Computer Science Review 53 (2024) 100654

Contents lists available at ScienceDirect

Computer Science Review

journal homepage: www.elsevier.com/locate/cosrev

Review article

A survey on detecting mental disorders with natural language processing: Literature review, trends and challenges

Arturo Montejo-Ráez [a,*], M. Dolores Molina-González [a], Salud María Jiménez-Zafra [a], Miguel Ángel García-Cumbreras [a], Luis Joaquín García-López [b]

[a] CEATIC. Department of Computer Science, Universidad de Jaén Campus Las Lagunillas, E-23071, Jaén, Spain
[b] Department of Psychology, Universidad de Jaén Campus Las Lagunillas, E-23071, Jaén, Spain

ARTICLE INFO

Keywords:
Mental disorders detection
Natural language processing
Machine learning
Survey

ABSTRACT

For years, the scientific community has researched monitoring approaches for the detection of certain mental disorders and risky behaviors, like depression, eating disorders, gambling, and suicidal ideation among others, in order to activate prevention or mitigation strategies and, in severe cases, clinical treatment. Natural Language Processing is one of the most active disciplines dealing with the automatic detection of mental disorders. This paper offers a comprehensive and extensive review of research works on Natural Language Processing applied to the identification of some mental disorders. To this end, we have identified from a literature review, which are the main types of features used to represent the texts, the machine learning algorithms that are preferred or the most targeted social media platforms, among other aspects. Besides, the paper reports on scientific forums and projects focused on the automatic detection of these problems over the most popular social networks. Thus, this compilation provides a broad view of the matter, summarizing main strategies, and significant findings, but, also, recognizing some of the weaknesses in the research works published so far, serving as clues for future research.

Contents

1. Introduction 2
2. Methodology 2
   2.1. Relevance of selected disorders 2
   2.2. Review considerations 3
   2.3. Approaches characterization 4
   2.4. Evaluation campaigns 4
3. Findings 5
   3.1. Datasets 5
   3.2. Algorithms 5
   3.3. Remarks on most interesting works 6
   3.4. Projects 9
4. Discussion 10
5. Conclusions 12
   Declaration of competing interest 12
   Data availability 12
   Acknowledgments 12
   Appendix A. Acronyms used for algoritms 13
   Appendix B. Summary tables of the review 13
   References 13

* Corresponding author.
E-mail addresses: amontejo@ujaen.es (A. Montejo-Ráez), mdmolina@ujaen.es (M.D. Molina-González), sjzafra@ujaen.es (S.M. Jiménez-Zafra), magc@ujaen.es (M.Á. García-Cumbreras), ljgarcia@ujaen.es (L.J. García-López).

https://doi.org/10.1016/j.cosrev.2024.100654
Received 30 January 2024; Received in revised form 31 May 2024; Accepted 13 June 2024
Available online 22 June 2024

1. Introduction

In the digital world, personal details and daily interactions with others are continuously exposed in web based social networks. Children and adolescents have been identified by World Health Organization (WHO) as being at particular risk of psychological distress [1]. Widespread access to the Internet and the use of Information and Communication Technologies (ICTs), including mobile devices, are generating important changes in interpersonal relationships and communication that are more evident among young people and adolescents [2,3]. Harassment, abuse, manipulation, extortion, bullying, incitement to suicide, social pressure, offensive messages... all these attacks have been leveraged by the new forms of social interaction promoted by the Internet and are boosting depression and anxiety, which are increasing faster than ever, with rising suicide rates. Suicide is likely to become a more pressing concern after the COVID-19 pandemic and has longer-term effects on the general population, the economy, and vulnerable groups [4]. Preventing suicide therefore needs urgent consideration by addressing emotional problems. Compared to estimates before the epidemic, reports already indicate a 3-fold increase in the prevalence of experience of mood and anxiety swings [5,6].

Online platforms can be considered the most widely used mass media today. These platforms offer the ideal anonymous channel to expose one's own feelings remaining immune and safe, among other facts. The digital 2023 report elaborated by We Are Social and Hootsuite [7] reveals that there are 4.76 billion social media users, equivalent to 59.4 percent of the world's population. The social network users worldwide have grown by 3 percent, with 137 million new users starting on social networks during 2022. On average, users spend about 2 h and 30 min per day connected to social media. These statistics reflect the growing use of social media and the importance of studying them due to the large number of users connected, and the time spent on them.

Human language is the main transmission medium involved in social interaction, also over these platforms, and it is time for current revolutionary algorithms in Natural Language Processing (NLP) to take part in this scenario and provide means to prevent and predict risky interactions, protecting the most fragile members of our digital societies. Human Language Technologies can help us to build more confident environments. Thanks to NLP, artificial intelligence solutions are able to model human language and use learned models to extract information and understand the meaning of text flowing through social networks. The combination of deep learning algorithms, which have shown impressive performances in many text related tasks, with linguistic resources and tools, enable the construction of monitoring systems for the early detection of signs of misbehavior like eating disorders, depression, bullying or suicide tendencies over social media. The main research question tackled by the studies under our focus is if NLP solutions are capable of detecting or estimating the risk of suffering mental disorders by processing textual data found in social media or other digital platforms.

The aim of this work is to provide an updated insight into the interdisciplinary research that is being performed to analyze social media communications in order to identify several disorders like eating disorders, depression, suicidal ideation and gambling addiction. The rise of these problems, leveraged by social networks on the Internet, have attracted the interest of governments and joined the efforts of researchers from disciplines like Psychology or Computer Science. This overview provides a wide revision of the state of the art in applying artificial intelligence for detecting these problems from social interactions in several known media like Facebook,[1] Reddit,[2] Twitter[3] or Instagram,[4] among others. Major approaches, scientific forums and evaluation campaigns are also covered to allow novel research in the area to rapidly catch up to the subject.

As a contribution from the analysis, we identify some challenges for future research: the need for further research on languages beyond English, to raise the quality of datasets with annotation by experts in the field, to explore multimodal data (text combined with visual information), explainability, and applicability in real world scenarios.

The rest of the paper is organized as follows: Section 2 describes the methodology followed in this review of the literature by selecting the disorders analyzed (Section 2.1), defining the objectives of the review and the strategy for compiling the references (Section 2.2), establishing the key aspects to be identified in the reviewed papers (Section 2.3), and enumerating the main scientific forums on the matter (Section 2.4). Next, in Section 3, it is provided the result of the review and analysis performed by highlighting our major findings, which has been structured by, first, describing how datasets are created and annotated (Section 3.1), then, by summarizing algorithms and methods found (Section 3.2). Then, within the same section, a more detailed description on the most interesting works, according to our criteria, is provided at Section 3.3. Section 3 ends with a summary of some projects found during our search of related works (Section 3.4). Section 4 is devoted to discussing findings and other appreciations collected during our review and related to data sources, features used, approaches and performances, confidence of these systems and, of course, their applicability in real world scenarios. Some brief conclusions serve as ending words to this paper in Section 5.

2. Methodology

To provide a clear and rational meta-analysis of the subject, we have performed a bibliographic review attending to most of the PRISMA recommendations [8]. Beyond the bibliographic review, we have also included a revision on related projects and real-world applications, as we find these aspects also relevant to shape a clear sight on detecting mental disorders in social media with NLP. The framework wherein the review fits in is composed of the following phases:

1. **Selection of disorders**. We have selected four disorders: depression, eating disorders, suicidal ideation and pathological gambling. This disorders are introduced in Section 2.1 to provide both a justification of its relevance and a better understanding of them.
2. **Bibliographic compilation process**. At this step, the search of relevant works is performed. In Section 2.2 we provide the details on how the references have been collected (search queries, bibliographics databases, curation criteria and so on).
3. **Characterization**. In this phase, after the preliminary review of the papers selected, a characterization scheme was defined. This is detailed in Section 2.3. From this scheme we have organized, for each disorder, the related papers, so the general overview on the matter can be formed and discussed.

2.1. Relevance of selected disorders

As stated before, social media plays an important role in mental health. They have become one of the main means of communication, and the large amount of information shared in them is a very valuable source of data to detect risks, disorders and attacks. Here are the ones that have raised major attention due to their prevalence and comorbidity:

Depression According to the Global Burden of Disease 2019 study[5] depression is a common mental disorder worldwide, with almost

[1] https://www.facebook.com/.
[2] https://www.reddit.com/.
[3] https://twitter.com.
[4] https://www.instagram.com.
[5] https://vizhub.healthdata.org/gbd-results/.

22% of the population from 15 to 19 years old under risk. Depression is different from usual mood fluctuations and short-lived emotional responses to challenges in everyday life. It can cause the affected person to suffer greatly and function poorly at work, at school and in the family. WHO [9] stated that during a depressive episode, the person experiences significant difficulty in personal, family, social, educational, occupational, and/or other important areas of functioning. At its worst, depression can lead to suicide. Depression is estimated to occur among 1.1% of adolescents aged 10–14 years, and 2.8% of 15-19-year-olds. Social withdrawal can exacerbate isolation and loneliness and it seems that internet addiction has a lot to do with it [10].

Eating Disorder According to American Psychiatric Association [11], many children and adolescents worry about their eating, weight, or body shape. But for some of them, these worries can lead to unhealthy eating or dieting, known as eating disorders. It is important to get children with eating disorders early evidence-based treatment because many disorders can lead to serious, even life-threatening medical problems. The Society of Clinical Child and Adolescent Psychology have stressed that eating behaviors under eating problems are usually harmful ways to control body shape or weight, and cause the child to view his or her own body negatively. There are a wide variety of eating disorders, including: anorexia nervosa, avoidant/restrictive food intake disorder, binge eating disorder, bulimia nervosa, pica and rumination disorder. As apparent health and wellness becomes a cultural focus, social networks have leveraged these problems [12,13].

Suicidal ideation Over 700,000 people die due to suicide every year. Suicide is one of the leading causes of death in young people [14]. Every suicide is a tragedy that affects families, communities and entire countries and has long-lasting effects on the people left behind. Recent studies by the Spanish National Institute of Statistics have shown that suicide is the leading cause of death in Spanish young people, with one death per day. The link between suicide and mental disorders (in particular, depression and alcohol use disorders) is well established. Experiencing conflict, disaster, violence, abuse, or loss and a sense of isolation are strongly associated with suicidal behavior. Suicide rates are also high amongst vulnerable groups who experience discrimination, such as refugees and migrants; indigenous peoples; lesbian, gay, bisexual, transgender, intersex (LGBTI) persons. By far, the strongest risk factor for suicide is a previous suicide attempt. Actually, the World Health Organization recognizes suicide as a public health priority.

Pathological gambling In addition to substance-related addictions, pathological gambling is also included, reflecting evidence that gambling behaviors activate reward systems similar to those activated by drugs, as they produce some behavioral symptoms similar to substance use disorders like mood disorders, panic disorder, obsessive-compulsive disorder, agoraphobia, substance use problems (which may precede or accompany gambling behavior) and suicidal attempts [15]. Pathological gambling also known as ludomania is the impulse to gamble regardless of its negative consequences [16]. Some signs of gambling addiction are: frequency of playing, rise of losses, compulsory behavior and the consumption of alcohol. One of the most characteristic symptoms of gambling addiction is that the gambler gambles to try to recoup losses (known as "chasing" one's losses), which, of course, aggravates the problem [17].

2.2. Review considerations

We have considered those studies that tackle detection of mental disorders by applying NLP techniques. Therefore, our interest is on those approaches that try to detect the presence of a mental disorder in a subject in the base of his/her written texts, usually posts and comments in social media platforms but also from clinical texts and online psychological platforms. Social media platforms have been found to be a major cause of mental illness in teenagers, specially in girls [18]. Regarding our review, there are similar reviews with a wider number of records checked [19], though our overview varies in the way we categorize main aspects of proposals found and further information regarding evaluation campaigns, projects and real-world applications. Besides, we provide more up to date references and consider pathological gambling, a disorder rarely covered in previous studies. The advances in NLP due to deep learning models in the last two years are relevant enough to complement previous revisions.

The objectives that our review aim to address are the following:

1. To understand how these mental disorders are tackled by the NLP community mainly over social networks.
2. To characterize the approaches by identify relevant aspects to enable a comparison among systems and proposals.
3. To define some categorical dimensions to group reviewed works according to relevant classification criteria.
4. To provide an updated state-of-the-art in NLP techniques for the detection of mental disorders.
5. To identify gaps, lack of research, major findings, main strategies and pending issues on those aspects by works up to date, so guidelines for further work can be proposed.

In the computer science environment, the most commonly used scientific and bibliographic databases are Web of Science (WoS), Scopus and Google Scholar. PubMED is rarely used because it is a free search engine that mainly accesses the MEDLINE medical database, whose data come mainly from private medical records and not from posts in social networks or online platforms. For this reason, we discarded PubMED.

After choosing WoS, Scopus and Google Scholar databases and interacting with their interface, we found that Google Scholar could not be searched in the abstract, and considering that the title and abstract contain the most relevant information of the paper, we also decided to discard it. The abstract should briefly state the objective of the research, the main results and the main conclusions. Therefore, the papers reviewed must include in their abstract the use of Natural Language Processing (NLP) or, otherwise, some algorithm implemented in the research. As can be seen in the glossary Appendix A, there is a huge diversity of algorithms, although most of them can be classified as machine or deep learning approaches.

In our first attempt to choose between the two remaining databases, we filtered by title with the keyword "depression" and abstract with "Natural Language Processing OR machine learning OR deep learning" found more papers in the Scopus database than in WoS. After this, we tried with the keyword "anorexia" in the title and in the abstract the same keywords as in the previous case, and so on, with "suicide" and "pathological gambling" and there was also a greater number of results in the Scopus database. For this reason, we decided to perform the final search in the Scopus scientific database. The search strategy consisted of an deep retrieval of the Scopus scientific database from its earliest records until April 2023.

Observing the high number of papers found for the keywords "depression" and "suicide", and spending some time reading the titles of some papers we found, we thought it would be advisable to limit the search with some additional keywords to better match the intent of our research since many of those papers were not related to the field we wanted.

Table 1
Keywords included and excluded in search queries.

