as 'anticipatory emotional states', which aligns most closely with the LIWC category "anticipation" (in the English version, these would match words like *hope, expect, await*).

Table 4.1: Average Emotion Scores by Class

Emotion	**Controls (M)**	**Depressed (M)**	**t(114)**	**ρ-value**
Gioia (joy)	0.39	0.32	3.43	0.00
Tristezza (sadness)	0.12	0.14	1.51	0.13
Rabbia (anger)	0.08	0.10	1.29	0.20
Disgusto (disgust)	0.03	0.03	0.31	0.76
Paura (fear)	0.15	0.17	1.01	0.31
Fiducia (trust)	0.31	0.28	2.18	0.03
Sorpresa (surprise)	0.21	0.18	1.79	0.08
Aspettativa (anticipation)	0.45	0.40	2.07	0.04

The linguistic analysis revealed measurable differences in how depressed and non-depressed individuals use language. Patients produced more first person pronouns (M = 12.02, SD = 9.80) and first person verbs (M = 12.89, SD = 7.14) than controls (M = 8.75, SD = 4.44;

M = 9.29, SD = 5.12), with both differences reaching statistical significance (t(114) = 2.40, $\rho = 0.018$; t(114) = 2.89, $\rho = 0.005$, respectively). The combined first person total was also significantly higher in the depressed group (M = 24.91, SD = 16.07) compared to controls (M = 18.04, SD = 8.22). Negation usage followed a similar trend, with depressed participants using more negations (M = 8.35, SD = 6.48) than controls (M = 5.53, SD = 3.48), a statistically significant difference (t(114) = 2.80, $\rho = 0.006$). Present tense verb usage was slightly higher in the depressed group (M = 24.45, SD = 16.00) compared to controls (M = 20.96, SD = 8.77), but this difference was not statistically significant (t(114) = 1.34, p = 0.18). No significant group differences were found for past tense usage, average sentence length, or overall word-sentence count. These results help quantify lexical and syntactic tendencies that distinguish depressive from non-depressive language use, particularly in markers of self focus and negation. The following table outlines the key findings from the t-test analysis.

Table 4.2: Feature Comparison Table

Feature	**Controls (M ± SD)**	**Patients (M ± SD)**	**t(114)**	**ρ-value**
First person pronouns	8.75 ± 4.44	12.02 ± 9.80	2.40	0.018
First person verbs	9.29 ± 5.12	12.89 ± 7.14	2.89	0.005
First person total	18.04 ± 8.22	24.91 ± 16.07	2.91	0.004
Negations	5.53 ± 3.48	8.35 ± 6.48	2.80	0.006
Past tense verbs	11.67 ± 6.02	11.18 ± 9.15	0.36	0.720
Present tense verbs	20.96 ± 8.77	24.45 ± 16.00	1.34	0.180
Avg. sentence length	22.38 ± 3.84	22.73 ± 21.13	0.13	0.890
Total words	403.85 ± 71.60	384.47 ± 233.97	0.56	0.580
Total sentences	18.47 ± 4.26	20.56 ± 12.26	1.16	0.250

4.2 Comparative Performance of SVM and BERTino

The baseline model combined TF-IDF vectors of the transcriptions with a selected set of linguistic and emotional features to evaluate their effectiveness in distinguishing depressed from non-depressed individuals. Logistic regression was first used to estimate feature importance. This ensured that only the most relevant features were included in the classification model, allowing for more efficient training and a performance outcome that better reflects the underlying data. Among the most influential predictors were *gioia* (joy) and average sentence length, followed by present and past tense usage, with *tristezza* (sadness) and total first person references also contributing meaningfully. Negation frequency had the lowest relative weight in the model. Building on these findings, a support vector machine with a linear kernel was trained using the same input. The model achieved an accuracy of 0.84 in a stratified 80/20 split, with strong precision and recall for both classes (F1 = 0.80 for non-depressed; F1 = 0.87 for depressed). Cross-validation confirmed the model's stability, with a mean accuracy of 0.8427 (± 0.0187) across five folds.

