

Article

SpADE-BERT: Multilingual BERT-Based Model with Trigram-Sensitive Tokenization, Tuned for Depression Detection in Spanish Texts

Abdiel Reyes-Vera ^{1,*} , Magdalena Saldana-Perez ¹ , Marco Moreno-Ibarra ¹  and Juan Pablo Francisco Posadas-Durán ² 

¹ Centro de Investigación en Computación (CIC), Instituto Politécnico Nacional (IPN), Mexico City 07738, Mexico; amagdasaldana@cic.ipn.mx (M.S.-P.); marcomoreno@cic.ipn.mx (M.M.-I.)

² Escuela Superior de Ingeniería Mecánica y Eléctrica (ESIME) Zacatenco, Instituto Politécnico Nacional (IPN), Mexico City 07738, Mexico; jposadasd@ipn.mx

* Correspondence: areyesve@ipn.mx

Abstract

This article proposes an automated approach, based on artificial intelligence techniques, for detecting indicators of depression in texts written in Spanish. Among the main contributions is the construction of a new specialized corpus, supervised by mental health professionals and based on the Beck Depression Inventory. Text processing included linguistic techniques such as lemmatization, stopword removal, and structural transformation using trigrams. As part of the work, SpADE-BERT was designed, a model based on multilingual BERT with a tokenization scheme adapted to incorporate trigrams directly from the input phase. This modification allowed for more robust interaction between the local context and semantic representations. SpADE-BERT was evaluated against multiple approaches reported in the literature, which employ algorithms such as logistic regression, support vector machines, decision trees, and Random Forest with advanced configurations and specialized preprocessing. In all cases, our model showed consistently superior performance on metrics such as precision, recall, and F1-score. The results show that integrating deep language models with adapted tokenization strategies can significantly strengthen the automated identification of linguistic signals associated with depression in Spanish texts.

Keywords: depression; natural language processing; multilingual BERT; trigram-based tokenization; transfer learning; Spanish clinical corpus



Academic Editor: Olufemi A. Omitaomu

Received: 14 October 2025

Revised: 2 January 2026

Accepted: 10 January 2026

Published: 1 February 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

1. Introduction

Natural language processing (NLP) and artificial intelligence (AI) have become increasingly relevant in the medical field, particularly as tools to support analysis, prediction, and clinical classification tasks. In the field of mental health, these technologies make it possible to detect linguistic patterns that may be associated with disorders such as depression, anxiety, or impulsivity. Techniques such as machine learning (ML) and deep learning (DL) have proven effective in the automatic classification of texts, thus facilitating the early identification of emotional or cognitive signals that may require professional attention [1–3].

Recent studies have analyzed social media posts, personal diaries, and open-ended questionnaire responses, demonstrating that written language can reflect relevant psychological states. This opens up the possibility of using spontaneous texts as a complementary source of information for screening clinical conditions in a non-invasive and scalable way.

However, a significant limitation of this approach is the availability of specific corpora in languages other than English. Most of the available datasets are constructed from Anglo-Saxon texts, which makes it difficult to generalize models to other languages. For this reason, it is essential to develop linguistic resources adapted to specific contexts, such as Spanish, that allow the full potential of NLP to be exploited in clinical applications. One of the most widely used tools for assessing depression is the Beck Depression Inventory (BDI), developed by psychologist Aaron T. Beck. This clinical instrument consists of 21 items covering emotional, cognitive, and somatic symptoms associated with depressive disorder. For each item, the patient must select the statement that best describes how they have felt during the last two weeks, thus generating a total score that allows the severity of depression to be classified into levels such as mild, moderate, or severe [4]. The BDI has demonstrated high validity and reliability in various clinical and research studies, establishing itself as a benchmark in the assessment of depression [4]. However, it should be emphasized that it does not constitute a medical diagnosis on its own, but rather an assessment tool that must be complemented by clinical interviews and other professional evaluations [5].

Although its usefulness in individual contexts is widely recognized, its widespread application has significant limitations. It requires adequate privacy, reading comprehension on the part of the patient, professional supervision, and sufficient time for analysis and interpretation. These requirements make it difficult to implement in settings with limited human resources or where a large number of people need to be assessed in a short period of time. This problem has encouraged the search for technological alternatives that allow for the automated and effective detection of possible signs of depression, without replacing clinical diagnosis, but facilitating initial screening in broader population contexts. Given the limitations inherent in the widespread application of the Beck Depression Inventory, this study proposes an automated tool based on artificial intelligence, aimed at the early detection of linguistic signs associated with depression in texts written in Spanish. Far from seeking to replace standardized clinical instruments, the proposal aims to offer a complementary screening mechanism that facilitates the preliminary identification of possible cases in contexts where direct psychological evaluation is not feasible [4,5].

The model developed is based on the multilingual BERT (mBERT) architecture, adapted using a specific preprocessing scheme that transforms texts into sequences of trigrams. This transformation allows the capture of relevant local semantic and syntactic patterns and is integrated into the model's tokenizer to optimize its performance in the classification task. Previous models have demonstrated the effectiveness of BERT and its variants in clinical text analysis and depressive symptom detection tasks [3,6], but they do not address the direct modification of the tokenizer or the explicit integration of trigrams as proposed in this study. In addition, a specialized corpus of Spanish texts related to depression was constructed, supervised and validated by mental health professionals, ensuring greater fidelity between the textual content and the associated clinical labels [7,8]. Such expert validation is rare in other datasets, especially in Spanish [9]. The results obtained show that the proposed model outperforms various traditional classifiers in metrics such as accuracy and F1-score, suggesting its viability as a support tool for automated screening [10–12]. This proposal represents a concrete contribution to the development of accessible and culturally adapted technologies that strengthen early diagnosis and mental health care. Various studies have explored alternative approaches to depression detection using machine learning and text analysis techniques, with promising results. Models such as Random Forest [13], AdaBoost [14,15], and K-Nearest Neighbors [16] have been used in classification tasks on texts extracted from social networks or clinical interviews. Likewise, feature extraction methods such as TF-IDF, Doc2Vec, and entropy measures have been applied to represent the texts before training [10,17]. In more recent studies, hybrid

strategies combining neural models with classical classifiers have been used, as in the case of BETO+SVM [18]. However, most of these studies have been developed on English-language corpora and lack direct validation by mental health specialists, which limits their applicability in real clinical contexts and in languages other than English. This research seeks to contribute to this gap with a clinically validated technical proposal designed specifically for the Spanish language.

2. State of the Art (Background or Related Work)

This section describes some of the work related to text analysis using machine learning and deep learning algorithms. As mentioned above, there is growing interest in developing an objective method for identifying and measuring the level of depression using artificial intelligence (AI) [19]. It also presents some datasets generated specifically for analyzing depression.

With the aim of assessing and analyzing the progress made in identifying mental disorders between 2017 and 2023, Figure 1 was created based on the information presented in [9]. The mental illnesses that have been detected using text analysis appear in the first column, which is linked to the second column describing the origin of the text analyzed, mainly from social media and interviews. Subsequently, the third column describes the languages in which the identification of mental illnesses was carried out and, finally, the fourth column shows the AI approach used to analyze the texts. The thickness of the lines indicates the prevalence of each combination. For Figure 1, 450 articles aimed at identifying mental disorders were considered, of which 130 focused on the detection of depression. Analyzing texts in English, 75 studies use ML techniques and 57 use Deep Learning, in contrast to 14 that use ML and 8 that use DL in Spanish. It is important to mention that in Figure 1, it can be clearly seen that the languages in which mental illness analysis has been most widely conducted are English and Chinese, while a small portion was conducted in Spanish. This reflects the fact that currently, there are few resources available in Spanish, which are necessary to analyze texts considering the characteristics of the language.

