Acadlore Transactions on AI and Machine Learning
https://www.acadlore.com/journals/ATAIML

Natural Language Processing for PTSD Detection: A Systematic Review and Bibliometric Analysis

Engin Seven*, Eylem Yucel

Department of Computer Engineering, Istanbul University—Cerrahpasa, 34320 Istanbul, Turkey

* Correspondence: Engin Seven (eseven41@gmail.com)

Received: 09-07-2025 **Revised:** 11-15-2025 **Accepted:** 11-24-2025

Citation: E. Seven, and E. Yucel, "Natural language processing for PTSD detection: A systematic review and bibliometric analysis," *Acadlore Trans. Mach. Learn.*, vol. 4, no. 4, pp. 273–300, 2025. https://doi.org/10.56578/ataiml040404.

Abstract: Post-traumatic stress disorder (PTSD) has been recognized as a critical global mental health challenge, and the application of natural language processing (NLP) has emerged as a promising approach for its detection and management. In this study, a systematic review was conducted to evaluate the quality, quantity, and consistency of research investigating the role of NLP in PTSD detection. Through this process, prior research was consolidated, methodological gaps were identified, and a conceptual framework was formulated to guide future investigations. To complement the systematic review, a bibliometric analysis was performed to map the intellectual landscape, assess publication trends, and visualize research networks within this domain. The systematic review involved a structured search across ScienceDirect, IEEE Xplore, PubMed, and Web of Science, resulting in the retrieval of 328 records. After rigorous screening, 56 studies were included in the final synthesis. Separately, a bibliometric analysis was conducted on 4,138 publications obtained from the Web of Science database. The findings highlight that NLP methods not only enhance the detection of PTSD but also support the development of personalized treatment strategies. Ethical and security considerations were also identified as pressing concerns requiring further attention. The results of this study underscore the significance of NLP in advancing PTSD research and emphasize its potential to transform mental health services. By identifying trends, challenges, and opportunities, this study provides a foundation for future research aimed at strengthening the role of NLP in clinical practice and mental health policy.

Keywords: Post-traumatic stress disorder; Natural language processing; Health system; Systematic review; Bibliometric analysis

1 Introduction

The World Mental Health Report 2022 issued by the World Health Organization (WHO) highlights significant global disparities in mental health care systems and outcomes. It is estimated that one in every eight individuals worldwide experiences a mental health disorder [1]. Furthermore, individuals suffering from such disorders are at an increased risk of premature mortality. Among children aged 2 to 17, high levels of emotional, physical, and sexual violence have been reported, which significantly elevates the risk of developing mental disorders in later stages of life [2]. Childhood represents a critical period in which adverse experiences can shape emotional and psychological development. These adverse experiences are linked to a heightened risk of disorders such as depression, anxiety, and PTSD [3]. Mental health conditions also pose substantial social and economic burdens. For instance, the global economic cost of mental disorders was estimated at USD 2.5 trillion in 2010, with projections reaching nearly USD 5 trillion by 2030 [4]. Environmental and societal stressors, including natural disasters (e.g., earthquakes and floods) and human-made crises (e.g., terrorism), are recognized as major contributors to the onset of mental health issues. The COVID-19 pandemic further exacerbated this burden; mental health disorders such as PTSD, anxiety, and depression increased by approximately 25% during the first year of the outbreak, placing additional strain on healthcare infrastructures worldwide [5].

PTSD is a complex clinical condition characterized by symptoms, including hypervigilance, avoidance behaviors, cognitive and mood disturbances, and heightened physiological reactivity following a traumatic event [6]. Prevalence rates vary across populations and contexts; for example, a study conducted in the United States and Canada reported PTSD prevalence between 6% and 9% [7]. Untreated PTSD can significantly impair social functioning and lead to comorbid psychiatric or physical health conditions. Common symptoms include attention deficits, disrupted sleep, emotional dysregulation, and anger outbursts, which collectively contribute to poor quality of life and may escalate to suicidal ideation. NLP, a subfield of artificial intelligence (AI), has emerged as a promising tool for analyzing unstructured text data, including clinical notes, survey responses, and social media content. NLP encompasses technologies such as speech and emotion recognition, and semantic analysis. Its applications in healthcare are increasingly diverse, spanning radiology reports, biomedical terminology mapping, and physician documentation. One of the earliest examples of human–machine interaction via NLP is the ELIZA program developed in 1966 [8]. In modern implementations, NLP often integrates machine learning (ML) and deep learning (DL) to convert raw text into computable features for classification or prediction tasks. WHO's 2022 report highlights significant global disparities in access to mental health services, particularly regarding limited availability of psychotherapy and shortages in trained professionals [1]. In this context, AI and NLP technologies have emerged as supportive tools for the early detection and monitoring of mental health symptoms, especially in resource-limited settings. This study aims to systematically review the current literature on how NLP techniques may contribute to addressing these systemic challenges in mental healthcare delivery.

The goal of this study is to systematically review and evaluate how NLP techniques have been employed to support the identification and classification of PTSD-related symptoms. This review also explores methodological trends, research gaps, and the potential integration of NLP-based tools into clinical workflows.

The research questions addressed in this study are as follows:

1. How have NLP techniques evolved specifically in the detection of PTSD?
2. What are the focal areas of NLP-based trauma research aimed at identifying PTSD?
3. Which NLP approaches have been used specifically for PTSD detection, and how were they applied?
4. What are the methodological challenges in using NLP for PTSD classification?
5. Can AI-based PTSD detection approaches be realistically integrated into healthcare systems?

The structure of this study is organized as follows: Section 2 outlines the methodology; Section 3 discusses PTSD and associated research methods; Section 4 reviews NLP methods used in PTSD-related studies; and Section 5 presents the results and discussion.

2 Research Methodology

In this systematic review, the primary objective was not to establish a direct diagnostic framework for PTSD using NLP techniques, but rather to assess the potential of NLP methods for detecting PTSD-related symptoms in unstructured text data. The focus was on identifying and extracting key symptoms and linguistic markers associated with PTSD, as reported in the literature. As such, the reviewed studies predominantly have utilized NLP for the detection and classification of PTSD symptoms rather than formal diagnostic assessment based on clinical criteria. The number of studies conducted in the field of mental health is increasing on a daily basis. A bibliometric analysis and systematic literature review in this field are anticipated to serve as a roadmap for researchers. A research project specifically focused on PTSD in the field of mental health can provide guidance on general trends, potential issues, and solutions. To this end, bibliometric analysis can be used to examine research trends and structural patterns in the literature [9]. In order to achieve this objective, a comprehensive and systematic analysis of a substantial corpus of studies was conducted. This analysis generated a map of the trends and latest developments. The general process of the literature review in this study is illustrated in Figure 1.