Table 4.3: Feature Importance Chart for SVM model

Feature	**Importance**
gioia	1.251782
avg_sentence_length	1.052469
present_tense_per_sentence	0.744978
past_tense_per_sentence	0.740573
tristezza	0.701813
first_person_total	0.635740
fiducia	0.460252
rabbia	0.235552
negations	0.063246

Table 4.4: Classification Report for SVM baseline model

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	0.89	0.73	0.80	11
1	0.81	0.93	0.87	14
Accuracy		0.84		25
Macro avg	0.85	0.83	0.83	25
Wtd. avg	0.85	0.84	0.84	25

Cross-Validation (5-fold) Accuracy: {0.880, 0.833, 0.833, 0.833, 0.833}
Mean: 0.843 ± 0.019

The transformer based classification model, fine-tuned using the BERTino architecture, demonstrated strong performance in distinguishing between depressed and non-depressed individuals. The training process spanned eight epochs, where an epoch refers to a complete pass through the entire training dataset. During each epoch, the model adjusts its internal parameters based on the errors it makes, gradually improving its predictions with each iteration. Across the eight epochs, both training and validation losses steadily decreased, reflecting effective learning and strong generalization. Specifically, the training loss declined from 0.6995 to 0.2353, while the validation loss dropped from 0.6284 to 0.3295. This consistent improvement across epochs indicates that the model was able to extract meaningful patterns from the text without overfitting. On the held out test set, the model achieved an accuracy of 92%, surpassing the baseline. The macro averaged F1-score reached 0.92, with a perfect recall of 1.00 for the depressed class, highlighting the model's ability to identify all depressed cases. Cross-validation was not applied here due to the high computational cost associated with retraining large transformer models multiple times. Due to the limited dataset and available resources, a stratified 80/20 split was used, and early stopping was applied to prevent overfitting, ensuring efficient and reliable training. The results indicate that transformer models are especially effective at capturing subtle emotional and linguistic cues present in free form narratives.

Table 4.5: Epochs results

Epoch	**Training Loss**	**Validation Loss**
1	0.70	0.63
2	0.60	0.56
3	0.51	0.48
4	0.47	0.42
5	0.32	0.38
6	0.28	0.35
7	0.25	0.34
8	0.24	0.33

Table 4.6: Classification Report for BERTino model

Class	**Precision**	**Recall**	**F1-Score**	**Support**
0	1.00	0.82	0.90	11
1	0.88	1.00	0.93	14
Accuracy		0.92		25
Macro avg	0.94	0.91	0.92	25
Wtd. avg	0.93	0.92	0.92	25

The result obtained here is comparable to the 84.5% accuracy reported by Tao et al. (2020), who used speech-based features extracted from read speech in the "Androids Corpus". Their model relied on paralinguistic markers, notably reading speed and silence duration, which were found to be reliable indicators of depression-related disfluency and slowed processing. These features, just three in total, improved classification accuracy by over 16 percentage points from a 68.2% baseline. In contrast, the present study used a substantially different modality, focusing on lexical, syntactic, and affective cues derived from transcribed language. Despite the absence of prosodic input, the model's performance slightly exceeded that of the speech-based approach. This finding highlights the strength of deep language models in capturing complex linguistic signals associated with depressive cognition and suggests that text alone can support high-accuracy depression classification.

Chapter 5

Discussion

5.1 Findings Overview

This study set out to investigate the role of emotional and cognitive linguistic features in distinguishing between depressed and non-depressed individuals, using both traditional ML and transformers. The baseline model, an SVM which integrated TF-IDF vectors with selected linguistic and emotional features, achieved solid performance, with the classifier reaching 84% accuracy and strong F1-scores across both classes. Notably, feature importance analysis revealed that emotional markers like *gioia* (joy), along with sentence structure and verb tense, were influential predictors. However, BERTino significantly outperformed the baseline, reaching a test accuracy of 92% and a perfect recall for the depressed class. This suggests that while handcrafted linguistic and emotional features are useful, transformer models are better equipped to capture complex, nuanced patterns in natural language that correlate with depression.