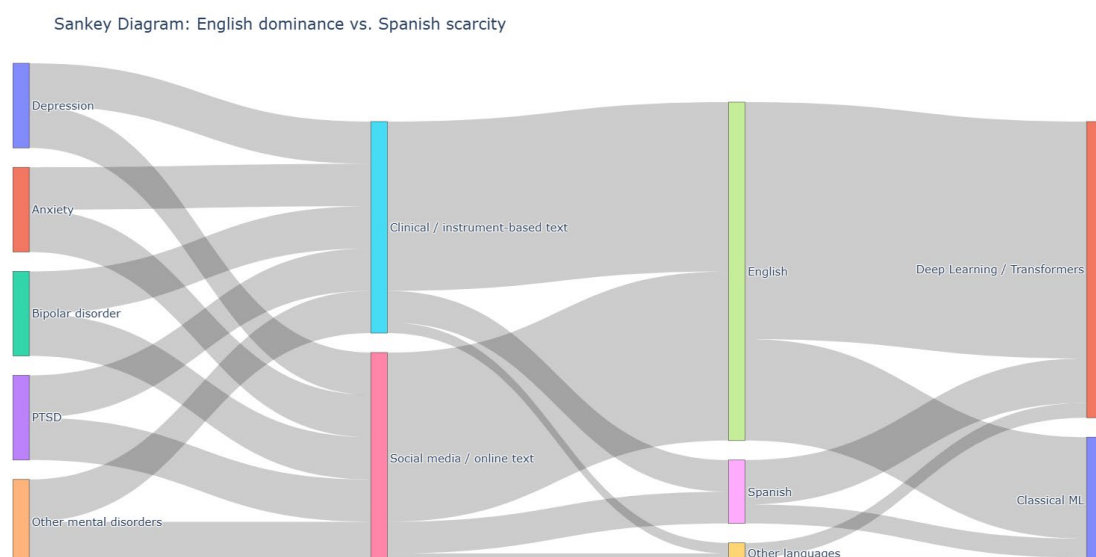


Figure 1. Sankey diagram, where each flow connects the analyzed aspects of the reviewed works, including Sector 1: mental disorder, Sector 2: origin of the samples, Sector 3: language, and Sector 4: AI technique, while the thickness between the sectors indicates the volume of each combination.

2.1. Construction and Use of Corpora for the Study of Depression

One of the challenges in studying depression through NLP is the collection of relevant data. In this context, Losada et al. [7] propose the creation of a dataset extracted from the social network Reddit, in which users with depression were identified by detecting self-reported phrases. Up to 1000 messages per user were collected, ultimately selecting 137 users with depression and 755 as a control group. In addition, this study introduced the task of “early prediction detection,” a metric that measures the time it takes for a system to validate a decision, which has influenced the development of more efficient models.

The importance of this dataset has transcended internationally, becoming a key reference in the eRisk competition, which since 2017 has promoted research in mental risk detection through PLN. Studies such as “Overview of MentalRiskES at IberLEF 2023” [8] have explored various strategies applied in eRisk, including models based on pre-trained neural networks, hybrid models, and conventional classifiers such as Random Forest, Logistic Regression, and K-Nearest Neighbors (K-NN).

Other studies have analyzed different approaches to data collection. Jiayi and Zhaowey [18] developed a web crawler to analyze the perception of depression in online communities, allowing psychologists to better understand users’ experiences and provide more effective interventions. Similarly, Tobin et al. [17] investigated the use of words in reports by young people with depression, applying frequency analysis, clustering, and keyword detection in textual corpora.

2.2. Machine Learning Algorithms in Depression Detection

Various machine learning models have been used to classify depression based on text. The most relevant are presented below:

Logistic Regression: It has been widely used in the classification of clinical texts, showing significant improvements when combined with vectorization techniques such as TF-IDF and Word2Vec [10–12].

Random Forest: Its ability to handle structurally complex and noisy data—that is, information containing outliers, inconsistencies, or irrelevant features that can hinder classification—has made it a benchmark model in the analysis of mental disorders. In recent studies [13,20,21], it has been combined with preprocessing and vector representation techniques to improve its performance in detecting depression.

Support Vector Machines (SVM): These have been used in combination with neural networks to improve the classification of clinical texts. For example, in [22], BETO, a BERT-based model, was used to extract textual representations and feed an SVM as the final classifier.

AdaBoost: Used in multiple studies on mental illness detection [14–16], its adaptive learning capability has shown good results on nonlinear data.

K-Nearest Neighbors (K-NN): Its approach based on the distance between instances has proven useful in the classification of psychological disorders, especially when combined with dimensionality reduction techniques [23–25].

Decision Trees: These have been widely used in the detection of symptoms of depression, integrated into ensemble methods such as Random Forest and Gradient Boosting to improve model accuracy [26–28].

A notable study in this area is “Exploring the Performance of Baseline Text Mining Frameworks for Early Prediction of Self-Harm Over Social Media” [10], which evaluated several classifiers in the eRisk corpus. This work analyzed models such as SVM, Random Forest, and Logistic Regression, combining them with feature extraction techniques such as Doc2Vec, TF-IDF, and entropy measures. It was found that the best accuracy was achieved with SVM and entropy-based features, achieving an F1 score of 0.89.

2.3. Pre-Trained Models and Multilingual Transfer in Depression Classification

With the advancement of deep learning models, the use of pre-trained models has become highly relevant in text classification [29]. One of the most promising approaches is the use of BERT (Bidirectional Encoder Representations from Transformers) and its multilingual variants [6,30,31].

In particular, Multilingual BERT (mBERT) has been widely studied due to its ability to handle texts in multiple languages. Unlike BERT, which was trained exclusively in English with Wikipedia and BooksCorpus corpora, mBERT was trained with texts from 104 languages, giving it significant cross-language transferability (zero-shot learning).

Recent studies [32] have shown that mBERT has improved performance in text classification, sentiment analysis, and entity detection tasks in multiple languages, surpassing traditional supervised learning methods. In architectural terms, mBERT has 768 hidden units, 12 transformer blocks, and 12 self-attention heads, optimizing the contextual representation of language.

In this work, we chose to use mBERT, given that the dataset included text in several languages, which required a model with multilingual capabilities. The tests performed confirmed that the use of mBERT with the AdamW optimizer significantly improved classification accuracy, thanks to its ability to capture deep semantic relationships in multiple languages.

3. Methodology

In this article, we propose the following three-stage methodology: corpus construction, preprocessing, and model creation. In the first stage, Corpus Creation, we will detail the steps followed for its construction, such as searching for associations that work with people suffering from depression to collect information, applying a test endorsed by psychologists, and, finally, collecting a written statement from people diagnosed with depression. In the second stage of preprocessing, we apply three techniques to the samples: stopword removal, lemmatization, and n-grams. Together, these techniques allow us to avoid redundancy in the text information. Finally, in the model creation stage, we used a combination of two elements: preprocessing and transfer learning. This allowed us to leverage the knowledge previously acquired by the model and refine it using preprocessing techniques to improve its performance.

3.1. Corpus Construction

The corpus construction process is shown in Figure 2. It is important to mention that we had the advice of psychologists at every step, who guided us on issues related to psychology, the management of participants in the sampling process, and the appropriate treatment of individuals.

Sampling was carried out through an individualized process (see Figure 3), in which each participant was evaluated by professional psychologists under controlled conditions using two components: (1) they wrote a free-form text response to the question “*How are you feeling today?*”, in which they described their mood, and (2) they completed the *Beck Depression Inventory*, developed by psychologist Aaron T. Beck [33].

The Beck Depression Inventory has been widely used and has proven its validity and effectiveness in numerous clinical and research studies [4]. It was chosen for its carefully designed structure, together with its multidimensional approach, which allows for a comprehensive assessment of depressive symptoms, including their cognitive, emotional, and physical aspects. When collecting samples, the confidentiality and privacy of each participant’s data was guaranteed. The Beck Depression Inventory consists of five levels. Levels 1 and 2 were considered non-depressive because, according to the test, these levels

indicate that the person does not need psychological help. In contrast, levels 3 to 5 are considered to require therapy and are therefore considered to be depressive.