	Included keywords	Excluded keywords
Title	`detection OR prevention OR recognition OR prediction OR identification OR detecting OR preventing OR recognizing OR predicting OR identifying OR detect OR prevent OR predict OR identify`	`survey AND review AND study AND EEG AND facial AND audio AND speech AND imaging AND multimodal AND voice AND phone`
Abstract	`NLP OR Natural Language Processing OR machine learning OR deep learning`	

Table 2
Number of papers found by disorder and year.

Mental disorder	Total	2023	2022	2021	2020	2019 and prior
Depression	249	10	84	50	37	48
Anorexia	14	0	0	3	4	7
Suicidal ideation	186	13	50	46	30	47
Pathological gambling	10	1	8	1	0	0

Although some notable advances have been made using EEG and different types of data such as facial or voice recognition through imaging, speech or audio for the detection of mental disorders, the data required for the application of NLP techniques are textual. Also, other medical papers make use of the phone as a receiver of information to track patients with mental disorders, but these ones are outside the scope of our survey. Besides, we did not want our paper to be influenced by other studies, surveys or reviews available to date in the literature. For these reasons, words related to the above are excluded from the search query.

Obviously, the search terms where mainly those related to the disorders targeted (depression, anorexia, pathological gambling and suicidal ideation). Also, for each disorder and to construct the search query, we included and excluded some words in the title and in the abstract fields, as can be seen in Table 1.

According to the keywords defined, a different query was used for each disorder, so related papers were retrieved:

1. `TITLE (depression AND (<included keywords>) AND NOT (<excluded keywords>) AND ABS (<included keywords>))`
2. `TITLE ((anorexia or eating) AND (<included keywords>) AND NOT (<excluded keywords>) AND ABS (<included keywords>))`
3. `TITLE ((suicidal OR harm-self OR suicide) AND (<included keywords>) AND NOT (<excluded keywords>) AND ABS (<included keywords>))`
4. `TITLE (gambling AND (<included keywords>) AND NOT (<excluded keywords>) AND ABS (<included keywords>))`

In Table 2, we show the number of papers found by disorder and year in any type of document.

Due to the large number of papers found for depression and suicidal ideation, we have reduced the number to 50 reviewed papers (5 most novel papers from 2023 and 20, 15 and 10 most cited papers from 2022, 2021 and 2020, respectively). We have reviewed all the papers found on anorexia and pathological gambling.

2.3. Approaches characterization

From the works revised, several key aspects were identified to characterize different approaches, to ease the comparison between systems and disorders, to evidence the rise of certain algorithms, identify the most studied sources and, with all this, to realize certain imbalance in the way these problems are been overcome by the research community. The aspects considered are described in Table 3.

Table 3
Characterization keys for the categorization of examined works.

Aspect	Possible values
Features This refers to the features used to model texts that are to be classified. Multimodal data could exists.	WE (word embeddings), BoW (bag-of-words), LF (linguistic features), LR (using linguistic resources, like lexicons), Raw text or Metadata
Algorithm Algorithms used to solve the task. This includes classical ones or deep learning neural networks.	See Appendix A
Approach This clusters the different algorithms into five categories.	DL-FT-LLM (Deep learning: fine-tuning over large language models, which refers to the process of taking a large language model and pre-training and adapting it to a specific task by fine-tuning its parameters with a specialized dataset), DL-Train (Deep learning: trained from samples, which refers to training deep learning models using data samples), SL (Shallow learning, i.e. classical approaches), Hybrid approaches or Other
Language The main language in which posted messages have been written	ISO 639-1 codes have been used to represent the targeted language, i.e. EN, ES, ZH
Source Where the data comes from	Reddit, Twitter, Facebook, Weibo, Clinical survey, personal notes
Annotation How ground-truth was stated, that is, how the presence of mental disorder in subject has been determined so samples in benchmark datasets are associated to a *gold standard*	Experts (psychological evaluation by professionals), By pattern (looking for specific expressions in messages), By group (according to the social network group where the messages come from), By manual review (using non-experts as annotators), Real Cases
Metrics Performance metrics used evaluate proposed approaches	F1, P (Precision), R (Recall), Acc (Accuracy), MAE (Mean Absolute Error), MSE (Mean Squared Error), R^2, ERDE (Early Risk Detection Error), RMSE (Root Mean Squared Error), AUC (Area Under the Curve), ROC, Sens (Sensivility), Spec (Specificity), PPV (Positive Predictive Value), TPR (True Positive Rate), PGSI (Problem Gambling Severity Index)

2.4. Evaluation campaigns

We have found that these mental health issues have not been addressed at NLP-related congresses until very recently. Nevertheless, during the last years several workshops have emerged to tackle the issue. Many of the references reviewed are contributions to these scientific forums. Table 4 shows the most relevant congresses and competitions for each of them and their first edition.

As can be seen, there are not many international forums dealing with these issues from the NLP point of view. One of the most relevant is eRisk@CLEF: Early risk prediction on the Internet.[6] Across the different editions, it has proposed tasks for the identification of eating disorders, gambling addiction and suicidal ideation. Regarding eating

[6] https://erisk.irlab.org/.

Table 4
Scientific forums.

Campaign name Language	Year Best approach	Data source Features	Annotation	Tasks
ERisk@CLEF English	2018 Hybrid	Reddit Extracted	Human	Early detection of signs of depression/Early detection of signs of anorexia
ERisk@CLEF English	2019 Classical ML	BDI Questionnaires Extracted	Diagnosed	Early detection of signs of anorexia/Early detection of signs of self-harm/Measuring the severity of the signs of depression
ERisk@CLEF English	2020 Hybrid	BDI Questionnaires Hybrid	Diagnosed	Early detection of signs of self-harm/Measuring the severity of the signs of depression
ERisk@CLEF English	2021 Hybrid	Reddit Hybrid	Human	Early Detection of signs of pathological gambling/Early detection of signs of self-harm/Measuring the severity of the signs of depression
ERisk@CLEF English	2022 Hybrid	Reddit Hybrid	Human	Early detection of signs of pathological gambling/Early detection of depression/Measuring the severity of the signs of eating disorders
OffensEval@SemEval English	2019 DL	Twitter End-To-End	Human	Offensive language identification/Automatic categorization of offense types/Offense target identification
OffensEval@SemEval English, Arabic, Greek, Danish and Turkish	2020 DL	Twitter End-To-End	Human	Offensive language identification/Automatic categorization of offense types/Offense target identification.
DETOXIS@IberLEF Spanish	2021 DL	NewsCom-TOX Corpus End-To-End	Diagnosed	Toxicity detection task/Toxicity level detection task
MeOffendEs@IberLEF Spanish and Mexican Spanish	2021 Hybrid	Twitter, YouTube and Instagram Hybrid	Human	Non-contextual multiclass classification for generic Spanish/Contextual multiclass classification for generic Spanish/Non-contextual binary classification for Mexican Spanish/Contextual binary classification for Mexican Spanish

disorders, apart from eRisk@CLEF, SemEval[7] and IberLEF[8] [20] forums address some o these disorders. As part of the last one, this year it has been organized a specific task called MentalRiskEs@IberLEF 2023 [21] covering, only on Spanish, eating disorders, depression, and anxiety.[9]

3. Findings

The most studied disorder was depression, with a total of 249 papers, followed by suicidal ideation, with 186 references. Far from those number was anorexia, with 14 works reviewed. More marginal was the interest for approaching pathological gambling, with a total of 10 works. Most of the papers for this last disorder came from the eRisk task, which attracted many researchers in its edition of year 2022. It could be expected a higher number of related contributions during the rest of year 2023, as the frame conference, CLEF, usually takes place in September. In any case, we can draw from Table 2 that the number of works tackling disorder detection over social media shows steady growth.

We now show our findings according to the different aspects identified and categorized. A detailed view for each disorder for all the revised papers is available in Tables B.5–B.10, placed at the end of this paper (Appendices sections).

3.1. Datasets

In Figs. 1–4 we can see the sources of data identified for research papers on anorexia, depression, suicide and gambling respectively. As can be noticed, most of the social media targeted where Facebook, Twitter, Reddit and Weibo.[10] This is interesting as, when it comes to young people, these platforms are not the preferred ones, which are Instagram, TikTok[11] and YouTube[12] according to Digital 2024: Global Overview Report.[13]

In Fig. 5 we can see, apart from the clear increase in the number of works on the subject of this overview, the hegemony of English (EN) as the most studied language. Anyhow, other languages like French (FR), Chinese (ZH) or Spanish (ES), are attracting the attention of the research community.

As last comment regarding datasets (see Fig. 6), we found very interesting the rising of data collections that have been annotated according to experts, that is, by using psychological scales to diagnose the presence of the disorder for each individual or any other rigorous method to determine the diagnosis.

3.2. Algorithms

According to the classification algorithm, interesting findings can be drawn from the reviewed works. With the advent of deep learning, most

[7] https://semeval.github.io/.
[8] http://sepln2023.sepln.org/iberlef/.
[9] https://sites.google.com/view/mentalriskes/.
[10] https://weibo.com/.
[11] https://www.tiktok.com.
[12] https://www.youtube.com.
[13] https://datareportal.com/reports/digital-2024-global-overview-report.

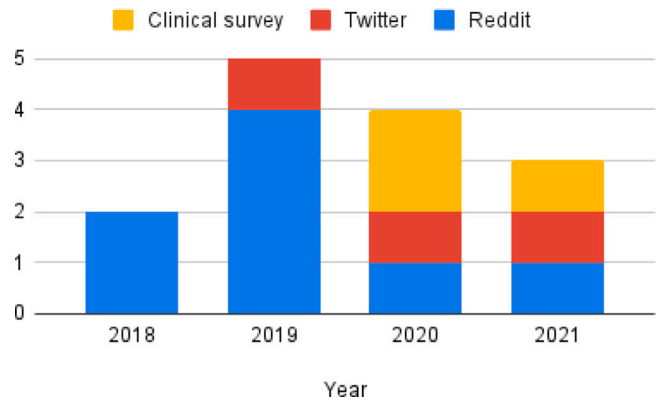

Fig. 1. Count of sources used for research papers targeting anorexia, from 2018 to 2021.

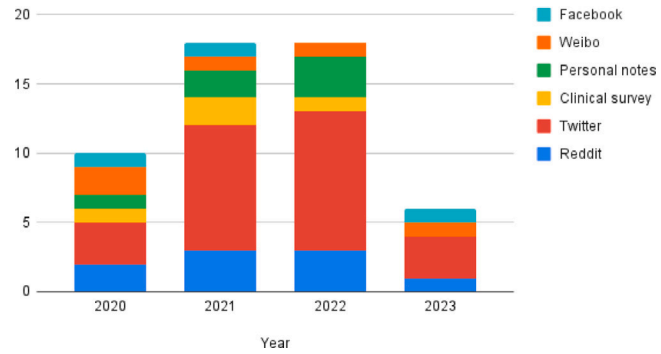

Fig. 2. Count of sources used for research papers targeting depression, from 2020 to 2023.

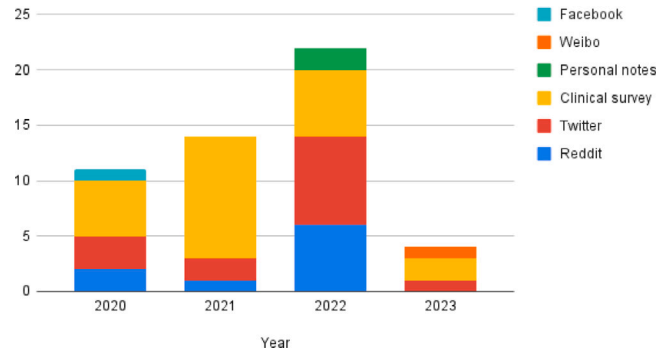

Fig. 3. Count of sources used for research papers targeting suicidal ideation, from 2020 to 2023.

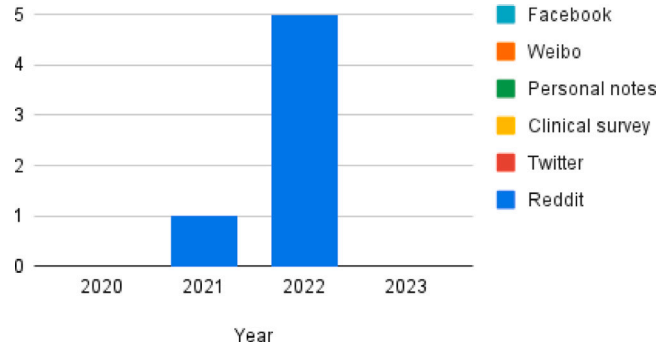

Fig. 4. Count of sources used for research papers targeting pathological gambling, from 2020 to 2023.

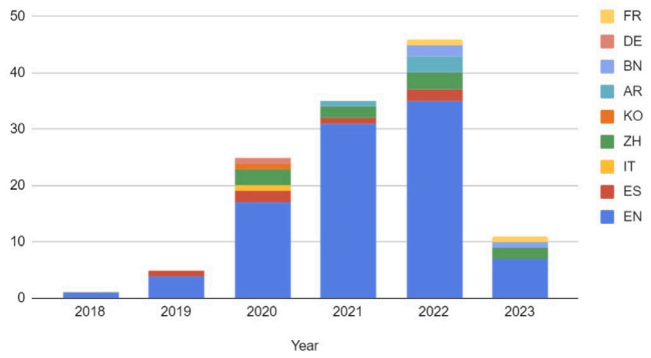

Fig. 5. Count of languages used for research papers targeting all disorders, from 2018 to 2023.

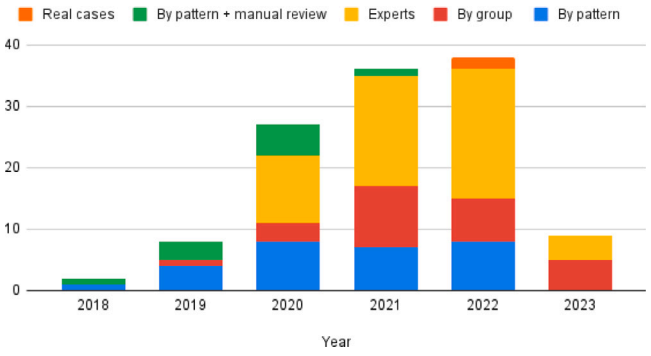

Fig. 6. Count of annotation method used for research papers, from 2018 to 2023.

of the current systems follow an *end-to-end* approach, which means that no feature engineering process is applied. Classical approaches in computational linguistics involve the processing of texts and words following syntactic and morpho-syntactic transformations, anaphora resolution, dependency parsing and other techniques to extract linguistic information [22]. Many features can be extracted from a text regarding its content, style, complexity, lexical diversity and alike. Then, features are filtered or transformed before the construction of the final vector of weights that will be fed to the classification/learning algorithm. On the contrary, end-to-end approaches do not consider linguistic information, but a fully statistical one. This is the case of current deep neuronal networks, where texts are segmented using frequency based tokenizers and resulting tokens are mapped to embeddings, which are part of the trainable network. Nevertheless, we can identify a third category in this classification dimension: hybrid approaches. Some works have found that merging linguistic information with deep encodings may lead to better performances.