5.2 Interpretation of Emotional and Linguistic Features

The comparative evaluation between models reflects both computational trends and theoretical frameworks outlined in the literature. The SVM classifier, trained on TF-IDF representations alongside emotion and cognitive-linguistic features, achieved a respectable accuracy of 84%, with a balanced performance across precision and recall. Feature importance analysis placed *gioia* (joy) and average sentence length among the top predictors, followed by present and past tense usage. While these results demonstrate that handcrafted features rooted in psychological theory can capture salient cues for depression detection, their predictive power remained modest. This finding is consistent with prior research indicating that emotional content, though relevant, may not be sufficient in isolation for distinguishing depressive language. For example, Trifu et al. (2024) showed that in Romanian, a morphologically rich language like Italian, first person singular pronouns were not as salient due to pronoun-drop phenomena. This aligns with the methodology employed, which accounted for both explicit pronouns and first person verb morphology, an approach crucial for languages where self reference can be syntactically implicit. The use of the total number of first person (pronouns + verbs) in this project reflects this linguistic nuance and builds on findings by Spruit et al. (2022) and Lorenzoni et al. (2024), who also found increased use of self referential forms and past tense verbs as indicators of ruminative thinking in depressed individuals. By contrast, BERTino achieved substantially higher performance, with a test accuracy of 92% and perfect recall for identifying depressed transcriptions. This gain highlights the strength of transformer models in capturing complex, context-informed patterns in language, particularly relevant for Italian, where morphology and word order can encode subtle meaning shifts. Contrary to the baseline, BERTino did not rely on predefined features but learned directly from the raw text, leveraging contextual embeddings to detect depressive cues even in the absence of overt linguistic markers. These results echo the growing body of work showing the advantages of transformer models for mental health detection. As shown by Trotzek et al. (2020) and Tejaswini et al. (2022), deep learning architectures outperform traditional classifiers by capturing dependencies, semantic nuance, and affective tone in ways that fixed lexicons like LIWC often miss. Model interpretability was further explored by generating a word cloud based on tokens weighted by attention scores from the most confidently classified depressed transcriptions.

Table 5.1: BERTino's attention: top depressed tokens

Interpretation	**Top Tokens**
Negative affect	*male, niente, non, poche*
Self-reference	*io, mia, mio*
Temporal focus	*anni, settimana, mesi, mattina, domenica, ultimi, trascorso*
Family	*figlia, bambine, mamma, moglie, nonna, padre, marito*
Discourse organization	*allora, quindi, insomma, perché, quando*
Routine	*lavoro, terapia, ospedale, pulizie, cambiamento*