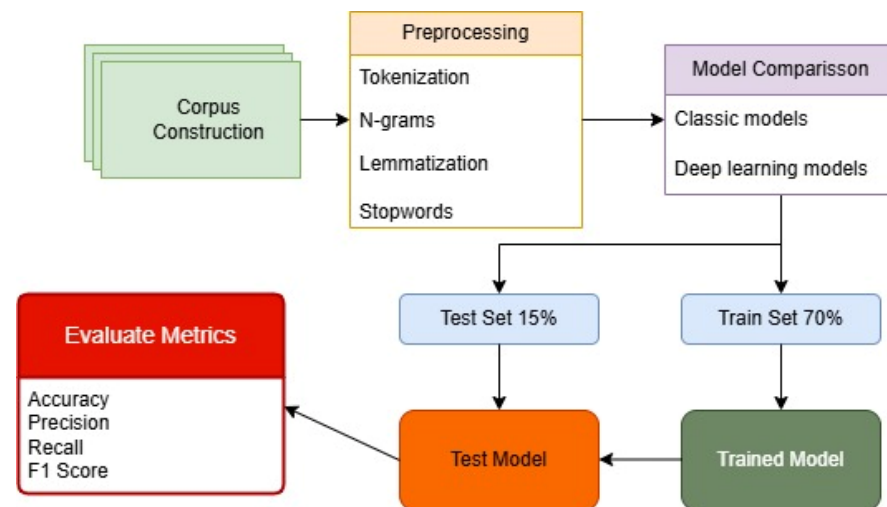


Figure 2. Diagram of the methodology used.

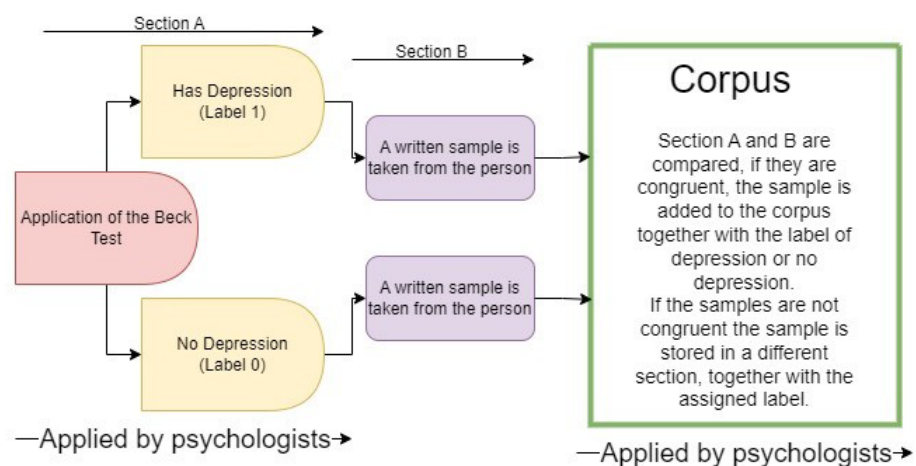


Figure 3. Corpus creation process.

The Beck Test assessment can be defined as follows:

Let P be a function that determines the degree of depression

$$P = \sum_{i=1}^n x_i$$

where

- P is the total score obtained on the test.
- x_i is the score for the i -th response (with values typically ranging from 0 to 3, depending on the severity of the reported symptoms).
- n is the total number of items in the test (21).

The range of P is from 0 to 63, given that there are 21 questions and each one can have a maximum value of 3 points. Thus, the degrees of depression are identified by classifying the values of P as follows:

- If $P \in [0, 13]$, then Level of depression = Minimal absence or mild depression
- If $P \in [14, 19]$, then Level of depression = Mild to moderate depression
- If $P \in [20, 28]$, then Level of depression = Moderate to severe depression

- If $P \in [29, 63]$, then Level of depression = Severe depression

The texts are analyzed by a psychologist, who evaluates the congruence between the Beck test and the “free text.” In this process, the expert reviews the text associated with each case and compares its content with the result obtained in the Beck Inventory. If both results are consistent, the text is labeled and stored in the main section of the corpus. If there is a discrepancy between the expert’s assessment and the test result, the text is labeled and stored in a secondary section of the corpus for further analysis.

The above can be defined as follows:

Let T be the free text response to the question “How are you feeling today?”.

Let P be the Beck Depression Test score associated with the individual.

Let $E(T, P)$ be the assessment of congruence made by the expert psychologist between the text T and the score P . Then:

- $E(T, P) = 1$, $E(T, P) = 1$ if the text and the score are congruent.
- $E(T, P) = 0$, $E(T, P) = 0$ if the text and score are incongruent.

If the result of the evaluation is congruent ($E(T, P) = 1$), then the text T is labeled as “congruent” and stored in the main section of the corpus.

If the evaluation result is inconsistent ($E(T, P) = 0$), then the text T is labeled as “inconsistent” and stored in the secondary section of the corpus.

The corpus can be consulted at the following link, see Figure 4.



Figure 4. Link to the dataset.

3.2. Preprocessing

During the text preprocessing phase, several techniques are applied to process and improve the quality of the collected data. These techniques include tokenization, lemmatization, stopword removal, and the use of n-grams. Below, we show the process by which words are received in this phase (see Figure 5).

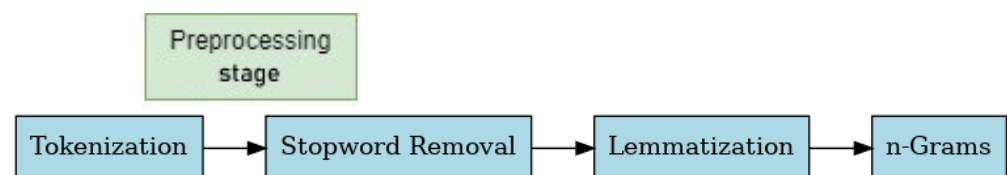


Figure 5. Stages of preprocessing.

Tokenization is the process of segmenting a text into smaller linguistic units, commonly referred to as *tokens*. These units can be words, characters, or subwords, depending on the approach taken. In natural language processing (NLP) tasks, tokenization is a fundamental step that allows unstructured text to be converted into a computationally manipulable sequence. This transformation facilitates statistical and semantic analysis, as it allows the occurrence, co-occurrence, or position of significant elements within the text to be identified. For this work, word-oriented tokenization was used, whereby each sentence

was divided into individual words. This segmentation prepares the text for subsequent stages of cleaning, lemmatization, and pattern analysis, such as n-gram generation. This process can be formally represented as follows:

- Let D be a natural language document composed of a sequence of characters.
- Let $\text{tokenize}(D)$ be a segmentation function that transforms D into a sequence of tokens.
- The resulting sequence is denoted as:

$$T = \text{tokenize}(D) = \{w_1, w_2, \dots, w_n\},$$

where w_i represents the i -th token of the document and n is the total number of tokens generated.

Stopwords are common and frequently used words in natural language, such as articles, pronouns, and conjunctions, which do not contribute significant semantic value to the analysis and can introduce noise into the data [34]. Therefore, to reduce the influence of irrelevant words in the analysis, stopwords are removed. The process consists of identifying and removing stopwords from the text, for which linguistic resources such as predefined lists of stopwords in Spanish were used. Preserving the most relevant and significant terms for the detection of depression, words such as “the,” “of,” “and,” “but,” “because,” and “this” are removed, as they do not contribute useful information to the analysis. This process is formalized as follows:

- Let $T = \{w_1, w_2, \dots, w_n\}$ be the original text, where w_i is the i -th word in the text, and n is the total number of words.
- Let S be the set of *stopwords*, i.e., common and insignificant words.