2.1 Bibliometric Analysis Method

Bibliometric analysis has been used for several decades to examine patterns in scientific publications [9]. The bibliometric analysis of publications, including articles, papers and books on a specific subject, allowed for the identification of emerging trends, collaborations and research components. Furthermore, it was employed to examine the intellectual structure of a subject area that is of interest to researchers. The existence of extensive databases such as Scopus and Web of Science has greatly enhanced the accessibility of a vast array of publications. Consequently, researchers have been able to conduct bibliometric analysis on big data in a more effective manner. The field of NLP has undergone significant advancements over the past decade. These developments have resulted in the production of a substantial corpus of articles, papers, and books. It is therefore evident that there is a growing need for bibliometric analysis in the field of NLP. Selecting the most appropriate parameters for a bibliometric analysis within a specific field is a crucial aspect of the process. To illustrate, the number of citations can be employed to determine the relative importance of each publication. Furthermore, the citation network between publications can be employed to analyze relationships between publications, thereby identifying the most influential articles in the field of interest. Additionally, data regarding the countries in which the publications are concentrated can be obtained. Bibliometric analysis encompasses a vast publication network within the applied database, and thus the collected information can be presented graphically. In this study, VOSviewer was utilized to construct and visualize co-authorship, co-citation, and keyword co-occurrence networks [10].

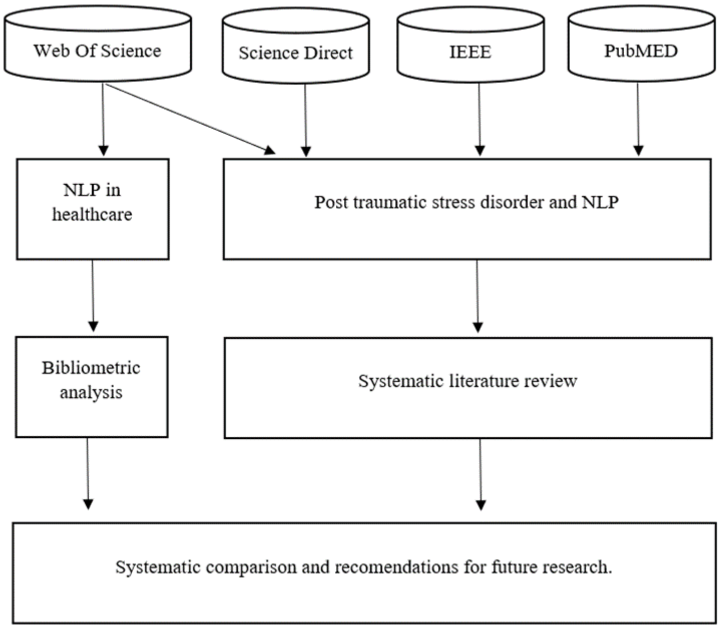

Figure 1. Literature review process

2.2 PRISMA Method

The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines provide an updated framework for reporting systematic reviews and meta-analyses [11]. While earlier systematic reviews in this field commonly followed the PRISMA 2009 guidelines [12], the present study adheres to the updated PRISMA 2020 framework to ensure improved transparency and reporting rigor [11]. The systematic review method is a frequently employed approach for empirically evaluating, developing a critical perspective, and summarizing data collected by researchers. Systematic reviews assist in determining future research priorities. Additionally, systematic reviews facilitate the synthesis of information about a research field and identify potential issues that may emerge in the future by utilizing the information obtained.

Following the PRISMA 2020 guideline, the article selection process was updated to provide detailed counts at each stage. Specifically, during the title and abstract screening, 198 articles were excluded due to irrelevance to PTSD or NLP. At the full-text assessment stage, 35 articles were excluded, primarily because they were reviews, did not apply NLP methods directly, or treated PTSD as a secondary topic.

3 Research Methodology for PTSD

Two research methods were selected for the examination and analysis of studies on the detection of PTSD using NLP methods. These are bibliometric analysis and a systematic literature review method, respectively.

3.1 Bibliometric Analysis and Results

This bibliometric analysis was conducted in accordance with the Guideline for Reporting Bibliometric Reviews of the Biomedical Literature (BIBLIO) [13] which outlines minimum reporting standards for bibliometric reviews in biomedical literature. The guideline was followed to structure the processes of database selection, search strategy development, inclusion/exclusion criteria definition, data extraction, performance analysis, and science mapping. In particular, the Web of Science database was selected, and search strategies were designed based on keywords (Table 1). The analysis covered aspects such as publication trends, country-wise distributions, co-authorship patterns, citation counts, institutional productivity, keyword co-occurrence, and document-based bibliographic coupling.

In the initial phase of this study, the bibliometric analysis method was employed. This approach enabled us to examine the trends in the utilization of NLP in the domain of health. The bibliometric analysis method entails the examination of publications sourced from pertinent databases on research within a particular field. It entails the analysis, mapping and visualization of these publications under specific parameters, thereby providing guidance for researchers regarding the field under investigation [11]. In the course of this research on the use of NLP in the detection of PTSD, the requisite information for the research query was gathered and the requisite analysis was conducted. To this end, the Web of Science database was used, which is one of the leading databases of abstracts and citations of major peer-reviewed publications, including journal articles, conference proceedings, books, and

Table 1. Search criteria for NLP in computer science

Database	Research Scope	Keywords	Language
Web of Science	Articles, papers, book chapters, inprogress documents, review articles, and early access items published between 2012 and May 2024	(“natural language processing” OR “Transformers” OR “language models” OR “text analysis”) AND (“healthcare” OR “electronic health records” OR “unstructured data”)	English

book chapters. A search was conducted in May 2024 utilizing the keywords outlined in Table 1. A total of 4,138 results were obtained from the research conducted by selecting all fields, including the title, author, and publication title. The corpus comprises 2,873 articles, 498 ongoing documents, 446 review articles, 143 early access items, 81 meeting abstracts, 72 editorial materials and 25 book chapters published between May 2012 and 2024. The data obtained from the study was subjected to evaluation under a number of different headings, including those relating to the increase in published material, the number of publications produced by individual countries, an analysis of the number of authors contributing to each publication, an analysis of the number of times each author has been cited, an analysis of the most productive organizations, an investigation into the co-occurrence of keywords, and a document-based bibliographic matching exercise.

For the bibliometric analysis, the Web of Science database was chosen over Scopus. This decision was made because Web of Science offers more standardized citation metadata, higher consistency in author and institutional affiliations, and strong compatibility with visualization tools such as VOSviewer. Although Scopus has broader coverage, it contains more duplicates and heterogeneous document types, which may compromise the precision of bibliometric mapping. Moreover, previous bibliometric guidelines (e.g., BIBLIO) frequently recommend Web of Science as the primary source for systematic bibliometric reviews in biomedical fields.