This can be recognized in Figs. 7–10. Despite the advent of deep learning approaches, where most of them are done by fine-tuning pre-trained large language models (DL-FT-LLM) or deep learning for training (DL-T), there is still room for shallow learning approaches (SL), also referred to as "classical machine learning" by many authors. Hybrid approach is also used.

3.3. Remarks on most interesting works

From all the revised papers, we have found some of them quite relevant for different reasons: the originality of the approach, the outstanding results, the sophistication of the method and alike. Here is our selection of papers:

López-Úbeda et al. [23] develop a system for the automatic detection of anorexia in textual information. It is worth mentioning that to do so, they first generated a corpus with tweets written in Spanish

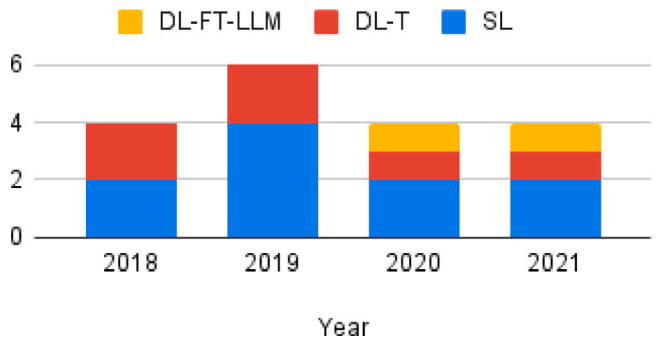

Fig. 7. Algorithms and approaches used in anorexia related works.

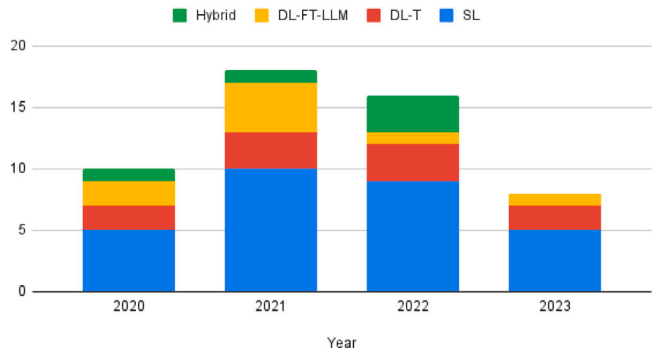

Fig. 8. Algorithms and approaches used in depression related works.

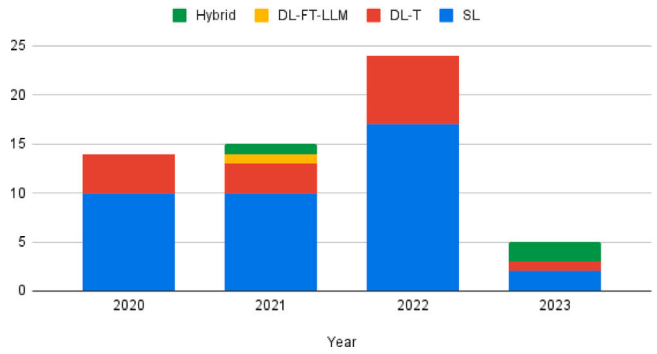

Fig. 9. Algorithms and approaches used in suicidal ideation related works.

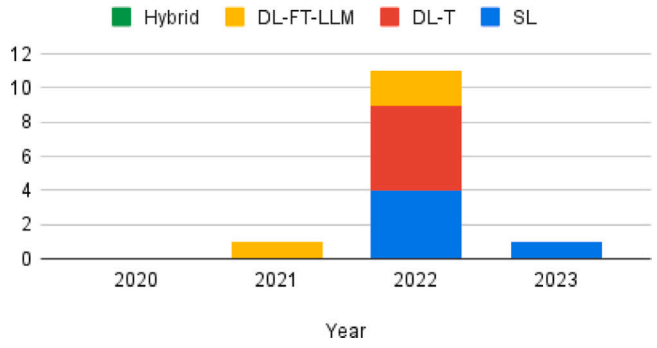

Fig. 10. Algorithms and approaches used in pathological gambling related works.

that included both people talking about anorexia and people talking about healthy eating habits. "Ana *y* mia" are the names used on websites promoting anorexia and bulimia to identify themselves. "Ana" is anorexia and "mia" is bulimia. On the one hand, they collected data regarding anorexia using as a query the hashtag #anaymia on Twitter, on the other hand they used as a query different hashtags related to food, nutrition, diet and healthy living inversely to anorexia, using the hashtags #realfood, #comidareal and #fitness.

Burkhardt et al. [24] evaluate the utility of emotion features extracted using a BERT-based model in comparison to emotions extracted using word counts as predictors of symptom severity in a large set of messages from text-based therapy sessions involving over 6,500 unique patients previously collected via the "Talkspace" platform over the course of 12 weeks, accompanied by data administered symptom scale measurements (PHQ-9 and GAD-7) every 3 weeks. Talkspace offers a paid service utilizing licensed and credentialed therapists to conduct asynchronous message based therapy conversations. Among the papers found, this is the only one that uses this online platform to collect text and generate a dataset.

Jacobucci et al. [25] use a simulation and an empirical example to demonstrate that pairing nonlinear and flexible machine learning approaches, such as random forests with the optimism-corrected bootstrap, provides highly inflated prediction estimates. They find no advantage for properly validated machine learning models over linear models. As features, they use demographic variables (e.g., age, sex), psychiatric diagnoses (e.g., bipolar disorder diagnosis, generalized anxiety disorder diagnosis), negative life events (e.g., childhood neglect), personality (i.e., neuroticism), and a history of suicidal behaviors.

Zhang et al. [26] propose a transformer-based model called TransformerRNN, which extracts contextual and long-term dependency information using a transformer encoder and a bidirectional short-term memory (BiLSTM) structure. They evaluate the model with classical models and a dataset of online sources (including 659 suicide notes, 431 last statements and 2000 neutral messages). TransformerRNN achieves 94.9% f1 performance, outperforming classical ML systems. The proposed model is effective in classifying suicide notes, which in turn can help develop suicide prevention technologies for social media. It is interesting the use of the Corpora dataset [27], where it is known that the note writer has died by suicide.

Gu et al. [28] use machine learning algorithms to extract textual features from Sina Weibo dataset[14] and build suicide risk prediction models to predict four dimensions of the Suicide Possibility Scale (hopelessness, suicidal ideation, negative self-evaluation, and hostility). With this model they detect symptoms of suicidal ideation in more detail. As features they use six dictionaries for linguistic feature extraction: The Weibo Five Basic Mood Lexicon, The Individualism/Collectivism Lexicon, Researchers built the Chinese suicide dictionary, The Chinese Version of the Moral Foundations Dictionary, The Moral Motivation Dictionary, the SCLIWC and the Simplified Chinese MicroblogWord Count tool.

Dai et al. [29] use a multimodal approach to automatic depression detection. Their experiments take a dataset of clinical interviews with video and audio, including transcripts. Subjects have been evaluated according to the PHQ8 psychological scale. From this information, audio, video and semantic features are generated and filtered using the minimum Redundancy Maximum Relevance (mRMR) approach. For semantic features, the LIWC categories were used. The F1 score obtained was of 0.67 on the test set.

The work by Lara et al. [30] study depression detection using the eRisk 2017 and 2018 task datasets as corpus. It is based on the Bag of Sub-Emotions (BoE) approach, which maps words to more fine-grained categories that are related to main emotions. Their approach, DeepBoSE, follows a similar approach, but relying on deep learning components. Therefore, words are represented by embeddings. Similarly, words related to emotions taken from a lexicon are also represented by embeddings. This allows to generate several sub-clusters for each emotion according to distances between these embeddings. How

[14] Available at https://github.com/bryant03/Sina-Weibo-Dataset.

Table B.5
Works about eating disorder detection (Anorexia).

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[31]	BoW, LF, LR, WE, Metadata	SL, CNN	SL, DL-T	EN	Reddit	By pattern + manual review	ERDE, F1
[32]	WE, BoW	SVM, LR, RF, AB, RNN	SL, DL-T	EN	Reddit	By pattern + manual review	ERDE, F1
[23]	BoW	SVM, MLP, DT, RF, NB, LR	SL	ES	Twitter	By pattern	F1
[33]	BoW	LSTM, SGDC	DL-T, SL	EN	Reddit	By pattern + manual review	ERDE, F1
[34]	WE	UMLS	DL-Train	EN	Reddit	By pattern + manual review	ERDE, F1
[35]	BoW	SS3	SL	EN	Reddit	By pattern + manual review	ERDE
[36]	BoW, WE	LR	SL	EN	Reddit	By group	Acc
[37]	LF, Metadata	RF, LR	SL	EN	Reddit	By pattern + manual review	ERDE
[38]	Metadata	SVM	SL	IT	Clinical survey	Experts	P, F1
[39]	WE	CNN	DL-Train	EN	Reddit	By pattern + manual review	Acc, P, R, F1
[40]	BoW, Metadata	RF, LR, DT, SVM	SL	EN	Clinical survey	Experts	Acc
[41]	Raw text	BERT	DL-FT-LLM	EN	Twitter	Experts	Acc
[42]	BoW, Metadata	SVM, LR, KNN	SL	EN	Clinical survey	Experts	AUC, Acc
[43]	WE, Raw text	LSTM, CNN, BiLSTM, XLM, BERT, BETO	DT-Train, DL-FT-LLM	ES	Twitter	By pattern	P, R, F1

Table B.6
Works about depression detection (2020 and 2021 years).

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[44]		DT, RF, NB, SVM, KNN	SL	EN	Clinical survey	Experts	Acc, F1
[45]	WE	CNN	DL-Train	EN	Reddit	By pattern + manual review	ERDE, F1
[46]	Metadata	DT, NB, SVM	SL	EN	Twitter		Acc, F1
[47]	BoW, Metadata	DL, SVM	Hybrid	ZH	Weibo	By pattern	P
[48]	BoW, LR		SL		Twitter	Experts	
[49]	LR, Metadata	lexicon-based, SVM, NB, DT	SL	EN	Facebook	Experts	*p*-value
[50]	WE, Metadata	BiLSTM	DL-Train	EN	Reddit	By pattern + manual review	ERDE, F1
[51]	Raw text	BERT, RoBERTa, XLNET	DL-FT-LLM	ZH	Weibo	By pattern + manual review	F1
[52]	Metadata	RF	SL	KO	Personal notes	Experts	AUC
[53]	Raw text	LSTM, CNN	DL-FT-LLM	EN	Twitter	By group	AUC
[54]	BoW LR,	LSVM, MLP, DT	SL	EN	Twitter	By pattern	Acc, F1, P, R
[55]	LR	LR, MLP, SVM, DT, GB	SL	EN	Twitter	By pattern	Acc, F1, P, R
[56]	Raw text	BERT, GRU BART, CNN	DL-FT-LLM	EN	Twitter	By pattern	Acc, F1, P, R
[57]	BoW	NB, NBT	SL	EN	Twitter	By polarity	Acc, R, P
[58]	Raw text	BERT, RoBERTa, DistilBERT, ELECTRA	DL-FT-LLM	EN	Reddit	By group	F1
[59]	WE, LR	BiLSTM, FFN	DL-Train	EN	Reddit	By group	Acc, F1, P, R
[30]	Raw tex	DeepBoSE	DL-FT-LLM	EN	Twitter	By pattern + manual review	Acc, F1, P, R
[60]	BoW	SVM, NB, KNN	SL	EN	Twitter	By group	Acc, F1, P, R
[61]	Raw text	ARABERT, MARBERT	DL-FT-LLM	AR	Twitter, personal notes, Web	By group	Acc, F1, P, R
[62]	BoW	SVM, RF	SL	EN	Twitter	By pattern	Acc, F1, P, R
[63]	WE	SVM, DNN, RF, NV, CNN	SL	EN	Twitter	By pattern	P, R, MSE
[64]	BoW	SVM, KNN, RF, XGBoost	SL	EN	Reddit	By group	Acc, P, R, AUC
[65]	WE	BiLSTM, FFN	DL-Train	ZH	Weibo	By group	Acc, P, R, F1
[66]	BoW, Metadata	NB, NN, DT, SVM, KNN	SL	TH	Facebook	Experts	Acc, P, R, F1
[67]	BoW	SVM, KNN, DT, RF, NB, LR	SL	RU	Vkontakte	By group	Acc, P, R, F1, AUC

a new text is mapped to these clusters, encoded and used to generate a final prediction is down following a deep learning architecture with several layers. The results obtained are as performant as the best results obtained in the eRisk shared tasks.

Another remarkable study is the one by Hemtanon et al. [66]. This work has not received a single citation at the time of our review, maybe because it is focused only on the Thai language. Nevertheless, the work introduces very interesting ideas and a robuts foundation. Firstly, participants have passed a clinical test, the PHQ9 questionnire, so a consistent diagnosis is performed. Secondly, they do not rely only on posts (taken from social network Facebook), but also in behavior features such as daily rates of publications, reactions and so on. Also a TF-IDF representation of the text in the post is calculated. The results of these two representations on different classical machine learning algorithms show that, although a high F-score can be obtained over textual data (0.88 using a simple feed-forware network), by using only behavioral features the prediction is just perfect (using the K-NN algorithm). This study shows that, when it comes to depression detection

Table B.7
Works about depression detection (2022 and 2023 years).