The text filtered by removing the stopwords T_{filtr} is the text T without the words belonging to the set S , as follows:

$$T_{\text{filtr}} = \{w_i \in T \mid w_i \notin S\}$$

Lemmatization is a linguistic process whose objective is to reduce words to their base form or lemma, so that the different forms of the same word are represented in a single form, in this case its lemma [35]. This technique is especially useful for normalizing vocabulary and grouping different variants of words into a unified representation. For example, words such as “running,” “ran,” and “runs” would be reduced to their base form “run.” This facilitates analysis by reducing redundancy and ensuring that different word variants are considered the same entity. The lemmatized text $T_{(\text{lem})}$ is the filtered text $T_{(\text{filtr})}$ where each word $w_{(i)}$ has been replaced by its lemma

$$T_{\text{lem}} = \{L(w_i) \mid w_i \in T_{\text{filtr}}\}$$

Therefore, $L(w_i)$ returns the lemmatized version of the word w_i .

Subsequently, a fundamental technique is used: n-grams, which are sequences of adjacent words in a text, and the number of segments represents the length of each n-gram. This technique has proven effective in capturing word combinations that could have a greater impact on the detection of depression [36]. N-grams of length 3 (trigrams) were used to improve performance with the dataset used, since capturing sequences of three consecutive words captures more complex contextual relationships in the text. This allowed for a more accurate analysis of the context and coherence of the expressions used in the Spanish documents [37]. To obtain the n-grams of size $k = 3$, we applied the function $N_{(k)}(T_{(\text{lem})})$, which returns a set of contiguous sequences of k words from the lemmatized text $T_{(\text{lem})}$:

$$N_k(T_{\text{lem}}) = \{(w_i, w_{i+1}, w_{i+2}) \mid w_i \in T_{\text{lem}} \text{ and } 1 \leq i \leq n - 2\}$$

At the end of this stage, we obtain the processed text, free of *stopwords*, with all words reduced to their base form through lemmatization and finally organized into strings of three words (*trigrams*). This allows us to standardize the text, achieving several benefits: the elimination of *stopwords* reduces noise by removing words without relevant meaning; lemmatization decreases linguistic variability by grouping words with the same root; and the use of *n-grams* facilitates the detection of key combinations within the text. With standardized texts, the model that is implemented will be able to identify patterns more easily.

Together, these techniques allow us to generalize the data and reduce its dimensionality, which improves efficiency in subsequent processes. In addition, by applying machine learning algorithms, these will be able to detect word sequences in short sentences with greater accuracy, thus optimizing the identification of relevant patterns in text analysis.

3.3. Model Creation

In our study, we reviewed several algorithms to address the task of classifying Spanish texts, with a special focus on detecting depression [2–5,7,13,14,18,20,28,38–45]. We considered classic classification methods such as support vector machines (SVM) [22,46,47], random forests [13,20,21], and logistic regression [10–12]. These algorithms have proven effective in various natural language processing (NLP) applications. However, after analyzing their capabilities and limitations, we decided to explore more advanced techniques, particularly neural networks, due to their ability to handle complex data and their high performance in classification tasks. Within this category, models based on *Transformer* architectures, widely adopted in the state of the art, stand out, as well as the use of *transfer learning* strategies, which allow pre-trained models on large corpora to be reused and effectively adapted to specific tasks with reduced volumes of annotated data [6,29,31,32,48–54].

In the field of deep learning, several techniques have been experimentally tested, including transfer learning, which allows the model to leverage previously acquired knowledge, thus facilitating the task of classifying information.

SpADE-BERT, Spanish Analysis for Depression with Enhanced BERT

During the development of our model, we identified that, when applying transfer learning, it was necessary to use the specific tokenizer corresponding to each pre-trained model. In this process, we experimented with different models and analyzed the performance of their tokenizers. Although some accepted the text resulting from our preprocessing, we considered the possibility of adjusting the tokenizer to better integrate with this processing and thus improve the model's performance.

The pre-trained models used consist of four stages: Input Embeddings, Positional Encoding, stacked encoders, and the encoder output. At each of these stages, we made adjustments to optimize the model's analysis. In the Input Embeddings stage, each word is assigned a position within the text, and that position is represented by a vector that allows contextual relationships between nearby words to be captured.

To align this stage with our preprocessing, we decided to modify the way these relationships are represented: instead of analyzing them word by word, they were grouped into strings of three words (*trigrams*). This decision aims to ensure that the model learns semantic and syntactic patterns consistent with the approach adopted in preprocessing. This modification is illustrated in Figure 6.

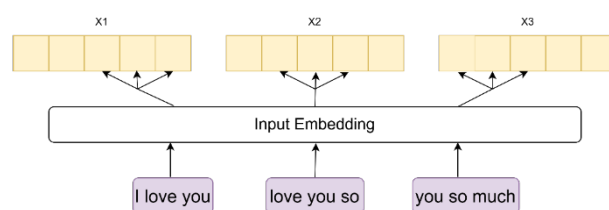


Figure 6. Input Embeddings.

In the second step, Positional Encoding, we seek to identify the order or position of words within the text. To achieve this, a signal is added to each word, in our case to each trigram, which allows its position to be encoded. This signal is generated using sine and cosine functions, which makes it possible to represent trigrams as vectors. These vectors reflect the specific position of each trigram within the sequence, making it easier for the model to detect relationships between them. The values generated by these functions range from -1 to 1 , which helps the model understand the relative location of each trigram within the text, as illustrated in Figure 7.

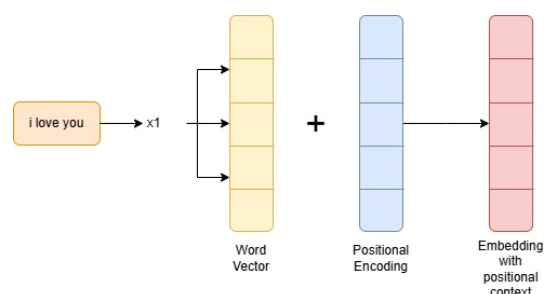


Figure 7. Positional Encoding.

In the third step, corresponding to the *stacked encoder layers*, multiple layers are applied that allow the model to process the trigrams in an increasingly deeper way. Each of these layers includes an attention mechanism that allows the model to focus on different parts of the text to identify contextual relationships between trigrams, as well as a neural network that transforms the information to generate more complex representations. By working with trigrams rather than individual words, these layers help capture key combinations of terms and language patterns that, in this particular case, could be associated with symptoms of depression. The accumulation of these layers allows the model to generalize better, understand the emotional context of the text, and learn to detect subtle signals throughout the processed sequence, see Figure 8.

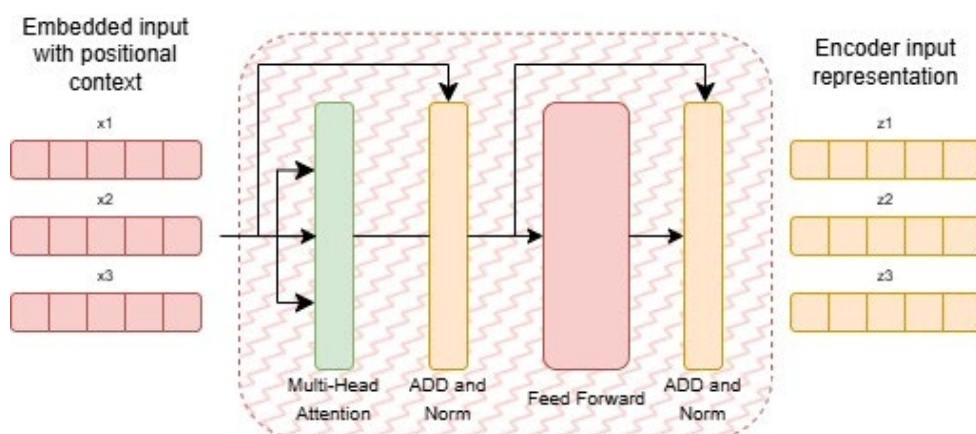


Figure 8. Stacked Encoder Layers.