In this study, a comprehensive search strategy was developed to identify publications related to the use of NLP in the detection of PTSD. The search was performed in the Web of Science database in May 2024. The following keywords were used: “post-traumatic stress disorder,” “PTSD,” “natural language processing,” and “NLP.” These terms were combined using Boolean operators (AND, OR) to create the final search string. The search was conducted across all fields, including title, author, keywords, and abstract. A total of 4,138 documents were retrieved using this strategy, covering the period from May 2012 to May 2024. The inclusion and exclusion criteria below were applied during the selection of documents. The inclusion criteria were (i) publications written in English; (ii) documents categorized as articles, review papers, conference papers, or book chapters; and (iii) studies addressing the use of NLP in the context of PTSD. The exclusion criteria were (i) publications not written in English; (ii) documents such as editorials, meeting abstracts, letters, or commentaries; and (iii) studies not directly addressing the relationship between PTSD and NLP.

3.1.1 Publication increase

Based on the documents obtained from the Web of Science database, 4,138 documents published in the field of NLP and health since 2012 were classified according to years. As seen in Figure 2, 212 documents were accessed in 2012 and more than 700 documents were published by May 2024. It is seen that the interest of researchers working in this field increases with each passing year. Rapid developments in AI have increased the possibilities of using this technology in the field of healthcare.

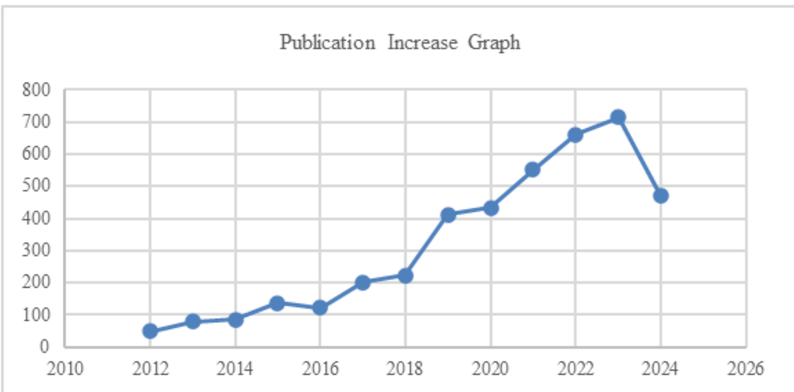

Figure 2. Annual publication numbers

3.1.2 Publication numbers by country

The distribution of more than 4,000 documents taken from the Web of Science database by country was examined. The most common country in these distributions is the United States. Figure 3 shows the distribution of the ten countries that published the most documents. Based on these publication numbers, it is seen that the United States is focusing on AI studies in the field of health. Pioneering studies have been carried out by the United States on the use of NLP in the healthcare sector.

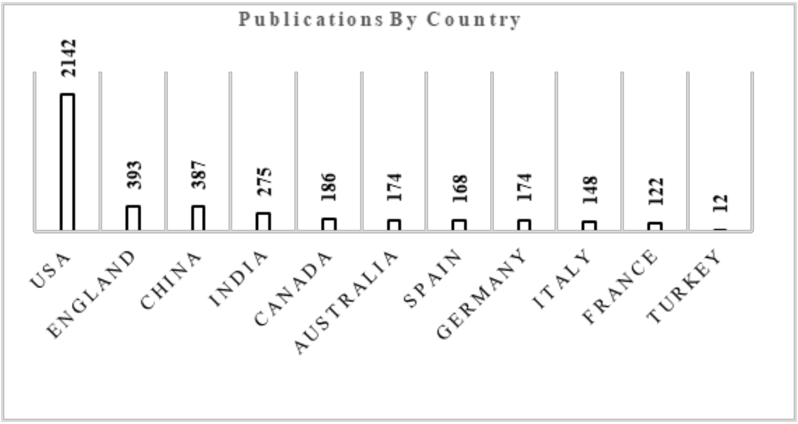

Figure 3. Number of publications by country

In addition, it is seen that the United Kingdom and China have contributed to the developments in this field with over 300 publications. The United States, on the other hand, published seven times more documents. A cross-validation between the bibliometric and systematic review findings reveals a strong degree of consistency. Countries that ranked highest in publication output in the bibliometric analysis, such as the United States and the United Kingdom, also contributed the majority of high-quality studies included in the systematic review, often relying on Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) criteria and advanced NLP methodologies. In contrast, some countries such as China, while highly productive in bibliometric counts, were less represented in the systematic review due to stricter methodological inclusion criteria. This comparison highlights both convergence and divergence across the two analyses, thereby strengthening the reliability of the conclusions of this study.

3.1.3 Co-authorship analysis

The act of co-authorship represents a tangible expression of scientific collaboration within a particular field of study. In the context of academic publishing, co-authorship refers to a publication where two or more individuals contribute to the work together [14]. It encourages academic creativity and improves the quality of scholarly work by providing a forum for the exchange of ideas and knowledge among authors. Using VOSviewer, a joint author analysis was conducted based on the criteria shown in Figure 4. Hongfang Liu is identified as one of the most prolific collaborative authors in this field, with 69 documents. The top 10 authors in the field are presented in Table 2.

3.1.4 Author citation analysis

An author citation analysis enables the relationship between studies in the relevant field to be identified. Researchers employ citations to establish a connection between their own studies and those in the existing literature, thereby facilitating their own contributions to the field. This approach allows the connections between studies on a given subject and between authors to be revealed [15]. A citation analysis was conducted on 4,138 documents retrieved from the Web of Science database. The most frequently cited author is Hongfang Liu, with 1,521 citations. The list of the most cited authors is presented in Table 3. Despite having only two publications, authors such as Katherina Chou and Greg Corrado have a high citation count of 1,291. Using VOSviewer, a joint author citation analysis was conducted based on the criteria shown in Figure 5.

3.1.5 Most productive organizations

Vosviewer was used to examine 4,138 documents retrieved using the relevant keywords. The organizations that support researchers in this field were listed. Table 4 shows the top ten organizations according to the number of documents. Harvard Med School is the most supportive institution, with 165 documents in this field. It has been observed that Mayo Clinic College of Medicine and Science allocates large amounts of resources to researchers on this subject.

In total, there are 3,266 organizations that support researchers in this field. While some of these organizations have published only one document, some have published 100 or more documents. Figure 6 also shows the map of all organizations.