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[68]	WE	CNN, BiLSTM	DL-Train	ZH	Clinical survey	Experts	Acc, F1
[69]	WE, Metadata	CNN-biLSTM	DL-Train	EN	Twitter	Experts	Acc, F1, P, R
[70]	Metadata	MLP	DL-Train	EN	Twitter	By pattern and manual review	P
[71]	Metadata	LR, RF, NB, SVM, LDA, KNN	SL	BN	personal notes via online	By pattern	Acc, F1, P, R
[72]	BoW	SVM, KNN, RF, LR, NB, AB	SL	AR	Twitter	Experts	Acc, F1, P, R
[73]	Metadata	CBPT	SL	EN	Twitter	Experts	Acc, F1
[74]	LR, Metadata	SVM, DT, NB, RF, DL	SL	TH	Twitter	Experts	Acc
[75]	BoW	6 Multinomial NB, SVM, DT, RF, KNN, LR	SL	EN	Twitter	By group	Acc, F1, P, R
[76]	LR	CNN, LSTM, BiLSTM	DL-train	EN	Twitter	By emotion	Acc
[77]	Metadata	Voting classifier (RF, KNN, MLP)	Hybrid	BN	Personal notes	Experts	Acc
[24]	LR	BERT	DL-FT-LLM	EN	Personal notes, messages	Experts	ROC
[78]	Raw text	Ensemble (LightGBM, XGBoost) + (ELECTRA, RoBERTa, DeBERTa)	Hybrid	EN	social media postings	By pattern and manual review	F1
[79]	WE	SVM, GB, LR, XGB, ExtraTree, Bagging, RF, AB	SL	EN	Twitter	Experts	Acc
[80]	Metadata	NN	MTL (DL FusionNet)	ZH	Weibo	Experts	F1
[81]	Raw Text	SVM, LR	SL	EN	clinical surveys	Experts	Acc
[82]	BoW, WE	SVM, RF	Hibrid ML	EN	Twitter	Experts	Acc
[83]	BoW	NB, DT, SVM, KNN	SL	EN	Twitter		Acc
[84]	Raw text	ALBERT	DL-FT-LLM	EN	Reddit	Experts	F1
[85]	Raw text, BoW	BERT, SVM, knowledge graph and textual representations	Hybrid	EN	Reddit	Experts	F1
[86]		SVM, RF, XGBoost	SL	EN	Reddit	Experts	F1
[87]	Raw text	S-RoBERTa + BiLSTM	DL-Train, DL-FT-LLM	EN	Bot	Experts	MAE, F1
[88]	BoW	SVM, HAN, NB, BiLSTM	SL, DL-Train	EN	Twitter, Reddit	By group	Acc
[89]	BoW	HAN	SL	EN	Twitter	By group	Acc, P, R, F1
[90]	BoW	RF, LR, DT, SVM, KNN	SL	BN	Facebook	By group	Acc, P, R, F1
[91]	BoW	LDA, SVM	SL		Twitter	By group	Acc
[92]	SCLIWC	LR	SL	ZH	Weibo	By group	F1, AUC

over social networks, many valuable information can be considered to enhance prediction performance.

Milintsevich et al. [87] tackle depression detection indirectly, by classifying text written by subjects according to major symptoms (eight in total). A multi-target hierarchical regression model to that end is trained and evaluated on the DAIC-WOZ corpus, which consists of interviews between a person and a virtual assistant. According to the answers given, the participants (200 in total) have been clinically diagnosed. The results obtained are significant, with a macro-F1 score of 0.739. This is a very interesting approach because, as symptoms are the target of the prediction, a more informative answer about the presence of the disorder can be reported.

3.4. Projects

Several research projects have focused on some of these problems in recent years. Actually, some of these projects are still under development. We summarize here those projects related to social media analysis for detection and prevention of certain disorders or risky behaviors:

- MENHIR[15] project stands for "Mental health monitoring through interactive conversations". It is funded by the European Program H2020, and it is participated by several research teams. The aim of the project is to research and develop conversational technologies to promote mental health and assist people with mental ill health (mainly mild depression and anxiety). One interesting research topic within the project is the construction of chatbots for mental health related dialogues and how natural language understanding techniques (NLU) can be applied in this scenario [149].
- AMIC[16] project studies the development of speech technologies for affective and inclusive communication. It is funded by Spanish Ministerio de Ciencia, Innovación *y* Universidades and FEDER founds. Regarding NLP, they have some contributions to sentiment analysis on Twitter [150] and irony detection [151].
- STOP[17] project studied the use of artificial intelligence to mental health issues related to social media, by looking for patterns of suicide [152], depression or eating disorders [153]. The project was funded by Universitat Pompeu Fabra (Spain) and finished in 2021. By analyzing messages exchanged in Instagram and Twitter, they were able to identify risky behaviors, suggesting calling to experts. A wide range of features were fed into the prediction model: word embeddings, bag of words, n-grams, behavioral features, tweet statistics and sociometrics features. In the case of suicidal ideation, the best systems were found to be the combination of psychological, sociometric and visual features with an SVM classifier.
- NetLang[18] project focus on cyberbullying and the language used to this end in social media for both English and Portuguese languages. The project was developed by the University of Minho, in Portugal (it ended in 2021). Within the project, a corpus[19] was built, containing labeled texts from YouTube and different newspapers.

[15] https://menhir-project.eu/.

[16] http://dihana.cps.unizar.es/~alfonso/amic.

[17] https://stop-project.github.io/.

[18] https://sites.google.com/site/projectnetlang/introduction.

[19] https://netlang-corpus.ilch.uminho.pt/cgi-enabled/corpus.cgi.

Table B.8
Works about suicidal ideation (2020 and 2021 years).

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[93]	Bow, WE	RF, SVM, XGBoost, LSTM, CNN, LSTM-CNN	SL, DL-Train	EN	Reddit	By group	Acc, P, R, F1
[94]	LF	CART, RF	SL, DL-Train	EN	Clinical	Experts	Acc, AUC
[95]	BoW	LR	SL, DL-Train	EN	Clinical	Experts	AUC
[96]	BoW, WE	SL, CNN	SL, DL-Train	ES	Twitter	By pattern	Acc, P, R, F1, AUC
[97]	Metadata	LR, lasso, ridge, RF	SL	EN, DE	Clinical	Experts	AUC, BS, Sens, PPV
[98]	LR	LR, DT, RF, GBR, SVM, MLP	SL	ZH	Clinical	Experts	P, R, F1, Acc, Sens, Spec, ROC AUC, PR AUC
[99]	BoW, LF, LR	NB, SVM, KNN	SL	EN	Reddit	By group	P, R, F1, Acc
[100]	BoW, LR, LF	SVM	SL	EN	Clinical	Experts	AUC, CI
[101]	BoW	NB, MNB, DT, LR, SVM, AB, RF	SL	EN	Twitter	By pattern, By manual review	P, R, Acc
[102]	BoW, LF, LR	KStar, SMO, RC, RT, RF	SL	EN, ES	Facebook, Twitter, Instagram, Blogs, Forums	By manual review	P, R, F1, ROC
[25]	LF	RF, LR	SL	EN	Clinical	Experts	AUC
[103]	LR, WE	NB, LASSO, RF, eXGB	SL	EN	Clinical	Experts	AUC, Sens, Spec
[104]	BoW, LF	SVM, NB, CNN	Hybrid	EN	Clinical	Experts	Acc, P, R, F1, AUC
[26]	LF, LR	TransformerRNN	DL-Train	EN	Clinical	Experts	F1
[105]	LF	RF	SL	EN	Clinical	Experts	P, Acc
[106]	LF	KNN, NN, avNNET, NB	SL	EN	Clinical	Experts	Acc, P, R, F1, AUC
[107]	LR	LDA, KNN, SVM, D-Tree	SL	EN	CLPsych 2021	Experts	F1, F2,TPR
[108]	Metadata	RF, KNN	SL	EN	Clinical	Experts	Acc
[109]	Metadata	RF	SL	EN	Clinical	Experts	AUC
[110]	Metadata	RF	SL	EN	Clinical	Experts	AUC
[111]	Raw text	BERT, S-BERT, GUSE	DL-FT-LLM	EN	Reddit	By group	Acc, P, R, F1, AUC
[112]	Raw text	BiLSTM	DL-Train	EN	Clinical	Experts	P, R, c-statistic
[113]	Metadata	CART	SL	EN	Clinical	Experts	Acc, F-Statistic, ROC-AUC, PR-AUC
[114]	Raw text	LSTM	DL-Train	EN	Twitter	By group	AUC
[115]	BoW	CSVM	SL	EN	Twitter	By group	Acc, P, R, F1,

- HATEMETER[20] project aimed to develop a hate speech tool for monitoring, analyzing and tackling Anti-Muslim hatred online [154]. Funded by the European Commission, it ended in January 2020, but the tool is still available.
- The Big Hug[21] project is an ambitious research effort focused on the early detection, over social networks, of different disorders (eating ones, anxiety, depression, suicidal ideation and gambling addiction) and misbehaviors (cyberbullying). The targeted population are mainly adolescents and is exploring warning strategies without loosing privacy and confidentiality of communications. It is funded by the Andalusian Government, in Spain and will end in 2023.
- The PRECOM[22] project is a continuation of Big Hug, but focused only on pathological gambling. Instead of working with social media data, the project relies on an approach based on conversational bots. It is funded by the Government of Spain and will end in 2023.

4. Discussion

In this section we summarize the main issues found during our review, organized in the following facets:

Source of data As usual, most of the work is in English, so it is worth highlighting whether languages different of English or multilingual solutions are explored. Nevertheless, in general, the existence of corpora for performing research in automatic disorders detection is scarce. Most of the known datasets has been prepared ad-hoc for certain evaluation campaigns. The data is difficult to collect, due to restrictions imposed by social platforms, privacy matters and legal issues.

Annotation The annotation of these resources is costly, even when using crowdsourcing approaches. But the most important issue is that expert annotation is not common, but desirable. A major obstacle in the advance of automatic detection systems is the lack of annotated data. Despite of the promising expectations of data augmentation solutions and pseudo-automatic annotation, annotating this data should be performed by experts (psychologists and psychiatrists, mainly), so we can effectively evaluate the real performance of these systems in a more rigorous way.

Features Although end-to-end approaches are populating the arena, still many systems rely on engineered features. Some of the features can be hybrid, merging visual encodings with textual ones, and there are many approaches integrating linguistic features (complexity, content-related, named entities, words from lexicons...) to end-to-end encodings.

Approaches and performances From the classical machine learning ones, linear regression and SVM prevail. Nevertheless, LLM based ones like BERT based ones and, even, GPT-based

[20] http://hatemeter.eu/.
[21] http://bighug.ujaen.es.
[22] http://precom.ujaen.es.

Table B.9
Works about suicidal ideation (2022 and 2023 years).

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[116]	BoW, WE	RF, BiLSTM	SL, DL-Train	EN	Twitter	By pattern	Acc, F1
[117]	Raw text, Metadata	LR, PR, DT, GB	SL	EN	Clinical	Experts	AUC
[118]	BoW	SVM	SL	EN	Clinical	Experts	AUC, Brier scores
[119]	Raw text, LF	GB	SL	EN	Clinical	Experts	P, R
[120]	Raw text, LR, Metadata	LR	SL		Twitter		Acc
[121]	Metadata	SVM, BERT, XLNet, RS, CR	SL, DL-Train	ES (Argentina)	Personal notes	Experts	AUC, P, IBS, Sens, Spec
[122]	BoW, WE, LR	CNN–BiLSTM, XGBoost	DL-Train	EN	Reddit	By group	Acc, P, R, F1
[123]	WE	SVM, CNN, LSTM, LSTM-CNN	SL, DL-Train	EN	Reddit	Experts, By manual review	Acc, F1
[124]	Metadata	RF	SL	EN	Clinical	Experts	AUC, ROC
[82]	BoW, WE	SVM, RF	SL	EN	Reddit	By group	Acc, P, R, F1
[125]	BoW, Raw text	SVM	SL	EN	Twitter	By manual review	Acc, F1
[126]	BoW, LR, Raw text	LR, RF, GB, KNN, SVM, AB, BERT, RoBERTa	SL, DL-Train	ES	Clinical	Experts	P, R, F1
[127]	Raw text, LR, Metadata	RF, BN, SVM, DT, AB, LR, XGBoost	SL	EN	Twitter, Youtube, Tumblr	Real cases	P, R, F1, AUC, MAP, p@5, p@10, p@15, p@21
[128]	WE	BERT	DL-Train	EN	Reddit	By manual review	micro-F1, macro-F1, weighted-F1
[129]	WE	SVM, XGBoost, NB	SL	EN	Reddit, Twitter	By group, By pattern	Acc
[130]	Raw text	LSTM	DL-Train	EN	Twitter	By pattern	Acc, P, R, F1
[131]	BoW	LR, RF, SVM, MNB	SL	AR	Suicide notes, Personal notes, Twitter	Real cases, by manual review	P, R, F1
[132]	BoW	SVM, LR, AB	SL	EN	Reddit	By pattern, By group	P, R, Acc
[133]	Metadata	LR, DT, SVM	SL	MS	Clinical	Experts	P, R, Acc, Spec
[134]	BoW, LR, LF	RF	SL	EN	Twitter	By pattern, By manual review	P, R, Acc
[28]	LF	LR	SL	ZH	Weibo	Experts	SPS
[135]	Metadata	SVM, KNN, CNN	Hybrid	EN	FER2013		Accuracy
[136]	Raw text, BoW	RF, CNN–BiLSTM	Hybrid	EN	Twitter	By pattern	Acc, P, R, F1
[137]	LF, LR	Transformer RNN	DL-Train	EN	Clinical	Experts	AUC-ROC, Sens, Spec
[138]		NB	SL	EN	Clinical	Experts	AUC, PPV, TPR

Table B.10
Works about pathological gambling.