No substantial modification was made to the Multi-Head Attention layer, as the trigram previously processed during the previous phases is used at this stage. From this point on, the model follows the classic Transformer architecture. The next step is the Add & Norm layer, where the information is normalized and the residual attention output is added. This operation helps stabilize training and mitigate problems such as gradient vanishing.

A particularly important layer in our case is the Feed Forward layer. After applying attention to each trigram, this layer transforms the information locally, i.e., it processes each trigram independently to generate a richer representation. Thanks to this nonlinear transformation, the model is able to capture more abstract and complex patterns that may be associated with typical word combinations in depressive expressions. Thus, not only is the global context provided by attention leveraged, but the semantic value of each trigram is reinforced, allowing the model to refine its ability to detect linguistic signals related to depression.

The last stage of our model corresponds to the encoder's output block, which provides the output embeddings resulting from all the processing previously applied to the trigrams. These embeddings contain deep representations that combine semantic, contextual, and structural information acquired throughout the model's multiple layers. Thanks to the attention mechanisms and feed forward layers, each trigram has been transformed into an enriched vector that reflects both its individual meaning and its relationship to the rest of the text.

Throughout this process, positional encoding has made it possible to preserve the order of appearance of the trigrams, which is essential for maintaining the coherence of the discourse and capturing linguistic nuances that may be key in detecting signs of depression. This encoding ensures that the model focuses not only on isolated words, but also on how they are combined and in what context they appear.

Once the encoder has generated the final representations, they are transferred to a stack of decoder layers, which interpret these embeddings and transform them into specific outputs. In our case, the expected output is a binary classification related to the detection of indicators of depression in the text.

To do this, a linear classifier is used, a fully connected layer that takes the vectors generated by the encoder as input and projects them into a lower-dimensional space, where each component represents a possible class. In this work, we define classes as text with signs of depression and text without signs of depression. This classifier acts as a bridge between the complex numerical representation that the model has learned and the practical interpretation we need for clinical-linguistic analysis.

Finally, a Softmax function is applied, which converts the classifier values into a probability distribution. This allows us not only to identify the most probable class, but also to know the model's degree of confidence in each prediction. This information can be particularly useful in sensitive contexts such as mental health, where not only the prediction itself matters, but also the confidence with which it is made.

This design, as a whole, allows the model to process Spanish texts robustly, taking into account key word combinations (trigrams), and offer informed predictions to support the analysis of possible cases of depression. Furthermore, by adjusting the architecture to work directly with three-word units, we aligned the entire model structure with our initial preprocessing, ensuring consistency and improving the overall effectiveness of the system. Figure 9 illustrates the internal structure of SpaDE-Bert.

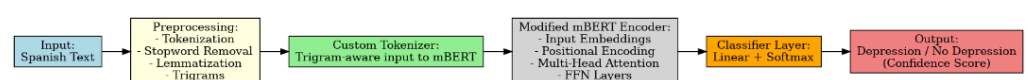


Figure 9. SpaDE-Bert Operation.

4. Experiments

In order to evaluate the performance of our proposed model, SpADE-BERT, several comparative experiments were conducted using a selection of classification algorithms widely reported in the literature. These algorithms include logistic regression [10–12], support vector machines (SVM) [22,46,47], Random Forest [13,20,21], AdaBoost [14–16], K-Nearest Neighbors (KNN) [23–25], and decision trees [26–28].

The choice of these models was not limited solely to the name of the algorithm, but also considered robust configurations described in the literature, including the use of traditional vector representations such as *Bag of Words*, and weighting schemes such as TF-IDF, as well as additional techniques such as pre-trained word embeddings (Word2Vec, GloVe), frequency-based attribute selection, dimensional reduction using PCA, and lexical normalization. These configurations were carefully replicated based on the referenced studies, prioritizing those that shared experimental methodologies similar to the proposal of this work.

To ensure the validity of the comparison, studies were selected whose experiments used methodological configurations similar to those used in this study, replicating these approaches under controlled conditions. In particular, two datasets were used: our clinical corpus in Spanish, and the standardized corpus by Losada et al. [7], commonly used in tasks of detecting depression from text. This dual-implementation allowed us to observe the performance of each algorithm under different data sources but with a homogeneous experimental methodology, which facilitates a fair comparison between both models and corpora.

In order to analyze the impact of the proposed preprocessing—particularly the generation of trigrams—experiments were conducted with each of the selected algorithms using two configurations: one with complete linguistic preprocessing (tokenization, stopword removal, lemmatization, and trigram formation), and another without applying this scheme. This comparison allowed us to observe the direct effect of preprocessing on the performance of each classifier. In several cases, there was a substantial improvement in metrics such as accuracy and F1-score when preprocessing was applied, suggesting that a richer semantic representation brings advantages even to models not based on neural networks.

The results presented in Table 1 correspond to this series of replicated experiments on our corpus, using both configurations with and without complete preprocessing, in order to evaluate the behavior of the classifiers under uniform conditions.

Table 1. Performance comparison of classical machine learning algorithms under different linguistic preprocessing configurations.

Algorithm	With Linguistic Preprocessing (Tokenization, Stopword Removal, Lemmatization, Trigrams)		Without Linguistic Preprocessing (Trigrams Only)	
	Accuracy	F1 Score	Accuracy	F1 Score
Logistic regression	0.87	0.87	0.76	0.77
Random forest	0.64	0.67	0.70	0.70
SVM	0.84	0.84	0.75	0.74
Adaboost	0.70	0.68	0.63	0.64
Nearest neighbor	0.67	0.72	0.63	0.59
Decision tree	0.81	0.76	0.75	0.78

Table 2 presents a comparison between the classic algorithms replicated in this work—such as logistic regression, SVM, Random Forest, among others—run on our clinical corpus and on the corpus of Losada et al. [7]. This comparison allows us to observe the behavior of each classifier under similar conditions but with different data sources. To

contrast these results, experiments reported in representative studies in the field were reviewed, such as the article, *A natural language processing-based risk prediction framework for pathological gambling* [12], in which several of these same algorithms were evaluated using the F1 metric. Although there are numerous additional studies, those that allow for consistent methodological comparisons were selected. It should be noted that the F1-score metric was prioritized in this comparison because it is the most commonly reported metric in the reviewed articles, which facilitates an objective analysis between approaches.

Table 2. Second experiments, compared to the state of the art.

Algorithm	Algorithms in Our Dataset	Algorithms in the Losada Dataset
	F1 Score	F1 Score
Logistic regression	0.87	0.67
Random forest	0.67	0.65
SVM	0.84	0.76
Adaboost	0.68	0.71
Nearest Neighbor	0.72	0.33
Decision tree	0.76	0.57

The results presented in Table 3 correspond to the experiments carried out with pre-trained neural network models, applied after the preprocessing described above. When analyzing the most prominent state-of-the-art proposals in depression detection using NLP, it was identified that a large proportion of the most recent works are based on Transformer-type architectures, particularly variants of the BERT model, due to their ability to capture deep contextual dependencies in language. Based on this, several relevant pre-trained models were selected: BERT-base, as it is the most influential and widely adopted model in NLP tasks; BETO, as it is specifically trained on Spanish corpora; DistilBERT, as it offers a lighter and more efficient version of BERT without losing too much performance; and mBERT, for its multilingual approach, useful for verifying the generalization of results in contexts with linguistic variability.

Table 3. Experiments with transformer models.