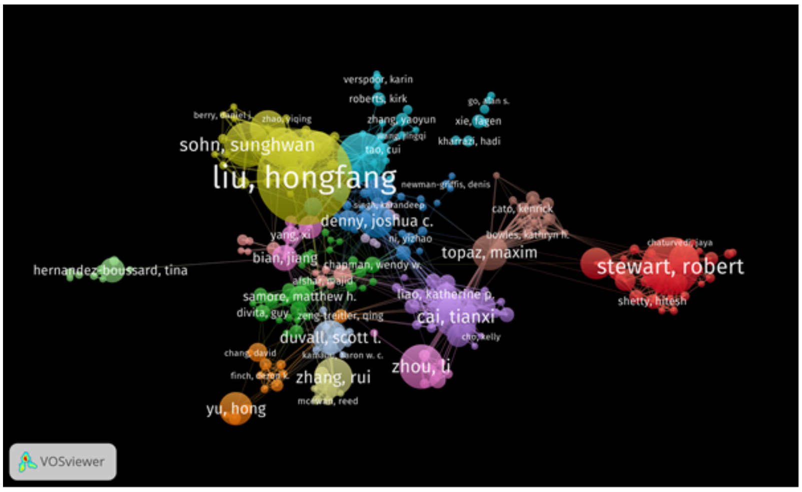

Figure 4. Graphical representation of the co-authorship analysis

Table 2. Search criteria for NLP in computer science

Author Name	Documents
Liu, Hongfang	69
Stewart, Robert	42
Xu, Hua	36
Sohn, Sunghwan	33
Zhou, Li	32
Cai, Tianxi	31
Zhang, Rui	28
Wang, Yanshan	27
Fu, Sunyang	26
Topaz, Maxim	26

Table 3. Search criteria for NLP in computer science

Authors	Citation Counts
Liu, Hongfang	1,521
Chou, Katherine	1,291
Corrado, Greg	1,291
Cui, Claire	1,291
Dean, Jeff	1,291
Depristo, Mark	1,276
Esteva, Andre	1,276
Kuleshov, Volodymyr	1,276
Ramsundar, Bharath	1,276
Robicquet, Alexandre	1,276
Thrun, Sebastian	1,276

3.1.6 Co-occurrence of keywords

Keywords are employed by researchers to provide a concise summary of their studies. The combination of keywords serves to guide researchers in their exploration of a given research field. Furthermore, the use of keywords has been shown to enhance the likelihood of an author being recognized within their respective field of expertise. In this study, the term “natural language processing” was identified as the most frequently occurring keyword, with a total of 1,655 repetitions. Table 5 shows the list of the most used keywords. Figure 7 presents a map illustrating the frequency of occurrence of keywords across 3,667 documents.

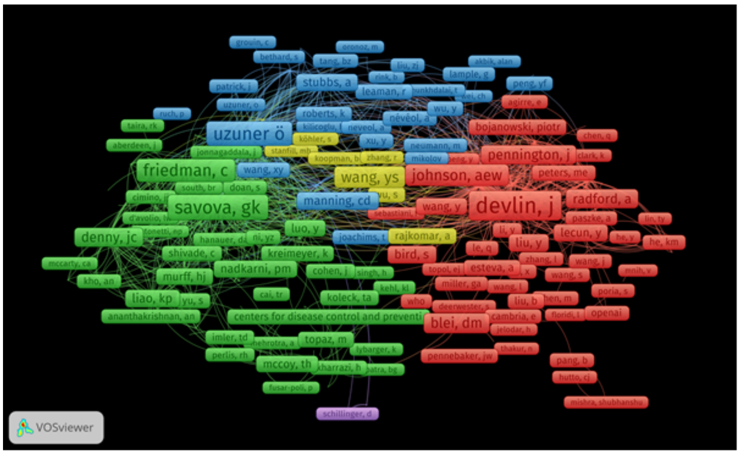

Figure 5. Map of the author citation analysis

Table 4. Search criteria for NLP in computer science

Establishment	Document	Citation Number
Harvard Med. Sch.	165	2,380
Mayo Clin.	141	2,837
Brigham & Women's Hosp	112	2,641
Univ. Utah	111	1,797
Massachusetts Gen. Hosp.	102	2,869
Vanderbilt Univ.	102	2,137
Stanford Univ.	96	2,894
Kings Coll London	91	1,648
Columbia Univ.	78	1,899
Univ. Penn	72	1,490

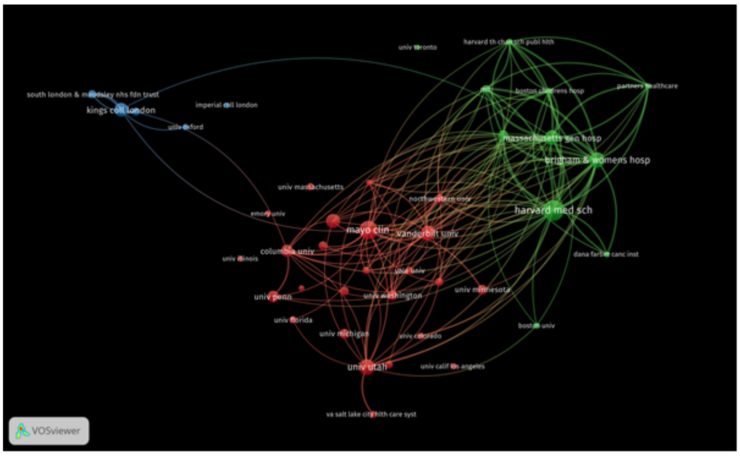

Figure 6. Graph of productive organizations

3.1.7 Document-based bibliographic matching

The analysis is conducted by examining the intertextual relationships between the documents. The referencing technique employed in a given document can be classified into two main categories: bibliographic coupling and co-citation. Bibliographic matching refers to the practice of citing the same publication in two different sources, whereas co-citation denotes the referencing of two distinct publications in a single source. The list of documents that received the highest number of bibliographic matching citations among the 3,667 documents is presented in Table 6. The document by Esteva et al. [16] has been cited by 1,276 different documents. In addition, the document by Jiang et al. [17] is expected to guide researchers working in this field by citing 1,150 different documents. A map of all documents is given in Figure 8.

Table 5. Search criteria for NLP in computer science

Keywords	Number of Co-occurrences
natural language processing	1,655
machine learning	545
electronic health records	509
deep learning	336
artificial intelligence	264
text mining	136
NLP	104
sentiment analysis	103
COVID-19	90

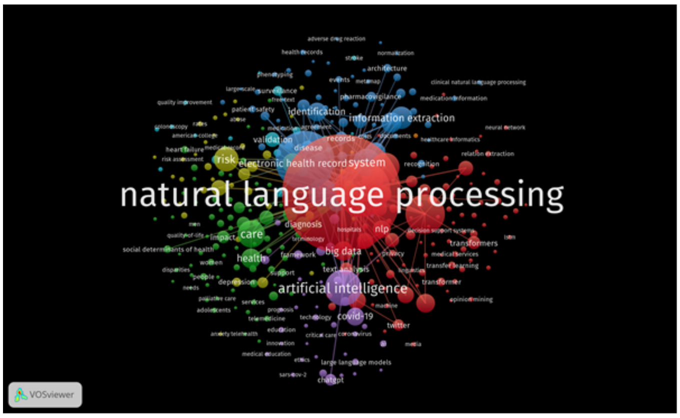

Figure 7. Co-occurrence of keywords in PTSD-related NLP studies

Table 6. Search criteria for NLP in computer science

Reference	Number of Citations
Estexa et al. [16]	1,276
Jiang et al. [17]	1,150
Lundervold and Lundervold [18]	869
Ting et al. [19]	525
Dharmage et al. [20]	492
Zhang et al. [21]	433
Lee and Yoon [22]	380
Price et al. [23]	364
Wang et al. [24]	334
Perera et al. [25]	310