Ref.	Features	Algorithm	Approach	Lang.	Source	Annotation	Metrics
[139]	Raw data	BERT	DL-FT-LLM	EN	Reddit	By manual review	P, R, F1
[140]	multimodal, WE	BiLSTM, CNN	DL-Train, DL-FT-LLM	ZH	Web pages	By group	Acc, P, R, F1
[141]	WE	TextRNN, CNN	DL-Train	ZH	Web pages	By group	P, R, F1
[142]	BoW	LinearSVC, LR, RF, KNN, DT, BERT, RoBERTa, ALBERT	SL, DL-Train	EN	Reddit	By manual review	P, R, F1
[143]	Raw data	RoBERTa	DL-FT-LLM	EN	Reddit	By manual review	P, R, F1
[144]	Metadata	RF, SVM, LR, NN	SL	FR	Web pages	By manual review	PGSI, AUROC
[145]	WE	RoBERTa, MiniLM	DL-Train	EN	Reddit	By manual review	P, R, F1
[146]	WE	ANN	SL	EN	Reddit	By manual review	P, R, F1
[147]	BoW, WE, LR, Raw text	AB, LR, RF, SVM, BERT, RoBERTa, Longformer, BioBERT	SL, DL-Train	EN	Reddit	By manual review	P, R, F1
[148]	Metadata		SL	FR	Web pages	By manual review	PGSI, AUC

prompted systems are raising. Regarding performances, even if some disorders, like eating related ones, have been found to allow for automatic detection with good performances, others, like depression, are still difficult to track and identify in initial states. Besides, most of the problems covered in this overview have not been explored in low-resourced languages (i.e. others than English), so effectiveness and conclusions cannot be always transferred. Certain disorders have not been explored in depth, as it happens with gambling addiction. And, for sure, new disorders and unwanted conducts may appear in the near future. A conclusion that can be drawn from the results reported by some systems, is that hybrid approaches are worth exploring. Despite the semantics contained in textual communications, visual information and sociometrics have been found to improve

prediction performance. We believe that techniques coming from time series analysis can also contribute to better early risk detection systems.

Confidence Another crucial aspect is trustworthiness. Parents, tutors and society in general will only rely on this detection systems if both conditions are met: accuracy and explainability. A trained algorithm could be accurate, reporting high levels of precision and recall, but if the reasons for a given prediction, specially when a high risk is detected, are not provided, the adoption of such a systems will not take place. This is an issue spotted in recent works [155], as a major one for NLP over the task of detection mental disorders. Although this is related to what has become *Safe Artificial Intelligence*. Explain causality and key factors in how algorithms behave will be a parallel task for these systems to enter the market.

Applicability As a last but, maybe, most important challenge, is to find a way to make these systems be part of real world applications to make a clear and, if possible, measurable impact in vulnerable communities. An interesting approach to tackle with people that match a risk profile when analyzing social media interactions is the one followed by the STOP project introduced in Section 3.4. A targeted campaign was launched over those profiles, offering a 24 h hotline for emotional support. Actually, the calls to such a service coming from social media increased by 60%. Actually, one of the conclusions of the Big Hug project was that most of the adolescents were reluctant to facilitate access to their activity in social networks, so parental monitoring was not an option.

The difficulties to propose non-invasive solutions and privacy related matters prevent a faster application of such systems in real-world environments. Regarding data protection and privacy, the deployment of real-world solutions has to follow a process which is not easy and close to what is requested to clinical products. Impact studies and evaluations over a significant number of the targeted population are imposed in order to ensure that the application is beneficial rather than harmful. The latter is a common worry for parents, as they are mainly afraid of systems alerting about a potential disorder which would need further psychological evaluations and that could cause, by the process itself, a trauma in a false positive adolescent. It is time to boost the research in this direction, so artificial intelligence emerges as shield for the society against these undesired situations.

5. Conclusions

This overview compiles the most relevant and novel research in early detection of several emotional disorders and problems related to social media interaction among young population mainly. Current statistics on the prevalence of disorders like depression, anxiety, eating disorders, suicidal ideation, gambling addiction show an increase after the pandemic situation suffered all over the world. Many efforts have been applied to mitigate such a negative tendency, including the exploration of prevention mechanisms by means of supervising algorithms. We have reviewed the most relevant scientific forums, summarized the state of the art in automatic detection using NLP approaches and listed related projects.

Some main challenges and topics to be tackled have been proposed. In this regard, we can summarize that deep learning techniques are gaining attraction, but still classic algorithms like SVM can lead to performing solutions. Integration of different features seems also promising, beyond pure end-to-end approaches. Some of the main challenges are related to the applicability of these results in real-world scenarios, due to the lack of explainability in most cases or data protection and privacy policies. When talking about early detection, depression remains still the most difficult disorder, maybe because it is needed to analyze subject's behavior for longer periods of time.

There is a large room for research in other languages different from English. To do so, well-annotated corpora (i.e. by experts) have to be generated. This will allow to better understand the multilingual capabilities of proposed methods. During our study, we found that focusing on social media could be a limiting factor, as the applicability of these platforms generates some worries about privacy and potential "alarm" effects. Research on the use of chatbots for mental health is not new, and these systems are opening the exploration of controlled environments for automatic detection systems [156–158]. Besides, research is moving towards symptom detection, as in the last edition of eRisk task,[23] rather than disorder detection, which provides a better understanding of the emotional, behavioral, and mental situation of a subject and ease the identification of comorbidities. We believe that major scientific advances will be come from these novel approaches, as large language models continue their evolution in their emotional, multimodal and empathic capabilities.

This paper is, therefore, a valuable contribution for those working on or entering to the subject, as it offers a wide oversight on this research area and enables a better positioning on the matter of study to conveniently continue potential further research.

Limitations

The speed at which new research is being produced and published has become overwhelming. This work may have missed relevant work worth mentioning. Besides, some papers do not clarify certain key aspects considered in our analysis. One of these key aspects is how to assert the quality of the annotation process, so this lack of detail could influence our evaluation. As our meta-study has been focused on textual data, systems dealing with voice or video-recorded records may have lied out of our scope, still representing a relevant source of research in the automatic detection of mental disorders.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgments

This work has been partially supported by projects CONSENSO (PID2021-122263OB-C21), MODERATES (TED2021-130145B-I00), SocialTOX (PDC2022-133146-C21) funded by Plan Nacional I+D+i from the Spanish Government, and project PRECOM (SUBV-00016) funded by the Ministry of Consumer Affairs of the Spanish Government. The research work conducted by Salud María Jiménez-Zafra has been supported by Action 7 from Universidad de Jaén under the Operational Plan for Research Support 2023–2024.

[23] https://erisk.irlab.org/.

Appendix A. Acronyms used for algoritms

AB	Adaptive Boosting
ALBERT	A little Bidirectional Encoder Representations from Transformers
ANN	Artificial Neural Networks
ARABERT	Arabic Bidirectional Encoder Representations from Transformers
avNNET	Neural Networks using model Averaging
BERT	Bidirectional Encoder Representations from Transformers
BETO	Spanish Bidirectional Encoder Representations from Transformers
BiLSTM	Bidirectional Long Short-Term Memory
BN	Bayesian Network
CART	Classification And Regression Tree
CBPT	Cost-sensitive Boosting Pruning Trees
CNN	Convolutional Neural Network
CR	Cox Regression
CSVM	Clustered Support Vector Machine
DeBERTa	Decoding-enhanced Bidirectional Encoder Representations from Transformers Approach
DeepBoSE	Deep Bag of Sub-Emotions
DistilBERT	Distilled BERT
DT	Decition Tree
ELECTRA	Efficiently Learning an Encoder that Classifies Token Replacements Accurately
ERT	Extremely Randomized Tree
eXGB	Ensemble eXtreme Gradient Boosting
FNN	Deep Feed Forward Neural
GD	Gradient Boost
GRU	Gated Recurrent Unit
GUSE	Google's Universal Sentence Encoding
HAN	Hierarchical Attention Networks
KNN	K-Nearest Neighbors
LASSO	Least Absolute Shrinkage and Selection Operator
LDA	Linear Discriminant Analysis
LightGBM	Light Gradient Boosted Machine
LR	Logistic Regression
LSTM	Long Short-Term Memory
MARBERT	Dialectal Arabic Bidirectional Encoder Representations from Transformers
MLP	MultiLayer Perceptron
MNB	Multinomial Naïve Bayes
MTL	MultiTask Learning
NB	Naïves Bayes
NN	Neural Networks
PR	Penalized Regression
RF	Random Forest
RoBERTa	Robustly Optimized Bidirectional Encoder Representations from Transformers Approach
RS	Random Survival Forest
SGDC	Stochastic Gradient Boost Classifier
SVM	Support Vector Machine
UMLS	Unified Medical Language System
USE	Universal Sentence Encoding
XGBoost	eXtreme Gradient Boosting
XLM	Cross-lingual Language Model
XLnet	Extension of transformer-XL (eXtra Long)

Appendix B. Summary tables of the review

See Tables B.5–B.10

References

[1] World Health Organization, Mental health of adolescents, 2021, https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health. (Accessed 20 May 2023).

[2] M. Garaigordobil, Prevalencia y consecuencias del cyberbullying: una revisión, Int. J. Psychol. Psychol. Ther. 11 (2) (2011) 233–254.

[3] D. Olweus, S.P. Limber, Some problems with cyberbullying research, Curr. Opin. Psychol. 19 (2018) 139–143, http://dx.doi.org/10.1016/j.copsyc.2017.04.012, URL https://www.sciencedirect.com/science/article/pii/S2352250X17301033. Aggression and violence.

[4] L. Sher, The impact of the COVID-19 pandemic on suicide rates, QJM: Int. J. Med. 113 (10) (2020) 707–712.

[5] J. Śniadach, S. Szymkowiak, P. Osip, N. Waszkiewicz, Increased depression and anxiety disorders during the COVID-19 pandemic in children and adolescents: A literature review, Life 11 (11) (2021) http://dx.doi.org/10.3390/life11111188.

[6] M.T. Hawes, A.K. Szenczy, D.N. Klein, G. Hajcak, B.D. Nelson, Increases in depression and anxiety symptoms in adolescents and young adults during the COVID-19 pandemic, Psychol. Med. (2021) 1–9.

[7] S. Kemp, Digital 2023: Global overview report, 2023, URL https://datareportal.com/reports/digital-2023-global-overview-report.

[8] M.J. Page, J.E. McKenzie, P.M. Bossuyt, I. Boutron, T.C. Hoffmann, C.D. Mulrow, L. Shamseer, J.M. Tetzlaff, D. Moher, Updating guidance for reporting systematic reviews: development of the PRISMA 2020 statement, J. Clin. Epidemiol. 134 (2021) 103–112.

[9] World Health Organization, Depression, 2022, https://www.who.int/es/news-room/fact-sheets/detail/depression. (Accessed 15 May 2023).

[10] T.A. Kato, N. Shinfuku, M. Tateno, Internet society, internet addiction, and pathological social withdrawal: the chicken and egg dilemma for internet addiction and hikikomori, Curr. Opin. Psychiatry 33 (3) (2020) 264–270.

[11] L.L. Hornberger, M.A. Lane, T.C.O. ADOLESCENCE, L.L. Hornberger, M. Lane, C.C. Breuner, E.M. Alderman, L.K. Grubb, M. Powers, K.K. Upadhya, S.B. Wallace, L.L. Hornberger, M. Lane, M. FRCPC, M. Loveless, S. Menon, L. Zapata, L. Hua, K. Smith, J. Baumberger, Identification and management of eating disorders in children and adolescents, Pediatrics 147 (1) (2021) e2020040279, http://dx.doi.org/10.1542/peds.2020-040279.

[12] R.J. Marks, A. De Foe, J. Collett, The pursuit of wellness: Social media, body image and eating disorders, Child. Youth Serv. Rev. 119 (2020) 105659.

[13] P. Aparicio-Martinez, A.-J. Perea-Moreno, M.P. Martinez-Jimenez, M.D. Redel-Macías, C. Pagliari, M. Vaquero-Abellan, Social media, thin-ideal, body dissatisfaction and disordered eating attitudes: An exploratory analysis, Int. J. Environ. Res. Public Health 16 (21) (2019) 4177.

[14] World Heatlh Organization, Suicide, 2021, https://www.who.int/news-room/fact-sheets/detail/suicide (Accesed 30 April 2023).

[15] J. Morrison, DSM-5® Guía Para el Diagnóstico Clínico, Editorial El Manual Moderno, 2015.

[16] J. Parapar, P. Martín-Rodilla, D.E. Losada, F. Crestani, Erisk 2021: pathological gambling, self-harm and depression challenges, in: European Conference on Information Retrieval, Springer, 2021, pp. 650–656.

[17] M. Chóliz, Ethical gambling: A necessary new point of view of gambling in public health policies, Front. Public Health 6 (2018) 12.

[18] CDC: Centers for Disease Control and Prevention, The youth risk behavior survey data summary & trends report: 2011–2021, 2023.

[19] T. Zhang, A.M. Schoene, S. Ji, S. Ananiadou, Natural language processing applied to mental illness detection: a narrative review, NPJ Digit. Med. 5 (1) (2022) 46.

[20] S.M. Jiménez-Zafra, F. Rangel, M.M.-y. Gómez, Overview of IberLEF 2023: Natural language processing challenges for spanish and other iberian languages, in: Proceedings of the Iberian Languages Evaluation Forum (IberLEF 2023), Co-Located with the 39th Conference of the Spanish Society for Natural Language Processing, SEPLN 2023, CEURWS. Org, 2023.

[21] A.M. Mármol-Romero, A. Moreno-Muñoz, F.M. Plaza-del Arco, M.D. Molina-González, M.T. Martín-Valdivia, L.A. Ureña-López, A. Montejo-Raéz, Overview of mentalriskes at IberLEF 2023: Early detection of mental disorders risk in spanish, Proces. Leng. Nat. 71 (2023) 329–350.

[22] R. Mitkov, The Oxford Handbook of Computational Linguistics, Oxford University Press, 2022.

[23] P. López-Úbeda, F. Plaza-Del-Arco, M. Díaz-Galiano, L. Alfonso Ureña-López, M.-T. Martín-Valdivia, Detecting anorexia in spanish tweets, in: International Conference Recent Advances in Natural Language Processing, RANLP, 2019-September, 2019, pp. 655–663, http://dx.doi.org/10.26615/978-954-452-056-4_077.

[24] H. Burkhardt, M. Pullmann, T. Hull, P. Areán, T. Cohen, Comparing emotion feature extraction approaches for predicting depression and anxiety, in: CLPsych 2022 - 8th Workshop on Computational Linguistics and Clinical Psychology, Proceedings, 2022, pp. 105–115.