Parameter	Options	Multilingual Berth	Berth	Beto	Distilberth	Spade-Bert
Lot size	[16, 32, 64]	16	32	32	64	16
Maximum sequence	[256, 512]	256	256	512	512	256
Learning rate	$[2 \times 10^{-5}, 3 \times 10^{-5}]$	2.00×10^{-5}	3.00×10^{-5}	2.00×10^{-5}	3.00×10^{-5}	2.00×10^{-5}
Period	[3, 4, 5]	5	3	4	5	3
Optimizer	[Adam, Adamw]	Adamw	Adam	Adam	Adamw	Adamw
Epsilon	$[1 \times 10^{-8}, \text{not applicable}]$	1.00×10^{-8}	not applicable	not applicable	1.00×10^{-8}	1.00×10^{-8}
Accuracy with preprocessing	--	0.8642	0.8312	0.8411	0.6477	0.9512
Accuracy without pre-treatment	--	0.7731	0.6254	0.7913	0.7645	0.8423

Given the widespread use of models based on Transformer architectures in natural language processing tasks, additional experiments were conducted using pre-trained models such as BERT-base, BETO, DistilBERT, and mBERT to establish a benchmark comparison. However, the central proposal of this work is SpADE-BERT, a model based on mBERT that incorporates a modified tokenization strategy to directly integrate trigrams into the model

input. This modification seeks to better capture the linguistic sequences associated with indicators of depression in texts written in Spanish.

The results of these experiments are presented in Table 3, where it can be seen that SpADE-BERT consistently outperforms the other models evaluated, achieving an accuracy of 95.12% with complete preprocessing. It should be noted that, in the case of SpADE-BERT, the “no preprocessing” option retains the trigram-based tokenization integrated directly into the model, but omits the previous linguistic processing steps such as stopword removal and lemmatization. This allows the impact of these external techniques to be isolated without removing the internal mechanism proposed by the model.

In order to clearly visualize the performance of the evaluated models, Figure 10 shows a direct comparison of the accuracy values obtained by each of them using complete preprocessing. The graph highlights that SpADE-BERT, the model proposed in this work, achieves the highest accuracy (0.9512), surpassing both traditional models and pre-trained variants of BERT. Even without external preprocessing, SpADE-BERT maintains competitive performance (0.8423), demonstrating that the incorporation of trigrams as part of the internal tokenizer contributes significantly to its classification ability. These results empirically validate the effectiveness of combining adapted linguistic representations with deep architectures for the detection of depressive symptoms in Spanish-language texts.

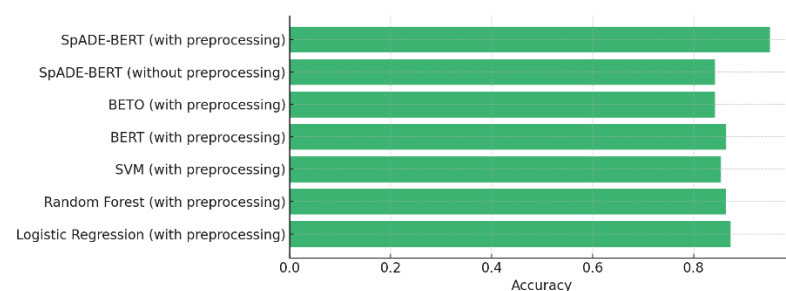


Figure 10. Comparison between models.

After observing the overall performance of the evaluated models, it is pertinent to analyze in greater detail the metrics obtained specifically by the proposed SpADE-BERT model, given that it performed best in most of the experiments conducted. Table 4 below summarizes the precision, recall, and F1-score values obtained by SpADE-BERT with and without complete linguistic preprocessing. This analysis allows us to quantify the real impact of preprocessing techniques on model performance and to delve deeper into the robustness of its behavior under different configurations.

Table 4. SpADE-Bert.

Model	Accuracy	Precision	Recall	F1-Score
SpADE-BERT (with pre-processing)	0.9515	0.9424	0.9477	0.945
SpADE-BERT (without preprocessing)	0.8423	0.8911	0.8836	0.8873

The results presented in Table 4 allow us to observe more clearly the effect of preprocessing on the performance of the SpADE-BERT model. When the complete preprocessing configuration is applied—including tokenization, lemmatization, stopword removal, and trigram formation—the model achieves an accuracy of 95.12%, along with high and balanced values in precision (94.24%), recall (94.77%), and F1-score (94.5%). These metrics

reflect not only the accuracy of the model, but also its ability to correctly identify positive cases without overfitting.

On the other hand, even without applying external preprocessing techniques, SpADE-BERT maintains competitive performance, with an F1-score of 88.73%, indicating that the proposed architecture—particularly its trigram-based tokenization integrated into the model—allows it to retain much of its discriminatory power. This robustness suggests that the model can adapt to real-world situations where text may not be structured or clean, which is especially useful in clinical contexts and applications in uncontrolled environments.

5. Conclusions

A new corpus was presented, compiled with the collaboration of psychologists and validated by them, ensuring its reliability and robustness as a resource for research in the field of depression detection. In general, the scarcity of available datasets in Spanish underscores the importance of creating and making datasets in this language available to the community. It was found that many of the most advanced datasets with the desired characteristics were unavailable or inaccessible. Attempts to contact the authors were unsuccessful. This highlights the need for Spanish-language data for research in this field. This paper presented an architecture based on Transformer models called SpADE-BERT, which incorporates a trigram-sensitive tokenization scheme specifically adapted for Spanish-language texts. As a fundamental part of this proposal, a new clinical corpus supervised by psychologists was developed, structured from free responses and scores from the Beck Depression Inventory. This dual validation, both automatic and professional, strengthens the reliability of the dataset and allows for progress in the detection of linguistic indicators of depression with a clinically relevant approach.

The experiments carried out showed that SpADE-BERT consistently outperforms widely used classical models such as SVM, Random Forest, and logistic regression, both in our corpus and in previously established sets such as that of Losada et al. Likewise, the proposed model obtained better results compared to pre-trained variants such as BERT-base, DistilBERT, BETO, and mBERT, reaching an accuracy of 0.9512, along with balanced metrics in precision (0.9424), recall (0.9477), and F1-score (0.945). Even without external preprocessing, the model maintained robust performance, highlighting the effectiveness of trigram integration from the tokenizer.

In addition to the performance achieved, this work shows that knowledge transfer from multilingual models such as mBERT to specific clinical domains in Spanish can be achieved effectively, even with limited resources. This approach not only takes advantage of prior massive training, but also facilitates specialization through curated and technically designed corpora. Unlike previous approaches that use only word-based representations or conventional embeddings, this work shows that a modification to the tokenizer can improve the capture of complex and relevant linguistic structures in clinical contexts. This modification represents a key methodological contribution to natural language processing in Spanish. Due to its accuracy, robustness, and adaptability, **SpADE-BERT** could be integrated into mental health-oriented technological tools, such as automatic screening systems or conversational assistants that operate on written Spanish. This positions it as a promising solution to support early detection processes in clinical and non-clinical settings.

Finally, it is recognized that, although the results are promising, it is necessary to validate this architecture on other Spanish-language datasets and evaluate its performance in uncontrolled environments. In the future, we plan to extend this line of work to texts from social networks and digital platforms, as well as integrate it into clinical diagnostic support tools.

Author Contributions: Conceptualization, A.R.-V. and J.P.F.P.-D.; methodology, A.R.-V.; software, A.R.-V. and J.P.F.P.-D.; validation, A.R.-V., M.S.-P. and J.P.F.P.-D.; formal analysis, A.R.-V., M.S.-P. and M.M.-I.; investigation, A.R.-V., M.S.-P. and J.P.F.P.-D.; resources, A.R.-V. and M.S.-P.; data curation, A.R.-V.; writing—original draft preparation, A.R.-V. and M.M.-I.; writing—review and editing, A.R.-V., M.S.-P., M.M.-I. and J.P.F.P.-D.; visualization, A.R.-V. and M.M.-I.; supervision, M.S.-P., M.M.-I. and J.P.F.P.-D.; project administration, A.R.-V., M.S.-P., M.M.-I. and J.P.F.P.-D.; funding acquisition, A.R.-V. All authors have read and agreed to the published version of the manuscript.