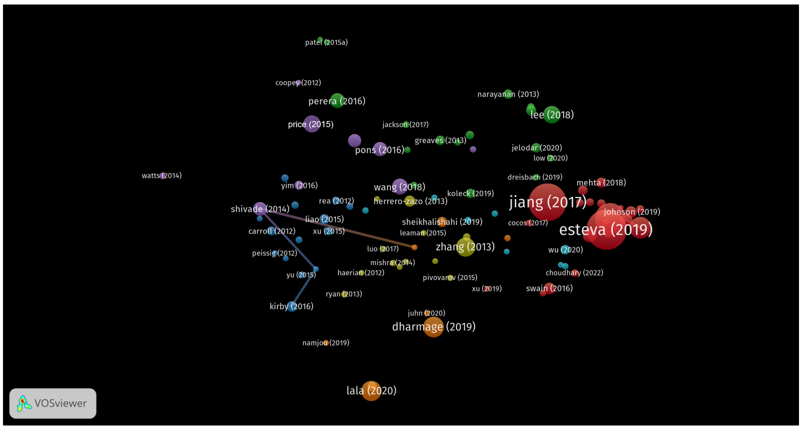

Figure 8. Bibliographic matching map of top-cited PTSD NLP studies

3.1.8 Limitations

This study has certain limitations. Firstly, the analysis was limited to publications indexed in the Web of Science database, potentially excluding relevant studies listed in other academic databases such as Scopus, PubMed, or IEEE Xplore. Secondly, only publications written in English were included, which may have led to the omission of significant research published in other languages. Finally, while bibliometric indicators such as citation counts and co-authorship analysis provide valuable quantitative insights, they may not fully capture the qualitative impact or methodological quality of the included publications.

3.2 Comprehensive Review of Advanced NLP Techniques

This section outlines key NLP techniques used in computer science. To classify textual data, words must first be converted into numerical vectors. The second step involves interpreting these vectors, and the final step is the assessment of model performance using classification algorithms. This section is divided into three parts: text representation, neural network architectures, and pre-trained language models. Text representation refers to the process of making natural language texts processable by computers [26], which can only handle numerical vectors. Thus, text must be transformed into mathematically expressible forms. These methods fall into two main categories: discrete and continuous representations, mainly differing in how they preserve the size and structure of data. These vector formats allow for mathematical operations and similarity calculations between texts.

3.2.1 Discrete text representation

Digitizing text for computer processing is a key concern in digital humanities. Researchers have proposed various methods to address this. Discrete text representation converts documents of different lengths into fixed-length vectors [27], typically by indexing unique words as a dictionary. Each vector has the same length as the dictionary, marking present words with "1" and others with "0" Common methods include the bag-of-words model, Term Frequency-Inverse Document Frequency (TF-IDF), and n-gram. The sparsity of one-hot vectors led to the Bag of Words (BoW) model, where texts are represented by a matrix with rows for sentences and columns for unique words with dimensions depending on sentence count and vocabulary size [28].

TF-IDF is a statistical method used to assess word significance in a document [26]. Term Frequency (TF) reflects how often a word appears, while Inverse Document Frequency (IDF) shows its rarity across the corpus. Words with high TF and IDF scores are considered important. TF-IDF is widely applied in information retrieval, text classification, recommendation systems, and text mining. Another key method is n-gram representation, which splits text into sequences of n words (e.g., bigrams for two words and trigrams for three) [27]. This approach retains word order and context, making it valuable in language modeling, sentiment analysis, automatic classification, and recommendation systems.

3.2.2 Continuous text representation

Continuous text representation mathematically maps words into a high-dimensional vector space to preserve semantic similarity [29, 30]. Models like Word to Vector (Word2Vec), Global Vectors for Word Representation (GloVe), and FastText learn these vectors from large text corpora, capturing meaning and context. This method is widely used in NLP tasks such as sentiment analysis, classification, and information retrieval [31, 32]. Latent Semantic Analysis (LSA) identifies semantic relationships in texts by transforming a term-frequency matrix into a lower-dimensional space using Singular Value Decomposition (SVD) [33]. It enables grouping of semantically similar words and documents, but its limitations lead to the development of advanced methods like Word2Vec. LSA reduces a term-frequency matrix to a lower-dimensional space using SVD, grouping semantically related words and documents, though it has limitations that lead to more advanced models like Word2Vec [33]. Latent Dirichlet Allocation (LDA) is a probabilistic model that identifies hidden topics in document collections [34]. It assumes that documents are mixtures of topics, and words are linked to topics probabilistically, often using algorithms like Gibbs sampling to uncover thematic structures.

Word2Vec generates word embeddings that capture semantic relationships by training on large corpora [31]. It includes Continuous Bag of Words (CBOW), which predicts a word from its context, and Skip-gram, which predicts context from a word. These vectors place semantically similar words closer in space and are widely used in NLP tasks like classification and sentiment analysis. Dynamic word representation allows word embeddings to adapt over time and across contexts, addressing the limitations of static vectors [35]. This is crucial for analyzing rapidly evolving language use, such as on social media, where words like ”cool” may shift in meaning depending on context. These models enhance language understanding by capturing temporal and contextual semantic shifts. Contextualized word embeddings, such as Context Vectors (CoVe), Embeddings from Language Models (ELMo), Generative Pre-trained Transformer (GPT), and Bidirectional Encoder Representations from Transformers (BERT), generate vectors that vary with context, reflecting a word's meaning dynamically. These are typically trained on large-scale language modelling tasks to create context-sensitive representations [36].

Context2Vec builds on this by learning word or phrase representations based on their surrounding context, offering deeper insight than traditional fixed embeddings [37]. Context2Vec uses neural networks, often Recurrent Neural Network (RNN) or Transformer-based, to generate word embeddings that reflect immediate context [38]. For example, in "The bank is by the river", it distinguishes between meanings of "bank" by analyzing surrounding words. This enhances performance in tasks like language modelling, sentiment analysis, and machine translation by capturing nuanced semantic relationships. CoVe, on the other hand, generates contextual embeddings using Bidirectional Long Short-Term Memory networks (BiLSTM) that process text in both directions [36]. This allows the model to consider both preceding and following words, producing richer word representations. CoVe has shown strong results in tasks like question answering, sentiment analysis, and named entity recognition by better capturing contextual meaning.

3.3 Neural Network Architectures

In recent years, there has been a notable increase in the utilization of artificial neural network models by researchers in the field of health-related text analysis. Amongst the most prevalent are Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), RNN and Transformer-based architectures.

3.3.1 CNN

Initially successful in image processing, CNNs have since proven highly effective in NLP. Designed to capture local patterns and dependencies in word sequences, CNNs apply convolution operations to extract key features from text. These operations help identify important structures, such as nearby word relationships. CNNs are widely used in NLP tasks like sentiment analysis, text classification, summarization, and language modeling, significantly enhancing structural understanding of textual data.