The table above summarizes the most prominent tokens identified by BERTino's attention mechanism in depressed transcripts, grouped by thematic interpretation and derived from the word cloud output. Several patterns emerge that align meaningfully with both linguistic theories and prior findings in depression detection. Notably, the prominence of familial terms such as *figlia* (daughter), *mamma* (mom), *moglie* (wife), *nonna* (grandmother), and *bambine* (little girls) suggests a strong focus on family related matters. This is consistent with previous literature indicating that depressed individuals often center their narratives around close relationships, possibly reflecting a sense of burden, loss, or emotional entanglement within the family sphere. The presence of words like *ospedale* (hospital), *terapia* (therapy), and *schiena* (back) further suggests that health related themes are prevalent in depressed speech, reinforcing psychological research linking depression with somatic concerns and medicalized self perception. Temporal expressions such as *anni* (years), *settimana* (week), *domenica* (Sunday), and *mattina* (morning) highlight an orientation toward time, which may reflect preoccupation with past experiences, both common features of depressive cognition. This aligns with prior studies that identified a strong association between depression and increased use of past tense verbs. Additionally, terms such as *non* (not), *niente* (nothing), and *male* (bad) stand out as high attention negative tokens, aligning with the linguistic profile of depression marked by negation and negative affect. Interestingly, discourse markers like *allora* (so), *quindi* (therefore), and *insomma* (more or less) are also prevalent. These tokens may reflect a more structured or explanatory speaking style, possibly compensating for difficulties in organizing thought or navigating emotionally charged narratives. The frequent appearance of *io* (I), despite Italian being a pro-drop language, further validates the model's sensitivity to self referential language. This could be interpreted as a linguistic manifestation of self focus, a hallmark of depressive cognition. Ultimately, the transformer's superior performance underscores its value for depression detection in Italian, particularly when combined with insights from linguistically informed baseline models. Together, these approaches offer complementary perspectives: while the baseline provides interpretable insights grounded in psychological theory, the transformer captures linguistic complexity that eludes manual feature extraction. The similarity in classification accuracy between the present study and Tao et al. (2020) (84.5%) is particularly notable given the difference in modalities, text versus speech. Tao et al.'s approach captured behavioral markers such as slowed reading and increased silences, which align with known neurocognitive impairments in depression. In contrast, this study focused on language structure and emotional content, highlighting how self referential language and affective tone serve as effective text-based proxies for depressive symptoms. These findings suggest that while speech based models may offer advantages in capturing prosodic markers, text based approaches offer a scalable and less invasive alternative, especially in digital mental health contexts.

Chapter 6

Conclusion

This study examined how depression manifests through language in Italian, adopting both a linguistically informed and computationally advanced approach. Features grounded in cognitive and affective theory were extracted using the ELIta emotion lexicon and syntactic information from the Stanza pipeline, with the goal of distinguishing between depressed and non-depressed speakers. These features informed a baseline SVM model, which performed reliably, but were ultimately surpassed by the BERTino transformer model. Achieving 92% accuracy and perfect recall on the depressed class, the transformer demonstrated a strong ability to capture nuanced linguistic signals. The contrast between models highlights the growing strength of deep learning in identifying psychological states from text, particularly in morphologically complex languages like Italian, where emotional and self referential cues may be subtle or dependent on the context. Notably, this work is the first to analyze depressive language in Italian purely via text and establishes a new benchmark beyond the baseline previously set (Tao et al., 2020) on the same corpus.

6.1 Limitations

A primary limitation of this study is the small size of the dataset (115 transcripts), which may constrain the generalizability of the models. Due to the computational cost of training transformer models, cross-validation was not performed for the BERTino approach, limiting robustness evaluation. Additionally, while deep learning models excel in performance, they lack transparency, making it challenging to interpret the basis of their predictions in clinical contexts. Lastly, although this study focused on lexical and syntactic features, it did not include prosodic or multimodal cues, which are known to play an important role in depression detection.

6.2 Future Directions

Future research should aim to expand the dataset and include more diverse speakers to improve model robustness and generalizability. Cross-validation and ablation studies could further validate feature effectiveness and help disentangle the contribution of different feature groups. Integrating multimodal data, such as voice tone or facial expressions, could also yield more holistic models of depression detection. Finally, developing explainable AI methods such as attention visualization, will be crucial for making these tools ethically usable in real world mental health applications.

Bibliography

Aldarwish, M. M., & Ahmad, H. F. (2017). Predicting depression levels using social media posts. In *Proceedings of the 2017 IEEE 13th International Symposium on Autonomous Decentralized Systems (ISADS)* (pp. 277–280). IEEE.

Al-Mosaiwi, M., & Johnstone, T. (2018). In an absolute state: Elevated use of absolutist words is a marker specific to anxiety, depression, and suicidal ideation. *Clinical Psychological Science, 6*(4), 529–542. https://doi.org/10.1177/2167702617747074

Amanat, A., Rizwan, M., Javed, A. R., Abdelhaq, M., Alsaqour, R., Pandya, S., & Uddin, M. (2022). Deep learning for depression detection from textual data. *Electronics, 11*(5), 676.