Funding: The authors has received funding from the SIP-IPN under grants 20250037, SECTEI/084/2023 and CM-SECTEI/040/2025.

Institutional Review Board Statement: Ethical review and approval were waived for this study because it involved the analysis of voluntarily provided textual data, without any clinical intervention, experimentation, or collection of sensitive personal identifiers. All data were anonymized prior to analysis, and informed consent was obtained from all participants.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data supporting the findings of this study are publicly available in an anonymized form to ensure participants' privacy and can be accessed at: https://github.com/abrey21/Depresion_dataset_spanish (accessed on 5 January 2026).

Acknowledgments: The authors would like to thank the Centro de Investigación en Computación (CIC) of the Instituto Politécnico Nacional (IPN) for the institutional support provided during the development of this research. Abdiel Reyes-Vera acknowledges the support of a doctoral scholarship from CONAHCYT (Consejo Nacional de Humanidades, Ciencias y Tecnologías, Mexico). The authors also thank the anonymous reviewers and the Academic Editor for their constructive comments, which significantly improved the clarity and quality of this manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Achary, R.; Shelke, C.J.; Shrivastava, V.K.; Paul, P.M.; Konda, S.; Billa, M. Depression detection based on NLP and ML techniques using text and speech recognition. In Proceedings of the International Conference on Trends in Sustainable Computing and Machine Intelligence, Bangkok, Thailand, 5–6 October 2023; Springer: Berlin/Heidelberg, Germany, 2023; pp. 357–372.
2. Nusrat, M.O.; Shahzad, W.; Jamal, S.A. Multi-class depression detection through tweets using artificial intelligence. *arXiv* **2024**, arXiv:2404.13104.
3. Tejaswini, V.; Sathya Babu, K.; Sahoo, B. Depression detection from social media text analysis using natural language processing techniques and hybrid deep learning model. *ACM Trans. Asian Low-Resour. Lang. Inf. Process.* **2024**, *23*, 1–20. [CrossRef]
4. Saberi, S.; Ahmadi, R.; Khakpoor, S.; Pirzeh, R.; Hasani, M.; Moradveisi, L.; Saed, O. Comparing the effectiveness of behavioral activation in group vs. self-help format for reducing depression, repetitive thoughts, and enhancing performance of patients with major depressive disorder: A randomized clinical trial. *BMC Psychiatry* **2024**, *24*, 516. [CrossRef]
5. Hajduska-Dér, B.; Kiss, G.; Sztahó, D.; Vicsi, K.; Simon, L. The applicability of the Beck Depression Inventory and the Hamilton Depression Scale in the automatic recognition of depression based on speech signal processing. *Front. Psychiatry* **2022**, *13*, 879896. [CrossRef]
6. Rothman, D. *Transformers for Natural Language Processing: Build, Train, and Fine-Tune Deep Neural Network Architectures for NLP with Python, Hugging Face, and OpenAI's GPT-3, ChatGPT, and GPT-4*; Packt Publishing Ltd.: Mumbai, India, 2022.
7. Losada, D.E.; Crestani, F. A collection of tests for research on depression and language use. In Proceedings of the International Conference of the Cross-Language Evaluation Forum for European Languages, Portorož, Slovenia, 23–28 May 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 28–39.
8. Mármol-Romero, A.M.; Moreno-Muñoz, A.; Plaza-del Arco, F.M.; Molina-González, M.D.; Martín-Valdivia, M.T.; Ureña-López, L.A.; Montejo-Raéz, A. Overview of mental risks in Iberlef 2023: Early detection of the risk of mental disorders in Spanish. *Nat. Lang. Process.* **2023**, *71*, 329–350.
9. Zhang, T.; Schoene, A.M.; Ji, S.; Ananiadou, S. Natural language processing applied to mental illness detection: A narrative review. *NPJ Digit. Med.* **2022**, *5*, 46. [CrossRef]
10. Basu, T.; Gkoutos, G.V. Exploring the performance of baseline text mining frameworks for early prediction of self-harm over social media. In Proceedings of the CLEF (Working Notes), Bucharest, Romania, 21–24 September 2021; pp. 928–937.

11. Lopes, R.P. Cedri at risk 2021: A naive approach for early detection of psycho-disorders on social media. In Proceedings of the CEUR Workshop Proceedings, College Station, TX, USA, 19–20 August 2021; pp. 981–991.
12. Talha, A.; Basu, T. A natural language processing-based risk prediction framework for pathological gambling. In Proceedings of the CLEF Working Notes, Thessaloniki, Greece, 18–21 September 2023; pp. 18–21.
13. Alhuzali, H.; Zhang, T.; Ananiadou, S. Predicting sign of depression via using frozen pre-trained models and random forest classifier. In Proceedings of the CLEF (Working Notes), Bucharest, Romania, 21–24 September 2021; pp. 888–896.
14. Gallegos Salazar, L.M.; Loyola-González, O.; Medina-Pérez, M.A. An explainable approach based on emotion and sentiment characteristics to detect people with mental disorders on social media. *Appl. Sci.* **2021**, *11*, 10932. [\[CrossRef\]](#)
15. Shahraki, A.; Abbasi, M.; Haugen, Ø. Boosting algorithms for network intrusion detection: A comparative evaluation of real adaboost, soft adaboost, and modest adaboost. *Artif. Intell. Eng. Appl.* **2020**, *94*, 103770. [\[CrossRef\]](#)
16. Wang, W.; Sun, D. The improved adaboost algorithms for imbalanced data classification. *Inf. Sci.* **2021**, *563*, 358–374. [\[CrossRef\]](#)
17. Tobin, C. Gender Identity Diversity in Individuals with Attention Deficit/Hyperactivity Disorder (ADHD). *Ment. Health Matters* **2023**, *10*, 12–13.
18. Shi, J.; Khoo, Z. Words for the hearts: A corpus study of metaphors in online depression communities. *Front. Psychol.* **2023**, *14*, 1227123. [\[CrossRef\]](#)
19. Sun, J.; Dong, Q.-X.; Wang, S.-W.; Zheng, Y.-B.; Liu, X.-X.; Lu, T.-S.; Yuan, K.; Shi, J.; Hu, B.; Lu, L.; et al. Artificial intelligence in psychiatry research, diagnosis, and therapy. *Asian J. Psychiatry* **2023**, *87*, 103705. [\[CrossRef\]](#)
20. Hacohe-Kerner, Y.; Manor, N.; Goldmeier, M.; Bachar, E. Detection of anorexic girls in blog posts written in Hebrew using a combined heuristic AI and NLP method. *IEEE Access* **2022**, *10*, 34800–34814. [\[CrossRef\]](#)
21. López-Vizcaino, M.; Nóvoa, F.J.; Fernández, D.; Cacheda, F. Time aware f-score for the evaluation of early detection of cybersecurity. *Appl. Sci.* **2024**, *14*, 574. [\[CrossRef\]](#)
22. Smys, S.; Raj, J.S. Analysis of deep learning techniques for early detection of depression on social media networks—A comparative study. *J. Trends Comput. Sci. Smart Technol.* **2021**, *3*, 24–39.
23. Aguilera, J.; Farías, D.I.H.; Ortega-Mendoza, R.M.; Montes-y Gómez, M. Detection of depression and anorexia in social networks as a one-class classification problem. *Appl. Intell.* **2021**, *51*, 6088–6103. [\[CrossRef\]](#)
24. Nedunchezian, P.; Mahalingam, M. Improving motivation for recovery from depression recommendation system (i-drmrs) in online social networks. *SN Inform.* **2022**, *3*, 166.
25. Xiong, L.; Yao, Y. Study on an adaptive thermal comfort model with k-nearest-neighbors (knn) algorithm. *Constr. Environ.* **2021**, *202*, 108026. [\[CrossRef\]](#)
26. AlHamed, F.; Ive, J.; Specia, L. Predicting moments of mood changes over time from imbalanced social media data. In Proceedings of the Eighth Workshop on Computational Linguistics and Clinical Psychology, Online, 15 July 2022; pp. 239–244.
27. Charbuty, B.; Abdulazeez, A. Classification based on the decision tree algorithm for machine learning. *J. Appl. Sci. Technol. Trends* **2021**, *2*, 20–28. [\[CrossRef\]](#)
28. de Jesús Titla-Tlatelpa, J.; Ortega-Mendoza, R.M.; Montes-y Gómez, M.; Villaseñor-Pineda, L. A profile-based sentiment-aware approach for depression detection in social media. *EPJ Data Sci.* **2021**, *10*, 54. [\[CrossRef\]](#)
29. Min, B.; Ross, H.; Sulem, E.; Veyseh, A.P.; Nguyen, T.H.; Sainz, O.; Agirre, E.; Heintz, I.; Roth, D. Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Comput. Surv.* **2023**, *56*, 1–40. [\[CrossRef\]](#)
30. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *arXiv* **2023**, arXiv:1706.03762. [\[PubMed\]](#)
31. Koroteev, M.V. BERT: A review of applications in natural language processing and understanding. *arXiv* **2021**, arXiv:2103.11943. [\[CrossRef\]](#)
32. Deode, S.; Gadre, J.; Kajale, A.; Joshi, A.; Joshi, R. L3Cube-IndicSBERT: A simple approach for learning cross-lingual sentence representations using multilingual BERT. *arXiv* **2023**, arXiv:2304.11434.
33. Beck, A.T. *Depression: Causes and Treatment*; University of Pennsylvania Press: Philadelphia, PA, USA, 1972.
34. Gupta, M.; Kumar, M. Employing NLP Model for Language Detector (Comprehensive Review). 2024. SSRN 4947447. Available online: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4947447 (accessed on 9 January 2026).
35. Pant, V.K.; Sharma, R.; Kundu, S. An overview of Stemming and Lemmatization Techniques. In *Advances in Networks, Intelligence and Computing*; CRC Press: Boca Raton, FL, USA, 2024; pp. 308–321.
36. del Valle, E.; de la Fuente, L. Sentiment analysis methods for politics and hate speech contents in Spanish language: A systematic review. *IEEE Lat. Am. Trans.* **2023**, *21*, 408–418. [\[CrossRef\]](#)
37. Ikegami, Y.; Tsuruta, S.; Kutics, A.; Damiani, E.; Knauf, R. Fast ML-based next-word prediction for hybrid languages. *Internet Things* **2024**, *25*, 101064. [\[CrossRef\]](#)
38. Li, Y.; Yang, X.; Zhao, M.; Wang, Z.; Yao, Y.; Qian, W.; Qi, S. Fpt-former: A flexible parallel transformer for depression recognition using multimodal measures based on audiovisual expert knowledge. *Int. J. Intell. Syst.* **2024**, *2024*, 1564574. [\[CrossRef\]](#)