3.3.2 LSTM

LSTM networks, a type of RNN, were developed to address the vanishing gradient problem and capture long-term dependencies in sequences [39]. They feature internal memory cells and three gates of forget, input, and output that regulate information flow. These gates allow LSTM to retain or discard information as needed, enabling learning across longer contexts. LSTMs are widely used in NLP tasks such as machine translation, sentiment analysis, speech recognition, and text generation, where understanding sequence and context is critical. Their ability to model syntax, semantics, and discourse has significantly improved language modelling.

3.3.3 RNN

RNNs are designed to model sequential data by retaining an internal state and using feedback loops to capture temporal dependencies [40]. Unlike feedforward networks, RNNs process sequences recursively, making them suitable for tasks involving variable-length inputs. However, they face the vanishing gradient problem, which limits their ability to learn long-term dependencies. To address this, enhanced versions like LSTM and Gated Recurrent Unit (GRU) include gating mechanisms that improve information retention over longer sequences. RNNs are widely applied in NLP, time series analysis, and speech recognition, particularly in language modeling, translation, sentiment analysis, and entity recognition.

3.3.4 Transformer-based language models

Transformer-based models, introduced by Vaswani et al. [41], revolutionized NLP by replacing recurrent and convolutional layers with a self-attention mechanism. This allows them to capture long-range dependencies and process sequences in parallel, boosting efficiency. Models like BERT and GPT have achieved state-of-the-art results in tasks such as classification, question answering, and machine translation [42]. Pre-trained on large corpora, they offer strong linguistic representations adaptable to various applications. Their success has led to widespread use in academia and industry, enhancing NLP capabilities across numerous domains.

3.4 Systematic Literature Review and Results

This systematic review was conducted in accordance with the updated PRISMA 2020 guidelines. The study aimed to synthesize the current state of research on the use of AI and NLP in the detection of PTSD. A PRISMA 2020 flow diagram was used to document the screening process, and a completed PRISMA 2020 checklist was provided as supplementary material. This review did not include a quantitative meta-analysis. Therefore, no formal data extraction of statistical outcomes (e.g., sample sizes, confidence intervals, and effect sizes) was performed. Instead, a thematic and bibliometric synthesis approach was applied. Each eligible study was manually reviewed and coded using a structured extraction framework. The following key data items were recorded for each study: title, first author, publication year, database source, dataset type, target task (e.g., classification or detection of PTSD), applied NLP or AI techniques (e.g., RNN, LSTM, and BERT), language of the dataset, and study context (e.g., clinical or social media). This information was manually organized into an Excel spreadsheet and served as the basis for qualitative comparison and thematic categorization. As no meta-analysis was conducted, formal statistical tests for reporting bias (e.g., funnel plot analysis) and formal risk of bias tools such as the Revised Cochrane Risk of Bias Tool for Randomized Trials (RoB 2) or Risk Of Bias In Non-randomized Studies—of Interventions (ROBINS-I) were not applied. However, potential sources of methodological limitations such as small sample sizes, dataset imbalance, insufficient reporting, and domain specificity were noted qualitatively during the screening process. These factors were taken into account when evaluating the robustness of each study's findings.

Similarly, a formal certainty of evidence assessment tool such as Grading of Recommendations Assessment, Development and Evaluation (GRADE) was not utilized due to the absence of aggregate quantitative outcomes. Nonetheless, the overall reliability of included studies was considered based on the consistency of reported results, methodological transparency, and dataset characteristics (e.g., annotation quality, linguistic variability, and population representativeness). A systematic literature review is a type of review article that aims to identify evidence-based clinical effectiveness on a given problem in order to provide recommendations for the resolution of specific research topics. The objective of a systematic literature review is to synthesize existing research, identify gaps in the existing research landscape, produce a coherent report, and create a research framework. In this study, databases such as ScienceDirect, IEEE, Web of Science and PubMed were scanned. A systematic literature review was conducted in accordance with the PRISMA method and meta-analysis. The following steps were followed for the systematic literature review in the study:

Step 1: Identifying a research question.
Step 2: Determining the database to be researched.
Step 3: Identifying keywords.
Step 4: Review of title, abstract and keywords.
Step 5: Literature eligibility criteria (inclusion and exclusion criteria).

3.4.1 Keywords selection

A systematic literature review was conducted to determine research topics. These were evaluated in relation to PTSD and AI, and search terms were identified. The research terms are provided in the table. The research was conducted using these keywords in relevant databases. The selected keywords are given in Table 7.

The research was conducted using databases that had been previously identified through the application of a keyword selection process. The relevant keywords were determined to be "post-traumatic stress disorder" "PTSD" "natural language processing" "recurrent neural network" "RNN" "NLP" "language models" and "Transformers". The research text was created by combining the keywords into two groups. The resulting search string was as follows: ("post-traumatic stress disorder" OR "PTSD") AND ("recurrent neural network" OR "RNN" OR "natural language processing" OR "NLP" OR "language models" OR "Transformers").

Table 7. Search criteria for NLP in computer science

Keywords
post-traumatic stress disorder
PTSD
recurrent neural network
RNN
natural language processing
language model
Transformers

3.4.2 Research questions

A systematic literature review is a process that begins with the formulation of research questions. These questions serve to guide the investigation and facilitate the analysis of the subject matter. Table 8 illustrates the research questions and objectives associated with the detection of PTSD using NLP methods.

3.4.3 Exclusion criteria and database results

a. Articles published before 2012 are not selected.
b. Articles that use PTSD disorder as a subheading of another disease are not selected.
c. Articles written in a language other than English are not selected.

Studies written in English were searched between January 2012 and May 2024 through ScienceDirect, IEEE, PubMed and Web of Science. Following the PRISMA guideline, 206 studies were identified from ScienceDirect, 12 from the IEEE database, 29 from PubMed, and 37 from Web of Science. A total of 328 documents were examined. The results obtained are shown in Table 9. The article selection process is shown in Figure 9. Although 29 records were initially retrieved from PubMed, only two articles met the inclusion criteria. Most PubMed results were clinical or biological studies not directly related to NLP, or general psychiatric research where PTSD was treated as a secondary aspect. In some cases, NLP was mentioned in the abstract, but a full-text review revealed insufficient application of NLP techniques for PTSD detection. Therefore, the low number of included articles is due to the characteristics of PubMed-indexed literature rather than overly strict selection, minimizing the risk of selection bias.