[25] R. Jacobucci, A. Littlefield, A. Millner, E. Kleiman, D. Steinley, Evidence of inflated prediction performance: A commentary on machine learning and suicide research, Clin. Psychol. Sci. 9 (1) (2021) 129–134, http://dx.doi.org/10.1177/2167702620954216.

[26] T. Zhang, A. Schoene, S. Ananiadou, Automatic identification of suicide notes with a transformer-based deep learning model, Internet Interv. 25 (2021) http://dx.doi.org/10.1016/j.invent.2021.100422.

[27] A.M. Schoene, G. Lacey, A.P. Turner, N. Dethlefs, Dilated lstm with attention for classification of suicide notes, in: Proceedings of the Tenth International Workshop on Health Text Mining and Information Analysis, LOUHI 2019, 2019, pp. 136–145.

[28] Y. Gu, D. Chen, X. Liu, Suicide possibility scale detection via sina weibo analytics: Preliminary results, Int. J. Environ. Res. Public Health 20 (1) (2023) http://dx.doi.org/10.3390/ijerph20010466.

[29] Z. Dai, H. Zhou, Q. Ba, Y. Zhou, L. Wang, G. Li, Improving depression prediction using a novel feature selection algorithm coupled with context-aware analysis, J. Affect. Disord. 295 (2021) 1040–1048, http://dx.doi.org/10.1016/j.jad.2021.09.001.

[30] J.S. Lara, M.E. Aragón, F.A. González, M. Montes-y Gómez, Deep bag-of-sub-emotions for depression detection in social media, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNAI, vol. 12848, 2021, pp. 60–72, http://dx.doi.org/10.1007/978-3-030-83527-9_5.

[31] M. Trotzek, S. Koitka, C.M. Friedrich, Word embeddings and linguistic metadata at the CLEF 2018 tasks for early detection of depression and anorexia, in: CLEF (Working Notes), 2018.

[32] S. Paul, J.S. Kalyani, T. Basu, Early detection of signs of anorexia and depression over social media using effective machine learning frameworks, in: CEUR Workshop Proceedings, Vol. 2125, 2018.

[33] A. Ranganathan, A. Haritha, D. Thenmozhi, C. Aravindan, Early detection of anorexia using RNN-LSTM and SVM classifiers, in: CEUR Workshop Proceedings, Vol. 2380, 2019.

[34] F.M. Plaza-Del-Arco, P. López-Úbeda, M.C. Díaz-Galiano, L. Alfonso Ureña-López, M. Teresa Martín-Valdivia, Integrating UMLs for early detection of sings of anorexia, in: CEUR Workshop Proceedings, Vol. 2380, 2019.

[35] S.G. Burdisso, M. Errecalde, M. Montes-Y-Gómez, UNSL at erisk 2019: A unified approach for anorexia, self-harm and depression detection in social media, in: CEUR Workshop Proceedings, Vol. 2380, 2019.

[36] H. Yan, E.E. Fitzsimmons-Craft, M. Goodman, M. Krauss, S. Das, P. Cavazos-Rehg, Automatic detection of eating disorder-related social media posts that could benefit from a mental health intervention, Int. J. Eat. Disord. 52 (10) (2019) 1150–1156, http://dx.doi.org/10.1002/eat.23148.

[37] F. Ramiandrisoa, J. Mothe, Early detection of depression and anorexia from social media: A machine learning approach, in: CEUR Workshop Proceedings, Vol. 2621, 2020.

[38] A. Astorino, R. Berti, A. Astorino, V. Bitonti, M. De Marco, V. Feraco, A. Palumbo, F. Porti, I. Zannino, Early detection of eating disorders through machine learning techniques, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, vol. 12096, 2020, pp. 33–39, http://dx.doi.org/10.1007/978-3-030-53552-0_5.

[39] H. Amini, L. Kosseim, Towards explainability in using deep learning for the detection of anorexia in social media, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, vol. 12089, 2020, pp. 225–235, http://dx.doi.org/10.1007/978-3-030-51310-8_21.

[40] S. Sadeh-Sharvit, E. Fitzsimmons-Craft, C. Taylor, E. Yom-Tov, Predicting eating disorders from internet activity, Int. J. Eat. Disord. 53 (9) (2020) 1526–1533, http://dx.doi.org/10.1002/eat.23338.

[41] J. Benítez-Andrades, M. García-Ordás, M. Russo, A. Sakor, L. Fernandes, M. Vidal, Empowering machine learning models with contextual knowledge for enhancing the detection of eating disorders in social media posts, 2023.

[42] H. Espel-Huynh, F. Zhang, J.G. Thomas, J.F. Boswell, H. Thompson-Brenner, A.S. Juarascio, M.R. Lowe, Prediction of eating disorder treatment response trajectories via machine learning does not improve performance versus a simpler regression approach, Int. J. Eat. Disord. 54 (7) (2021) 1250–1259, http://dx.doi.org/10.1002/eat.23510.

[43] P. López-úbeda, F.M. Plaza-Del-arco, M.C. Díaz-Galiano, M.-T. Martín-Valdivia, Article how successful is transfer learning for detecting anorexia on social media? Appl. Sci. (Switzerland) 11 (4) (2021) 1–16, http://dx.doi.org/10.3390/app11041838.

[44] A. Priya, S. Garg, N. Tigga, Predicting anxiety, depression and stress in modern life using machine learning algorithms, Procedia Comput. Sci. 167 (2020) 1258–1267, http://dx.doi.org/10.1016/j.procs.2020.03.442.

[45] M. Trotzek, S. Koitka, C.M. Friedrich, Utilizing neural networks and linguistic metadata for early detection of depression indications in text sequences, IEEE Trans. Knowl. Data Eng. 32 (3) (2020) 588–601, http://dx.doi.org/10.1109/TKDE.2018.2885515.

[46] H.S. Alsagri, M. Ykhlef, Machine learning-based approach for depression detection in twitter using content and activity features, IEICE Trans. Inf. Syst. E103D (8) (2020) 1825–1832, http://dx.doi.org/10.1587/transinf.2020EDP7023.

[47] Y. Ding, X. Chen, Q. Fu, S. Zhong, A depression recognition method for college students using deep integrated support vector algorithm, IEEE Access 8 (2020) 75616–75629, http://dx.doi.org/10.1109/ACCESS.2020.2987523.

[48] R.U. Mustafa, N. Ashraf, F.S. Ahmed, J. Ferzund, B. Shahzad, A. Gelbukh, A multiclass depression detection in social media based on sentiment analysis, Adv. Intell. Syst. Comput. 1134 (2020) 659–662, http://dx.doi.org/10.1007/978-3-030-43020-7_89.

[49] J. Hussain, F.A. Satti, M. Afzal, W.A. Khan, H.S.M. Bilal, M.Z. Ansaar, H.F. Ahmad, T. Hur, J. Bang, J.-I. Kim, G.H. Park, H. Seung, S. Lee, Exploring the dominant features of social media for depression detection, J. Inf. Sci. 46 (6) (2020) 739–759, http://dx.doi.org/10.1177/0165551519860469.

[50] F.M. Shah, F. Ahmed, S.K. Saha Joy, S. Ahmed, S. Sadek, R. Shil, M.H. Kabir, Early depression detection from social network using deep learning techniques, in: 2020 IEEE Region 10 Symposium, TENSYMP 2020, 2020, pp. 823–826, http://dx.doi.org/10.1109/TENSYMP50017.2020.9231008.

[51] X. Wang, S. Chen, T. Li, W. Li, Y. Zhou, J. Zheng, Q. Chen, J. Yan, B. Tang, Depression risk prediction for chinese microblogs via deep-learning methods: Content analysis, JMIR Med. Inform. 8 (7) (2020) http://dx.doi.org/10.2196/17958.

[52] K.-S. Na, S.-E. Cho, Z. Geem, Y.-K. Kim, Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm, Neurosci. Lett. 721 (2020) http://dx.doi.org/10.1016/j.neulet.2020.134804.

[53] N. Shetty, B. Muniyal, A. Anand, S. Kumar, S. Prabhu, Predicting depression using deep learning and ensemble algorithms on raw twitter data, Int. J. Electr. Comput. Eng. 10 (4) (2020) 3751–3756, http://dx.doi.org/10.11591/ijece.v10i4.pp3751-3756.

[54] R. Chiong, G.S. Budhi, S. Dhakal, F. Chiong, A textual-based featuring approach for depression detection using machine learning classifiers and social media texts, Comput. Biol. Med. 135 (2021) http://dx.doi.org/10.1016/j.compbiomed.2021.104499.

[55] R. Chiong, G.S. Budhi, S. Dhakal, E. Cambria, Combining sentiment lexicons and content-based features for depression detection, IEEE Intell. Syst. 36 (6) (2021) 99–105, http://dx.doi.org/10.1109/MIS.2021.3093660.

[56] H. Zogan, I. Razzak, S. Jameel, G. Xu, DepressionNet: Learning multi-modalities with user post summarization for depression detection on social media, in: SIGIR 2021 - Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, 2021, pp. 133–142, http://dx.doi.org/10.1145/3404835.3462938.

[57] K.A. Govindasamy, N. Palanichamy, Depression detection using machine learning techniques on twitter data, in: Proceedings - 5th International Conference on Intelligent Computing and Control Systems, ICICCS 2021, 2021, pp. 960–966, http://dx.doi.org/10.1109/ICICCS51141.2021.9432203.

[58] K. Malviya, B. Roy, S. Saritha, A transformers approach to detect depression in social media, in: Proceedings - International Conference on Artificial Intelligence and Smart Systems, ICAIS 2021, 2021, pp. 718–723, http://dx.doi.org/10.1109/ICAIS50930.2021.9395943.

[59] L. Ren, H. Lin, B. Xu, S. Zhang, L. Yang, S. Sun, Depression detection on reddit with an emotion-based attention network: Algorithm development and validation, JMIR Med. Inform. 9 (7) (2021) http://dx.doi.org/10.2196/28754.

[60] P. Verma, K. Sharma, G.S. Walia, Depression detection among social media users using machine learning, Adv. Intell. Syst. Comput. 1165 (2021) 865–874, http://dx.doi.org/10.1007/978-981-15-5113-0_72.

[61] M. El-Ramly, H. Abu-Elyazid, Y. Mo'men, G. Alshaer, N. Adib, K. Eldeen, M. El-Shazly, CairoDep: Detecting depression in arabic posts using BERT transformers, in: Proceedings - 2021 IEEE 10th International Conference on Intelligent Computing and Information Systems, ICICIS 2021, 2021, pp. 207–212, http://dx.doi.org/10.1109/ICICIS52592.2021.9694178.

[62] F. Azam, M. Agro, M. Sami, M. Abro, A. Dewani, Identifying depression among Twitter users using sentiment analysis, in: 2021 International Conference on Artificial Intelligence, ICAI 2021, 2021, pp. 44–49, http://dx.doi.org/10.1109/ICAI52203.2021.9445271.

[63] R. Martins, J. Almeida, P. Henriques, P. Novais, Identifying depression clues using emotions and AI, in: ICAART 2021 - Proceedings of the 13th International Conference on Agents and Artificial Intelligence, Vol. 2, 2021, pp. 1137–1143.

[64] N. Jagtap, H. Shukla, V. Shinde, S. Desai, V. Kulkarni, Use of ensemble machine learning to detect depression in social media posts, in: Proceedings of the 2nd International Conference on Electronics and Sustainable Communication Systems, ICESC 2021, 2021, pp. 1396–1400, http://dx.doi.org/10.1109/ICESC51422.2021.9532838.

[65] X. Hu, J. Shu, Z. Jin, Depression tendency detection model for weibo users based on Bi-LSTM, in: 2021 IEEE International Conference on Artificial Intelligence and Computer Applications, ICAICA 2021, 2021, pp. 785–790, http://dx.doi.org/10.1109/ICAICA52286.2021.9497931.

[66] S. Hemtanon, S. Aekwarangkoon, N. Kittiphattanabawon, Detection of depression-positive thai facebook users using posts and their usage behavior, Lect. Notes Netw. Syst. 251 (2021) 77–87, http://dx.doi.org/10.1007/978-3-030-79757-7_8.

[67] S. Shekerbekova, M. Yerekesheva, L. Tukenova, K. Turganbay, Z. Kozhamkulova, B. Omarov, Applying machine learning to detect depression-related texts on social networks, Commun. Comput. Inf. Sci. 1393 (2021) 161–169, http://dx.doi.org/10.1007/978-981-16-3660-8_15.

[68] Y. Liu, X. Lu, D. Shi, J. Yuan, Depression severity level classification using multitask learning of gender recognition, in: 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA ASC 2021 - Proceedings, 2021, pp. 1317–1322.

[69] H. Kour, M.K. Gupta, An hybrid deep learning approach for depression prediction from user tweets using feature-rich CNN and bi-directional LSTM, Multimedia Tools Appl. 81 (17) (2022) 23649–23685, http://dx.doi.org/10.1007/s11042-022-12648-y.

[70] H. Zogan, I. Razzak, X. Wang, S. Jameel, G. Xu, Explainable depression detection with multi-aspect features using a hybrid deep learning model on social media, World Wide Web 25 (1) (2022) 281–304, http://dx.doi.org/10.1007/s11280-021-00992-2.

[71] M. Nayan, M. Uddin, M. Hossain, M. Alam, M. Zinnia, I. Haq, M. Rahman, R. Ria, M. Haq Methun, Comparison of the performance of machine learning-based algorithms for predicting depression and anxiety among university students in Bangladesh: A result of the first wave of the COVID-19 pandemic, Asian J. Soc. Health Behav. 5 (2) (2022) 75–84, http://dx.doi.org/10.4103/shb.shb_38_22.

[72] D. Musleh, T. Alkhales, R. Almakki, S. Alnajim, S. Almarshad, R. Alhasaniah, S. Aljameel, A. Almuqhim, Twitter arabic sentiment analysis to detect depression using machine learning, Comput. Mater. Contin. 71 (2) (2022) 3463–3477, http://dx.doi.org/10.32604/cmc.2022.022508.