39. Shah M, N.; Kumar, R.; Sathianathan, R.; Gupta, M.; Manjula, S.; Kumar M, K. Clinicians' views on depression and the role of escitalopram in the treatment of depression. *Int. J. Neuropsychiatr. Dis.* **2024**, *21*, 61–67. [\[CrossRef\]](#)
40. Valbona, G.; Zhubi, E.; Bytyci, P.; Malazogu, E.; Hoxha, R. The use of the Beck Depression Inventory for the assessment of depressive symptoms in epilepsy: A single-center experience in Kosovo. *Ital. J. Med.* **2024**, *18*, 1703. [\[CrossRef\]](#)
41. Voros, V.; Fekete, S.; Tenyi, T.; Rihmer, Z.; Szili, I.; Osvath, P. Untreated depressive symptoms significantly worsen quality of life in old age and may lead to the misdiagnosis of dementia: A cross-sectional study. *Ann. Gen. Psychiatry* **2020**, *19*, 52. [\[CrossRef\]](#) [\[PubMed\]](#)
42. Zhang, Y.; Liao, H.; Gu, J.; Wang, J. Anxiety and depression related to child maltreatment in adolescents: Comparison of the multiple individual risk model, the cumulative risk model, and latent profile analysis. *Child Abuse Neglect.* **2022**, *128*, 105630. [\[CrossRef\]](#) [\[PubMed\]](#)
43. Cuesta Mosquera, E.L.; Picón Rodríguez, J.P.; Pineida Parra, P.M. Current trends in depression, risk factors, and substance abuse. *J. Am. Health* **2022**, *5*, 114. Available online: <https://www.jah-journal.com/index.php/jah/article/view/114> (accessed on 9 January 2026).
44. Vázquez-Salas, A.; Hubert, C.; Portillo-Romero, A.; Valdez-Santiago, R.; Barrientos-Gutiérrez, T.; Villalobos, A. Depressive symptoms in Mexican adolescents and adults: Ensanut 2022. *Public Health Mex.* **2023**, *65*, s117–s125.
45. Lee, B. National, state-level, and county-level prevalence estimates of adults aged ≥ 18 years self-reporting a lifetime diagnosis of depression—United States, 2020. *MMWR. Morb. Mortal. Wkly. Rep.* **2023**, *72*, 644–650. [\[CrossRef\]](#)
46. López-Vizcaíno, M.F.; Nóvoa, F.J.; Carneiro, V.; Cacheda, F. Early detection of cyberbullying on social media networks. *Future Gener. Comput. Syst.* **2021**, *118*, 219–229. [\[CrossRef\]](#)
47. Zhang, T.; Yang, K.; Ji, S.; Ananiadou, S. Emotion fusion for mental illness detection from social media: A survey. *Inf. Fusion* **2023**, *92*, 231–246. [\[CrossRef\]](#)
48. Turchin, A.; Masharsky, S.; Zitnik, M. Comparison of bert implementations for natural language processing of narrative medical documents. *Inform. Med. Desbloqueada* **2023**, *36*, 101139. [\[CrossRef\]](#)
49. Bab, N.B.; Rahman, R.N.A.R.A.; Mohamed, S.; Radzi, N.A.M.; Yusof, N. Effectiveness of oral health interventions among children and adolescents with mental disorders: A systematic review. *Spec. Care Dent.* **2024**, *44*, 1002–1025. [\[CrossRef\]](#)
50. Spring, R.; Johnson, M. The possibility of improving the automated calculation of lexical richness measures for EFL writing: A comparison of the LCA, NLTK, and Spacy tools. *System* **2022**, *106*, 102770. [\[CrossRef\]](#)
51. Broche-Pérez, Y.; Fernández-Castillo, E.; Reyes Luzardo, D.A. Psychological consequences of quarantine and social isolation during the COVID-19 pandemic. *Cuba. J. Public Health* **2021**, *46*, e2488.
52. Bermúdez Sabel, H. Natural language processing and text fixation. Experiences in the constitution of a diachronic corpus of sonnets. In *Editar el Siglo de Oro en la Era Digital*; Universitat Autònoma de Barcelona: Barcelona, Spain, 2024; pp. 161–174. Available online: <https://ddd.uab.cat/record/292186> (accessed on 9 January 2026).
53. Riofrio, C.H.E.; Chóez, A.C.; Gamboa, J.Z. Methods for extracting comments from the social network Twitter for use in Natural Language Processing. *Polo Conoc. Rev. Científico-Prof.* **2021**, *6*, 104–123.
54. Freitas, C.; Souza, E.; Castro, M.C.; Cavalcanti, T.; da Silva, P.F.; Cordeiro, F.C. Linguistic resources for domain-specific NLP: Petrolês. *Linguamática* **2023**, *15*, 51–68. [\[CrossRef\]](#)

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.