Table 8. Search criteria for NLP in computer science

Number	Question	Purpose
1	What are the challenges associated with the utilization of NLP for the classification of PTSD?	Identifying the challenges experienced and determining solutions
2	Can AI be used for the detection of PTSD in the healthcare system?	Determining its impacts on the healthcare system
3	What are the datasets used in PTSD detection? What are the challenges of these datasets?	Evaluating the methods used based on specific criteria
4	What are the AI techniques used in solving PTSD? What are the opportunities and challenges of these techniques in mental health disorder detection using NLP?	Identifying the challenges and opportunities in mental health disorder detection using NLP techniques

Table 9. Search criteria for NLP in computer science

Database Name	Number of Articles Found	Number of Articles Included	Number of Excluded Articles
ScienceDirect	240	27	213
IEEE	15	8	7
Web of Science	44	19	25
PubMed	29	2	27

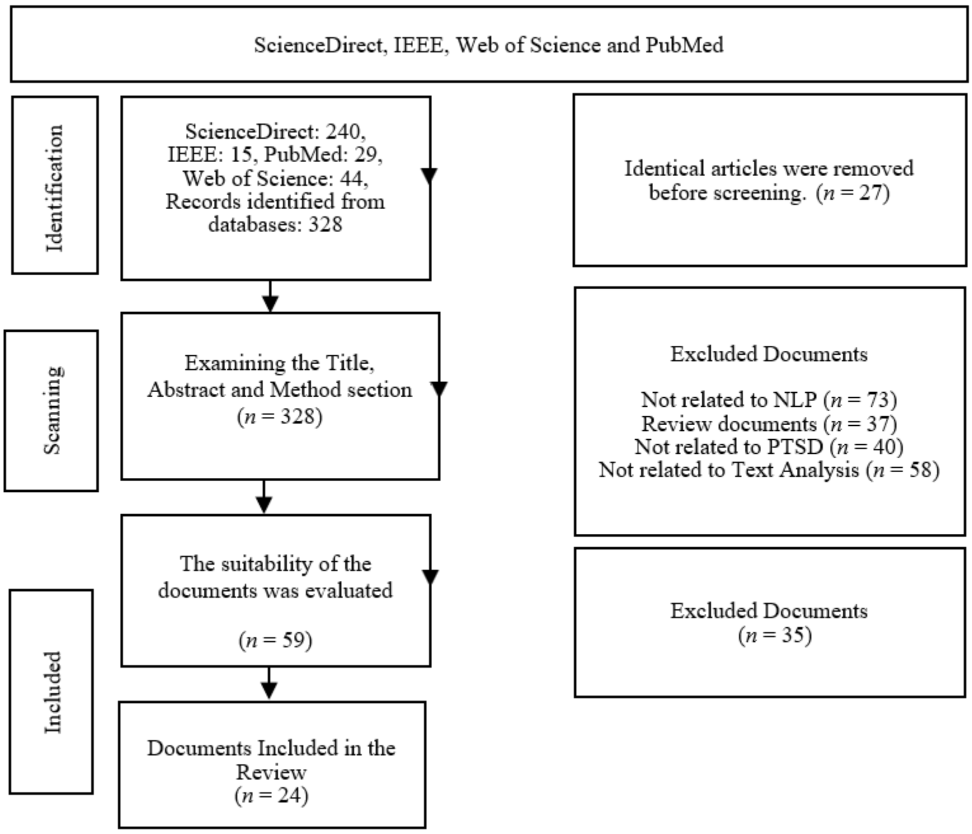

Figure 9. Article selection process

4 NLP Methods Used in PTSD

A categorization and examination of NLP methods used in the detection of PTSD was conducted. Furthermore, an investigation was undertaken into the evolution of NLP methods between 2012 and 2024, with a particular focus on their utilization in the early detection of the disease.

4.1 Classification of NLP Techniques Used in PTSD

This analysis examines approaches used in research on PTSD by comparing NLP methods and techniques used in various studies. The results obtained are shown in Table 10.

Table 10. NLP techniques

Reference	NLP Techniques	Strengths	Weaknesses	Comparative performance / applicability
[43]	Text mining, Support Vector Machine (SVM), Logistic Regression (LR), text cleaning, tokenization, lemmatization, and stop word removal	Effective at identifying PTSD features from narratives	Limited generalizability due to small sample size (159 narratives)	SVM and LR achieve moderate accuracy (∼ 70–75%) in small narrative datasets; good for pilot studies but not scalable
[44]	SVM, Naive Bayes (NB), text cleaning, tokenization, entity recognition, and sentiment analysis	Useful for quality assessment of therapy notes and automated classification	May not capture nuanced meanings in therapy notes	SVM generally outperforms NB in precision; suitable for structured text (therapy notes) but weaker in nuanced semantics
[45]	NLP, entity recognition, standardization, NB, and Decision Tree (DT)	Good for high-level conceptualization of medical notes	Less effective for detailed text analysis and requires substantial manual labeling	NB and DT are simple, interpretable methods; effective for structured Electronic Health Records (EHR) data but weaker on unstructured text
[46]	NLP keyword extraction. Algorithmic matching. Validation with EMR codes.	Comprehensive analysis of psychotherapy usage	Limited by the specificity of the text data used (e.g., psychotherapy notes)	RF shows better generalization than LR on psychotherapy notes; LR remains robust with smaller, structured data
[47]	SVM, NB, text cleaning, speech recognition, language modeling, and data coding	Effective for realtime distress detection in crisis situations	Speech recognition accuracy may impact the results	SVM performs better than NB (∼75% F1) when ASR accuracy is high; applicable in real-time crisis systems
[48]	LR, NB, text cleaning, data standardization, and missing data analysis	Effective for assessing psychotherapy implementation in a clinical setting	May not generalize to non-Veterans Affairs (VA) settings or other types of psychotherapy	NB is faster for categorical data; LR is more robust with continuous data; both are limited to VA datasets

Reference	NLP Techniques	Strengths	Weaknesses	Comparative performance / applicability
[49]	NN, SVM, text cleaning, data normalization, and statistical methods	Innovative use of cognitive computing methods for differential diaznosis	Complex methodology, potentially high computational requirements	NN > 80% F1 on clinical records; better than SVM in large datasets, but requires strong hardware
[50]	SVM, LR, text cleaning, tokenization, lemmatization, and stop word removal	Effective for automated PTSD screening from self-narratives	Limited by data diversity and sample size	SVM slightly outperforms LR in short self-narratives; accuracy ∼70–75%, limited generalizability
[51]	Heuristic algorithms, DT, text cleaning, template extraction, and entity recognition	Useful for novel template identification from medical records	Pilot study, limited sample size (5,000 records)	Effective for exploratory tasks; low scalability and generalization
[52]	Deep Neural Network (DNN), CNN, text cleaning, NLP, and data normalization	Advanced neural network techniques, good for complex pattern recognition	Requires extensive computational resources and data normalization	CNN/DNN achieve >80–85% accuracy in psychiatric notes; best for large text corpora and require GPUs
[53]	LR, NB, text cleaning, NLP, and data standardization	Large-scale analysis, applicable to a national healthcare system	Data quality and consistency issues across a large system	Both scale well; LR offers stability, NB is faster but less robust with noisy data
[54]	LR, NB, text cleaning, NLP, and data standardization	Comprehensive assessment of progress and outcomes in veterans	May not be applicable to nonveteran populations	Similar to [42]; applicability restricted to VA data
[55]	LR, SVM, text cleaning, sentiment analysis, and topic modeling	Focus on explainability and trustworthiness in Twitter analysis	Limited to Twitter data, which may not capture all aspects of PTSD	SVM handles sparse Twitter features better; LR is easier to interpret but less accurate
[56]	LR, NB, text cleaning, NLP, and data standardization	Longitudinal analysis over a significant time period	May be affected by changes in data collection methods over time	Stable over time but sensitive to changes in data capture; NB is simpler, LR is more robust
[57]	Transformerbased models, multi-aspect transfer learning, text cleaning, language modeling, multilingual embeddings	Effective for detecting mental disorders in lowresource languages; strong performance on social media data	Limited generalizability to clinical datasets; performance depends heavily on pre-trained transformer quality	Transformer models outperform traditional ML baselines; strong semantic sensitivity but computationally heavy
[58]	NN, LR, text cleaning, NLP, and sentiment analysis	Advanced AI methods for predicting longterm outcomes, effective sentiment analysis	Requires substantial computational resources and highquality data	NN achieves >80% Fl in longitudinal datasets; LR is less accurate but efficient on smaller sets