[73] L. Tong, Z. Liu, Z. Jiang, F. Zhou, L. Chen, J. Lyu, X. Zhang, Q. Zhang, A. Sadka, Y. Wang, L. Li, H. Zhou, Cost-sensitive boosting pruning trees for depression detection on Twitter, IEEE Trans. Affect. Comput. (2022) http://dx.doi.org/10.1109/TAFFC.2022.3145634.

[74] J. Angskun, S. Tipprasert, T. Angskun, Big data analytics on social networks for real-time depression detection, J. Big Data 9 (1) (2022) http://dx.doi.org/10.1186/s40537-022-00622-2.

[75] R.J. Lia, A.B. Siddikk, F. Muntasir, S.S.M.M. Rahman, N. Jahan, Depression detection from social media using Twitter's tweet, Stud. Comput. Intell. 994 (2022) 209–226, http://dx.doi.org/10.1007/978-3-030-87954-9_9.

[76] P. Bhat, A. Anuse, R. Kute, R. Bhadade, P. Purnaye, Mental health analyzer for depression detection based on textual analysis, J. Adv. Inf. Technol. 13 (1) (2022) 67–77, http://dx.doi.org/10.12720/jait.13.1.67-77.

[77] A. Pramanik, M.H.I. Bijoy, M.S. Rahman, Depression-level prediction during COVID-19 pandemic among the people of Bangladesh using ensemble technique: MIRF stacking and MIRF voting, Lect. Notes Netw. Syst. 437 (2022) 71–87, http://dx.doi.org/10.1007/978-981-19-2445-3_6.

[78] W.-Y. Wang, Y.-C. Tang, W.-W. Du, W.-C. Peng, NYCU TWD@LT-EDI-ACL2022: Ensemble models with VADER and contrastive learning for detecting signs of depression from social media, in: LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop, 2022, pp. 136–139.

[79] N. Reseena Mol, S. Veni, A stacked ensemble technique with glove embedding model for depression detection from tweets, Indian J. Comput. Sci. Eng. 13 (2) (2022) 586–595, http://dx.doi.org/10.21817/indjcse/2022/v13i2/221302088.

[80] Y. Wang, Z. Wang, C. Li, Y. Zhang, H. Wang, Online social network individual depression detection using a multitask heterogenous modality fusion approach, Inform. Sci. 609 (2022) 727–749, http://dx.doi.org/10.1016/j.ins.2022.07.109.

[81] K. Srinath, K. Kiran, S. Pranavi, M. Amrutha, P.D. Shenoy, K. Venugopal, Prediction of depression, anxiety and stress levels using dass-42, in: 2022 IEEE 7th International Conference for Convergence in Technology, I2CT 2022, 2022, http://dx.doi.org/10.1109/I2CT54291.2022.9824087.

[82] H. Kour, M. Gupta, Depression and suicide prediction using natural language processing and machine learning, Lect. Notes Netw. Syst. 370 (2022) 117–128, http://dx.doi.org/10.1007/978-981-16-8664-1_11.

[83] S.S. Nair, A. Ashok, R. Divya Pai, A. Hari Narayanan, Detection of Early Depression Signals Using Social Media Sentiment Analysis on Big Data, in: Lecture Notes on Data Engineering and Communications Technologies, vol. 75, 2022, pp. 413–422, http://dx.doi.org/10.1007/978-981-16-3728-5_31.

[84] S. Esackimuthu, H. Shruthi, R. Sivanaiah, S. Angel Deborah, R. Sakaya Milton, T. Mirnalinee, SSN_MLRG3 LTE-DI-ACL2022 depression detection system from social media text using transformer models, in: LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop, 2022, pp. 196–199.

[85] I. Tavchioski, B. Koloski, B. Škrlj, S. Pollak, E8-IJS@LT-EDI-ACL2022 - BERT, automl and knowledge-graph backed detection of depression, in: LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop, 2022, pp. 251–257.

[86] H. Sharen, R. Rajalakshmi, DLRG@LT-EDI-ACL2022:Detecting signs of depression from social media using XGBoost method, in: LTEDI 2022 - 2nd Workshop on Language Technology for Equality, Diversity and Inclusion, Proceedings of the Workshop, 2022, pp. 346–349.

[87] K. Milintsevich, K. Sirts, G. Dias, Towards automatic text-based estimation of depression through symptom prediction, Brain Inform. 10 (1) (2023) http://dx.doi.org/10.1186/s40708-023-00185-9.

[88] S. Dalal, S. Jain, M. Dave, An investigation of data requirements for the detection of depression from social media posts, Recent Pat. Eng. 17 (3) (2023) http://dx.doi.org/10.2174/1872212117666220812110956.

[89] D.S. Khafaga, M. Auvdaiappan, K. Deepa, M. Abouhawwash, F.K. Karim, Deep learning for depression detection using Twitter data, Intell. Autom. Soft Comput. 36 (2) (2023) 1301–1313, http://dx.doi.org/10.32604/iasc.2023.033360.

[90] Z.N. Vasha, B. Sharma, I.J. Esha, J. Al Nahian, J.A. Polin, Depression detection in social media comments data using machine learning algorithms, Bull. Electr. Eng. Inform. 12 (2) (2023) 987–996, http://dx.doi.org/10.11591/eei.v12i2.4182.

[91] P. Samanta, P. Kumar, S. Dutta, M. Chatterjee, D. Sarkar, Depression Detection from Twitter Data Using Two Level Multi-modal Feature Extraction, in: Lecture Notes on Data Engineering and Communications Technologies, vol. 137, 2023, pp. 451–465, http://dx.doi.org/10.1007/978-981-19-2600-6_32.

[92] W. Pan, X. Wang, W. Zhou, B. Hang, L. Guo, Linguistic analysis for identifying depression and subsequent suicidal ideation on weibo: Machine learning approaches, Int. J. Environ. Res. Public Health 20 (3) (2023) http://dx.doi.org/10.3390/ijerph20032688.

[93] M. Tadesse, H. Lin, B. Xu, L. Yang, Detection of suicide ideation in social media forums using deep learning, Algorithms 13 (1) (2020) http://dx.doi.org/10.3390/a13010007.

[94] J. Gradus, A. Rosellini, E. Horváth-Puhó, A. Street, I. Galatzer-Levy, T. Jiang, T. Lash, H. Sørensen, Prediction of sex-specific suicide risk using machine learning and single-payer health care registry data from Denmark, JAMA Psychiatry 77 (1) (2020) 25–34, http://dx.doi.org/10.1001/jamapsychiatry.2019.2905.

[95] C. Su, R. Aseltine, R. Doshi, K. Chen, S. Rogers, F. Wang, Machine learning for suicide risk prediction in children and adolescents with electronic health records, Transl. Psychiatry 10 (1) (2020) http://dx.doi.org/10.1038/s41398-020-01100-0.

[96] D. Ramírez-Cifuentes, A. Freire, R. Baeza-Yates, J. Puntí, P. Medina-Bravo, D. Velazquez, J. Gonfaus, J. Gonzàlez, Detection of suicidal ideation on social media: Multimodal, relational, and behavioral analysis, J. Med. Internet Res. 22 (7) (2020) http://dx.doi.org/10.2196/17758.

[97] M. Miché, E. Studerus, A. Meyer, A. Gloster, K. Beesdo-Baum, H.-U. Wittchen, R. Lieb, Prospective prediction of suicide attempts in community adolescents and young adults, using regression methods and machine learning, J. Affect. Disord. 265 (2020) 570–578, http://dx.doi.org/10.1016/j.jad.2019.11.093.

[98] G.-M. Lin, M. Nagamine, S.-N. Yang, Y.-M. Tai, C. Lin, H. Sato, Machine learning based suicide ideation prediction for military personnel, IEEE J. Biomed. Health Inf. 24 (7) (2020) 1907–1916, http://dx.doi.org/10.1109/JBHI.2020.2988393.

[99] F. Shah, F. Haque, R. Un Nur, S. Al Jahan, Z. Mamud, A hybridized feature extraction approach to suicidal ideation detection from social media post, in: 2020 IEEE Region 10 Symposium, TENSYMP 2020, 2020, pp. 985–988, http://dx.doi.org/10.1109/TENSYMP50017.2020.9230733.

[100] J. Pestian, D. Santel, M. Sorter, U. Bayram, B. Connolly, T. Glauser, M. DelBello, S. Tamang, K. Cohen, A machine learning approach to identifying changes in suicidal language, Suicide Life-Threat. Behav. 50 (5) (2020) 939–947, http://dx.doi.org/10.1111/sltb.12642.

[101] S. Rabani, Q. Khan, A. Ud Din Khanday, Detection of suicidal ideation on Twitter using machine learning & ensemble approaches, Baghdad Sci. J. 17 (4) (2020) 1328–1339, http://dx.doi.org/10.21123/bsj.2020.17.4.1328.

[102] R. Acuña Caicedo, J. Gómez Soriano, H. Melgar Sasieta, Assessment of supervised classifiers for the task of detecting messages with suicidal ideation, Heliyon 6 (8) (2020) http://dx.doi.org/10.1016/j.heliyon.2020.e04412.

[103] F. Tsui, L. Shi, V. Ruiz, N. Ryan, C. Biernesser, S. Iyengar, C. Walsh, D. Brent, Natural language processing and machine learning of electronic health records for prediction of first-time suicide attempts, JAMIA Open 4 (1) (2021) http://dx.doi.org/10.1093/jamiaopen/ooab011.

[104] M. Cusick, P. Adekkanattu, T. Campion Jr., E. Sholle, A. Myers, S. Banerjee, G. Alexopoulos, Y. Wang, J. Pathak, Using weak supervision and deep learning to classify clinical notes for identification of current suicidal ideation, J. Psychiatr. Res. 136 (2021) 95–102, http://dx.doi.org/10.1016/j.jpsychires.2021.01.052.

[105] J. Gradus, A. Rosellini, E. Horváth-Puhó, T. Jiang, A. Street, I. Galatzer-Levy, T. Lash, H. Sørensen, Predicting sex-specific nonfatal suicide attempt risk using machine learning and data from danish national registries, Am. J. Epidemiol. 190 (12) (2021) 2517–2527, http://dx.doi.org/10.1093/aje/kwab112.

[106] D. Lekkas, R. Klein, N. Jacobson, Predicting acute suicidal ideation on instagram using ensemble machine learning models, Internet Interv. 25 (2021) http://dx.doi.org/10.1016/j.invent.2021.100424.

[107] N. Wang, F. Luo, Y. Shivtare, V. Badal, K. Subbalakshmi, R. Chandramouli, E. Lee, Learning models for suicide prediction from social media posts, in: Computational Linguistics and Clinical Psychology: Improving Access, CLPsych 2021 - Proceedings of the 7th Workshop, in Conjunction with NAACL 2021, 2021, pp. 87–92.

[108] S. Kim, H.-K. Lee, K. Lee, Detecting suicidal risk using MMPI-2 based on machine learning algorithm, Sci. Rep. 11 (1) (2021) http://dx.doi.org/10.1038/s41598-021-94839-5.

[109] S.-E. Cho, Z. Geem, K.-S. Na, Development of a suicide prediction model for the elderly using health screening data, Int. J. Environ. Res. Public Health 18 (19) (2021) http://dx.doi.org/10.3390/ijerph181910150.

[110] G. Harman, D. Kliamovich, A. Morales, S. Gilbert, D. Barch, M. Mooney, S. Ewing, D. Fair, B. Nagel, Prediction of suicidal ideation and attempt in 9 and 10 year-old children using transdiagnostic risk features, PLoS ONE 16 (5 May) (2021) http://dx.doi.org/10.1371/journal.pone.0252114.

[111] A. Haque, V. Reddi, T. Giallanza, Deep learning for suicide and depression identification with unsupervised label correction, in: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), in: LNCS, vol. 12895, 2021, pp. 436–447, http://dx.doi.org/10.1007/978-3-030-86383-8_35.

[112] Z. Xu, Y. Xu, F. Cheung, M. Cheng, D. Lung, Y. Law, B. Chiang, Q. Zhang, P. Yip, Detecting suicide risk using knowledge-aware natural language processing and counseling service data, Soc. Sci. Med. 283 (2021) http://dx.doi.org/10.1016/j.socscimed.2021.114176.

[113] J. Edgcomb, T. Shaddox, G. Hellemann, J.O. Brooks III, Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness, J. Psychiatr. Res. 136 (2021) 515–521, http://dx.doi.org/10.1016/j.jpsychires.2020.10.024.

[114] S. Gollapalli, G. Zagatti, S.-K. Ng, Suicide risk prediction by tracking self-harm aspects in tweets: NUS-IDS at the clpsych 2021 shared task, in: Computational Linguistics and Clinical Psychology: Improving Access, CLPsych 2021 - Proceedings of the 7th Workshop, in Conjunction with NAACL 2021, 2021, pp. 93–98.

[115] M. Sharma, B. Pant, V. Singh, S. Kumar, STP:Suicidal tendency prediction among the youth using social network data, Adv. Intell. Syst. Comput. 1162 (2021) 161–169, http://dx.doi.org/10.1007/978-981-15-4851-2_17.

[116] R. Haque, N. Islam, M. Islam, M. Ahsan, A comparative analysis on suicidal ideation detection using NLP, machine, and deep learning, Technologies 10 (3) (2022) http://dx.doi.org/10.3390/technologies10030057.

[117] M. Nock, A. Millner, E. Ross, C. Kennedy, M. Al-Suwaidi, Y. Barak-Corren, V. Castro, F. Castro-Ramirez, T. Lauricella, N. Murman, M. Petukhova, S. Bird, B. Reis, J. Smoller, R. Kessler, Prediction of suicide attempts using clinician assessment, patient self-report, and electronic health records, JAMA Netw. Open 5 (1) (2022) http://dx.doi.org/10.1001/jamanetworkopen.2021.44373.

[118] J. Cohen, J. Wright-Berryman, L. Rohlfs, D. Trocinski, L. Daniel, T. Klatt, Integration and validation of a natural language processing machine learning suicide risk prediction model based on open-ended interview language in the emergency department, Front. Digit. Health 4 (2022) http://dx.doi.org/10.3389/fdgth.2022.818705.

[119] V. Rozova, K. Witt, J. Robinson, Y. Li, K. Verspoor, Detection of self-harm and suicidal ideation in emergency department triage notes, J. Am. Med. Inform. Assoc. 29 (3) (2022) 472–480, http://dx.doi.org/10.1093/jamia/ocab261.