American Psychiatric Association. (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). https://doi.org/10.1176/appi.books.9780890425787

Beck, A. T. (1967). *Depression.* Harper and Row.

Burdisso, S. G., Errecalde, M., & Montes-y Gómez, M. (2019). A text classification framework for simple and effective early depression detection over social media streams.

Expert Systems with Applications, 133, 182–197.

Di Palma, E. (2024). ELIta: A new Italian language resource for emotion analysis. In *Proceedings of the 10th Italian Conference on Computational Linguistics (CLiC-it 2024)* (pp. 297–307). Pisa, Italy: CEUR Workshop Proceedings.

Kim, N. H., Kim, J. M., Park, D. M., Ji, S. R., & Kim, J. W. (2022). Analysis of depression in social media texts through the Patient Health Questionnaire-9 and natural language processing. *Digital Health, 8*, 20552076221114204. https://doi.org/10.1177/20552076221114204

Lorenzoni, G., Tavares, C., Nascimento, N., Alencar, P., & Cowan, D. (2024). Assessing ML classification algorithms and NLP techniques for depression detection: An experimental case study. *arXiv*, 2404.04284. https://arxiv.org/abs/2404.04284

Mendula, M., Gabrielli, S., Finazzi, F., Dompe, C., & Delucis, M. (2024). Unveiling mental health insights: A novel NLP tool for stress detection through writing and speaking analysis to prevent burnout. *AHFE International, 122*, 164–174.

Morales, M. R., & Levitan, R. (2016). Speech vs. text: A comparative analysis of features for depression detection systems. In *Proceedings of the 2016 IEEE Spoken Language Technology Workshop (SLT)* (pp. 136–143). IEEE.

Munizza, C., Argentero, P., Coppo, A., Tibaldi, G., Di Giannantonio, M., Picci, R. L., & Rucci, P. (2013). Public beliefs and attitudes towards depression in Italy: A national survey. *PLOS ONE, 8*(5), e63806. https://doi.org/10.1371/journal.pone.0063806

Spruit, M., Verkleij, S., de Schepper, K., & Scheepers, F. (2022). Exploring language markers of mental health in psychiatric stories. *Applied Sciences, 12*(4), 2179.

Tao, F., Esposito, A., & Vinciarelli, A. (2020). Spotting the traces of depression in read speech: An approach based on computational paralinguistics and social signal processing. In *Proceedings of Interspeech 2020* (pp. 1828–1832). https://doi.org/10.21437/Interspeech.2020-2888

Tao, F., Esposito, A., & Vinciarelli, A. (2023). The Androids Corpus: A new publicly available benchmark for speech-based depression detection. In *Proceedings of Interspeech 2023* (pp. 4149–4153). https://doi.org/10.21437/Interspeech.2023-894

Tejaswini, V., Sathya Babu, K., & Sahoo, B. (2024). Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. *ACM Transactions on Asian and Low-Resource Language Information Processing, 23*(1), 1–20.

Tølbøll, K. B. (2019). Linguistic features in depression: A meta-analysis. *Journal of Language Works – Sprogvidenskabeligt Studentertidsskrift, 4*(2), 39.

Trifu, R. N., Nemeș, B., Herta, D. C., Bodea-Hategan, C., Talaș, D. A., & Coman, H. (2024). Linguistic markers for major depressive disorder: A cross-sectional study using an automated procedure. *Frontiers in Psychology, 15*, 1355734.

Trotzek, M., Koitka, S., & Friedrich, C. M. (2020). Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences. *IEEE*

Transactions on Knowledge and Data Engineering, 32(3), 588–601. https://doi.org/10.1109/TKDE.2018.2885515

World Health Organization. (n.d.). Depression. https://www.who.int/news-room/fact-sheets/detail/depression

Zucco, C., Calabrese, B., & Cannataro, M. (2017). Sentiment analysis and affective computing for depression monitoring. In *Proceedings of the 2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)* (pp. 1988–1995). IEEE.