[120] M. Chatterjee, P. Samanta, P. Kumar, D. Sarkar, Suicide ideation detection using multiple feature analysis from Twitter data, in: 2022 IEEE Delhi Section Conference, DELCON 2022, 2022, http://dx.doi.org/10.1109/DELCON54057.2022.9753295.

[121] L. Grendas, L. Chiapella, D. Rodante, F. Daray, Comparison of traditional model-based statistical methods with machine learning for the prediction of suicide behaviour, J. Psychiatr. Res. 145 (2022) 85–91, http://dx.doi.org/10.1016/j.jpsychires.2021.11.029.

[122] T. Aldhyani, S. Alsubari, A. Alshebami, H. Alkahtani, Z. Ahmed, Detecting and analyzing suicidal ideation on social media using deep learning and machine learning models, Int. J. Environ. Res. Public Health 19 (19) (2022) http://dx.doi.org/10.3390/ijerph191912635.

[123] S. Renjith, A. Abraham, S. Jyothi, L. Chandran, J. Thomson, An ensemble deep learning technique for detecting suicidal ideation from posts in social media platforms, J. King Saud Univ. - Comput. Inf. Sci. 34 (10) (2022) 9564–9575, http://dx.doi.org/10.1016/j.jksuci.2021.11.010.

[124] S. Colic, J. He, J. Richardson, K. Cyr, J. Reilly, G. Hasey, A machine learning approach to identification of self-harm and suicidal ideation among military and police veterans, J. Mil. Veteran Fam. Health 8 (1) (2022) 56–67, http://dx.doi.org/10.3138/JMVFH-2021-0035.

[125] H. Metzler, H. Baginski, T. Niederkrotenthaler, D. Garcia, Detecting potentially harmful and protective suicide-related content on Twitter: Machine learning approach, J. Med. Internet Res. 24 (8) (2022) http://dx.doi.org/10.2196/34705.

[126] J. Martinez-Romo, B. Reneses, I. Martínez-Capella, L. Araujo, J. Sevilla-Llewellyn-Jones, G. Seara-Aguilar, Detecting signs of non-suicidal self-injury in psychiatric medical reports using language analysis [detección de indicios de autolesiones no suicidas en informes médicos de psiquiatría mediante el análisis del lenguaje], Proces. Leng. Nat. 69 (2022) 129–140, http://dx.doi.org/10.26342/2022-69-11.

[127] A. Mbarek, S. Jamoussi, A. Hamadou, An across online social networks profile building approach: Application to suicidal ideation detection, Future Gener. Comput. Syst. 133 (2022) 171–183, http://dx.doi.org/10.1016/j.future.2022.03.017.

[128] J. Li, X. Chen, Z. Lin, K. Yang, H. Leong, N. Yu, Q. Li, Suicide risk level prediction and suicide trigger detection: A benchmark dataset, HKIE Trans. Hong Kong Inst. Eng. 29 (4) (2022) 268–282, http://dx.doi.org/10.33430/V29N4THIE-2022-0031.

[129] V. Desu, N. Komati, S. Lingamaneni, F. Shaik, Suicide and depression detection in social media forums, Smart Innov. Syst. Technol. 283 (2022) 263–270, http://dx.doi.org/10.1007/978-981-16-9705-0_26.

[130] R. Kancharapu, A. Srinagesh, M. Bhanusridhar, Prediction of human suicidal tendency based on social media using recurrent neural networks through LSTM, in: Proceedings - 2022 International Conference on Computing, Communication and Power Technology, IC3P 2022, 2022, pp. 123–128, http://dx.doi.org/10.1109/IC3P52835.2022.00033.

[131] O. Benlaaraj, I. El Jaafari, A. Ellahyani, I. Boutaayamou, Prediction of suicidal ideation in a new arabic annotated dataset, in: Proceedings - 2022 9th International Conference on Wireless Networks and Mobile Communications, WINCOM 2022, 2022, http://dx.doi.org/10.1109/WINCOM55661.2022.9966481.

[132] A. Chadha, A. Gupta, Y. Kumar, Suicidal ideation detection on social media: A machine learning approach, in: Proceedings of International Conference on Technological Advancements in Computational Sciences, ICTACS 2022, 2022, pp. 685–688, http://dx.doi.org/10.1109/ICTACS56270.2022.9988722.

[133] N. Nordin, Z. Zainol, M. Mohd Noor, C. Fong, Explainable machine learning models for suicidal behavior prediction, in: ACM International Conference Proceeding Series, 2022, pp. 118–123, http://dx.doi.org/10.1145/3545729.3545754.

[134] Y. Lim, M. Lee, Y. Loo, Towards a machine learning framework for suicide ideation detection in Twitter, in: 2022 3rd International Conference on Artificial Intelligence and Data Sciences: Championing Innovations in Artificial Intelligence and Data Sciences for Sustainable Future, AiDAS 2022 - Proceedings, 2022, pp. 153–157, http://dx.doi.org/10.1109/AiDAS56890.2022.9918782.

[135] R. Punithavathi, S. Thenmozhi, R. Jothilakshmi, V. Ellappan, I. Tahzib Ul, Suicide ideation detection of covid patients using machine learning algorithm, Comput. Syst. Sci. Eng. 45 (1) (2023) 247–261, http://dx.doi.org/10.32604/csse.2023.025972.

[136] B. Priyamvada, S. Singhal, A. Nayyar, R. Jain, P. Goel, M. Rani, M. Srivastava, Stacked CNN - LSTM approach for prediction of suicidal ideation on social media, Multimedia Tools Appl. (2023) http://dx.doi.org/10.1007/s11042-023-14431-z.

[137] H. Lu, A. Barrett, A. Pierce, J. Zheng, Y. Wang, C. Chiang, C. Rakovski, Predicting suicidal and self-injurious events in a correctional setting using AI algorithms on unstructured medical notes and structured data, J. Psychiatr. Res. 160 (2023) 19–27, http://dx.doi.org/10.1016/j.jpsychires.2023.01.032.

[138] Y. Barak-Corren, V. Castro, S. Javitt, M. Nock, J. Smoller, B. Reis, Improving risk prediction for target subpopulations: Predicting suicidal behaviors among multiple sclerosis patients, PLoS ONE 18 (2 February) (2023) http://dx.doi.org/10.1371/journal.pone.0277483.

[139] A.-M. Bucur, A. Cosma, L.P. Dinu, Early risk detection of pathological gambling, self-harm and depression using BERT, in: CLEF (Working Notes), 2021, pp. 938–949.

[140] C. Wang, M. Zhang, F. Shi, P. Xue, Y. Li, A hybrid multimodal data fusion-based method for identifying gambling websites, Electronics (Switzerland) 11 (16) (2022) http://dx.doi.org/10.3390/electronics11162489.

[141] C. Wang, P. Xue, M. Zhang, M. Hu, Identifying gambling websites with co-training, in: Proceedings of the International Conference on Software Engineering and Knowledge Engineering, SEKE, 2022, pp. 598–603, http://dx.doi.org/10.18293/SEKE2022-106.

[142] T.-A. Dumitraşcu, CLEF erisk 2022: Detecting early signs of pathological gambling using ML and DL models with dataset chunking, in: CEUR Workshop Proceedings, Vol. 3180, 2022, pp. 883–893.

[143] A.M. Mármol-Romero, S.M. Jiménez-Zafra, F.M. Plaza-del Arco, M.D. Molina-González, M.-T. Martín-Valdivia, A. Montejo-Ráez, SINAI at erisk@ CLEF 2022: Approaching early detection of gambling and eating disorders with natural language processing, in: CLEF (Working Notes), 2022, pp. 961–971.

[144] B. Perrot, J.-B. Hardouin, E. Thiabaud, A. Saillard, M. Grall-Bronnec, G. Challet-Bouju, Development and validation of a prediction model for online gambling problems based on players' account data, J. Behav. Addict. 11 (3) (2022) 874–889, http://dx.doi.org/10.1556/2006.2022.00063.

[145] A.-M. Bucur, A. Cosma, L.P. Dinu, P. Rosso, An end-to-end set transformer for user-level classification of depression and gambling disorder, in: CLEF (Working Notes), 2022, pp. 851–863.

[146] H. Fabregat, A. Duque, L. Araujo, J. Martinez-Romo, Uned-nlp at erisk 2022: Analyzing gambling disorders in social media using approximate nearest neighbors, in: CLEF (Working Notes), 2022, pp. 894–904.

[147] H. Srivastava, N. Lijin, S. Sruthi, T. Basu, NLP-IISERB@ erisk2022: Exploring the potential of bag of words, document embeddings and transformer based framework for early prediction of eating disorder, depression and pathological gambling over social media, in: CLEF (Working Notes), 2022, pp. 972–986.

[148] S. Kairouz, J.-M. Costes, W. Murch, P. Doray-Demers, C. Carrier, V. Eroukmanoff, Enabling new strategies to prevent problematic online gambling: A machine learning approach for identifying at-risk online gamblers in France, Int. Gambl. Stud. (2023) http://dx.doi.org/10.1080/14459795.2022.2164042.

[149] M. Kraus, P. Seldschopf, W. Minker, Towards the development of a trustworthy chatbot for mental health applications, in: J. Lokoč, T. Skopal, K. Schoeffmann, V. Mezaris, X. Li, S. Vrochidis, I. Patras (Eds.), MultiMedia Modeling, Springer International Publishing, Cham, 2021, pp. 354–366.

[150] J.Á. González, L.-F. Hurtado, F. Pla, Self-attention for Twitter sentiment analysis in spanish, J. Intell. Fuzzy Systems 39 (2) (2020) 2165–2175.

[151] J.Á. González, L.-F. Hurtado, F. Pla, Transformer based contextualization of pre-trained word embeddings for irony detection in Twitter, Inf. Process. Manage. 57 (4) (2020) 102262, http://dx.doi.org/10.1016/j.ipm.2020.102262.

[152] D. Ramírez-Cifuentes, A. Freire, R. Baeza-Yates, J. Puntí, P. Medina-Bravo, D.A. Velazquez, J.M. Gonfaus, J. Gonzàlez, et al., Detection of suicidal ideation on social media: multimodal, relational, and behavioral analysis, J. Med. Internet Res. 22 (7) (2020) e17758.

[153] D. Ramírez-Cifuentes, A. Freire, R. Baeza-Yates, N. Sanz Lamora, A. Álvarez, A. González-Rodríguez, M. Lozano Rochel, R. Llobet Vives, D.A. Velazquez, J.M. Gonfaus, J. Gonzàlez, Characterization of anorexia nervosa on social media: Textual, visual, relational, behavioral, and demographical analysis, J. Med. Internet Res. 23 (7) (2021) e25925, http://dx.doi.org/10.2196/25925.

[154] M. Laurent, Project hatemeter: helping NGOs and social science researchers to analyze and prevent anti-muslim hate speech on social media, Procedia Comput. Sci. 176 (2020) 2143–2153.

[155] M. Garg, C. Saxena, U. Naseem, B.J. Dorr, NLP as a lens for causal analysis and perception mining to infer mental health on social media, 2023, arXiv preprint arXiv:2301.11004.
[156] M.R. Haque, S. Rubya, An overview of chatbot-based mobile mental health apps: insights from app description and user reviews, JMIR mHealth uHealth 11 (1) (2023) e44838.
[157] A. Abilkaiyrkyzy, F. Laamarti, M. Hamdi, A. El Saddik, Dialogue system for early mental illness detection: Towards a digital twin solution, IEEE Access (2024).
[158] P. Kaywan, K. Ahmed, A. Ibaida, Y. Miao, B. Gu, Early detection of depression using a conversational AI bot: A non-clinical trial, PLoS ONE 18 (2) (2023) e0279743.

Arturo Montejo-Ráez is an Associate Professor at the University of Jaén (Spain). He holds a European Ph.D. in Computer Science from the University of Granada. He is a Spanish Society for Natural Language Processing member and founder of the spin-off Yottacode S.L. His scientific activity focuses on deep learning for NLP, analyzing stereotype bias in language models and explainability in foundational language models. He is also the Lead Researcher of several regional and national projects in Spain, some related to the early detection of mental disorders from texts.

M. Dolores Molina-González received the B.S. degree in telecommunication engineering from the University of Valencia (Spain), in 2005 and the Ph.D. degree in computer sciences engineering from University of Jaén (Spain), in 2014. She is associate professor in the Department of Engineering of Telecommunication at University of Jaén. Her current research interests include Natural Language Processing, Sentiment and Emotion Analysis, Offensive Language detection, and Mental Disorder Detection. In addition, she is author or co-author of more than 30 scientific publications. She is a technical reviewer for several journals. Also, she is a member of Spanish Society for Natural Language Processing (SEPLN).

Salud María Jiménez-Zafra holds a Ph.D. in Computer Science (2019), a Specialist in Internet Information Processing (2014), a Master Degree in Computer Science (2013) and a Bachelors' Degree in Computer Science and Statistics (2011) from Universidad de Jaén. She is author of more than 75 research works and 2 books. She has organized 22 shared tasks, workshops and conferences of international relevance in the NLP area. The scientific contributions she has carried out have led her to be honored with 9 research awards. Her research is currently focused on negation processing, resource generation, hate/hope speech identification and mental disorders detection.

Miguel Ángel García-Cumbreras, Ph.D. degree in computer sciences engineering from University of Jaén (Spain), in 2009. With more than 20 years of teaching and research experience, he is the author and co-author of more than 55 papers in journals, 50 contributions to congresses, participated in 40 R+D+i projects and grants, and has 23 agreements with companies. He is a partner promoter of two university spinoff and currently shares his work in the Vice-Rectorate for Lifelong Learning, Educational Technologies and Teaching Innovation. The main current research topics on which he is working are Fake News, Hope Speech and Mental Disorder Detection.

Luis Joaquin Garcia-López is Professor of Psychology at the University of Jaén. His lines of work are the promotion of health and emotional well-being of young people as well as the early detection and intervention of young people with (or at risk of) emotional problems. The candidate has almost 100 indexed publications. His H-index places him at Level A. He has been part of more than 15 research actions funded by national and international entities, emphasizing work with other disciplines. He serves as a reviewer of research projects at national, regional and international level.