Reference	NLP Techniques	Strengths	Weaknesses	Comparative performance / applicability
[59]	Neural networks, language modeling, text cleaning, sentiment analysis	Effective in predicting longterm trauma-related outcomes; sentiment features significantly improve prediction	Requires sufficiently large training datasets; interpretability of deep models is limited	Neural models outperform simpler baselines in most scenarios; LR is not reported in the original study
[60]	Social media analytics, text analysis, and emoji analysis	Interdisciplinary approach combining linguistics and cyber psychology, bilingual analysis (English and Arabic)	-	Strong in exploratory social/cyber psychology; weak comparability with clinical data
[61]	Bollo. TF-IDF, logistic regression (LR), random forest (RF), Naive Bayes (NB), multilayer perceptron (MLP) neural network	Combines structured EMR data with NLP features; interpretable ML models work well on sparse EHR text	Performance limited by EMR sparsity and missing data; MLP requires more computation	RF and MLP perform better than LR/NB in EHR classification; Boll + TF-IDF remain interpretable and stable
[62]	LSTM, RF, survey data cleaning, missing data imputation, text normalization, and data labeling	Focus on a specific and important issue, effective for handling sequential data	Limited to a specific population and data type	LSTM is >80% accuracy on sequence data; RF is stronger with tabular/structured inputs
[63]	SVM, NB, text cleaning, NLP, entity recognition, and text normalization	Large-scale data from national healthcare systems, effective for largescale analysis	Data quality and consistency issues across a large system	SVM is generally more accurate; NB is faster but less robust with heterogeneous data
[64]	MentalBERT and processing of EMR data	High precision and recall in capturing PTSD from EMR data, combining structured and unstructured data.	Model performance may vary depending on data quality, requiring extensive preprocessing of EMR data.	BERT-based models have >85% Fl, state-of-the-art; strong but compute-intensive
[65]	ML modeling analysis and data mining	Personalized approach, focus on postpartum PTSD, presented at many conferences	Limited to women's health context, specific to birth experiences	Promising in postpartum data; not easily generalizable outside women's health
[66]	LDASSes (dynamic topic modelling)	Identification of time-sensitive suicide risk variables improved over the REACHVET model	Requires large data sets, may not generalize well to smaller cohorts	Works best in longitudinal VA datasets; not reliable for small cohorts

He et al. [43] proposed the use of TF-IDF and n-gram models, in conjunction with verbal feature analysis and text mining techniques, to effectively detect the symptoms of PTSD. While this approach is effective in detecting PTSD symptoms based on self-narratives, it is constrained to self-reported narratives alone. The MedCat framework, developed by Fodeh et al. [45], employed an ontology-based approach to conceptualize medical notes, thereby facilitating a high level of conceptualization and the interpretation of medical terms. However, there are some limitations to the applicability of this framework. Shiner et al. [44] examined the use of evidence-based psychotherapy through analysis of electronic medical records and treatment notes, utilizing word frequency analysis and text classification techniques. While this method provides information on evidence-based practices, it remains dependent on accurate documentation. Shiner et al. [44] automatically classified psychotherapy note texts using SVM and LR techniques with the aid of algorithms in ML. However, with this technique offering an automatic, objective quality assessment method, accuracy in text classification, including comprehensiveness, should be guaranteed. Howard et al. [49] did research on cognitive computational methods for various diagnoses of PTSD, such as semantic analysis and pattern recognition techniques. While these methods offer more accurate diagnoses, they are not without difficulties in terms of complexity and applicability. Watts et al. [48] examined the implementation of evidence-based psychotherapies in VA specialty clinics, analyzing both electronic medical records and clinical notes. Although this approach does give an indication that evidence-based practices are effective, there is a need to ensure that such practices are sustainable and documented.

In the work of Pacula et al. [47], audio signal processing and text mining techniques were employed to automatically detect the signs of psychological distress in crisis line conversations. While this method detects symptoms with both speed and accuracy, crisis line conversations are inherently complex, posing a great challenge. Tran and Kavuluru [52] used a DNN for the prediction of mental states from psychiatric notes and retrieved high accuracies with the use of both RNN and CNN models. However, such models require training on large amounts of data. Another novel approach was the heuristic algorithms to identify new patterns from VA text data obtained from electronic medical records; however, the algorithm's applicability and generalizability are big challenges [51]. In a nutshell, all these papers make a great contribution to the detection and analysis of PTSD, as well as other mental disorders, by applying different NLP techniques. Yet, each method has its strengths and weaknesses depending on the specific context in which it would be put to application, together with the dataset. Therefore, their combination and continuous improvement will go hand in glove with more comprehensive and accurate results.

Several of the reviewed studies in Section 4.1 incorporated internationally recognized standardized diagnostic tools for assessing PTSD. For instance, Shiner et al. [59] utilized both the PTSD Checklist for DSM-IV (PCL-IV) and the PCL-5 (aligned with DSM-5 criteria); Sawalha et al. [63] applied the PCL-C and discussed its correlation with Clinician-Administered PTSD Scale for DSM-5 (CAPS-5), the clinical gold standard; Bartal et al. [65] employed the PCL-5 for postnatal PTSD symptom evaluation. In addition, Zafari et al. [61] classified PTSD cases using the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) code 309.81, and Howard et al. [49] integrated structured clinical interviews such as CAPS, Structured Clinical Interview for DSM Disorders (SCID), the Diagnostic Interview Schedule, Fourth Edition (DIS-IV), and the Composite International Diagnostic Interview (CIDI). Although not all articles referenced ICD-11 directly, the continued use of ICD-9/10 across large health systems provides clinically valid diagnostic frameworks. Therefore, a significant portion of the analyzed literature demonstrates alignment with established diagnostic standards such as DSM and ICD, adequately supporting PTSD classification within this review context.

The included studies varied in their definition of PTSD symptoms. While the majority relied on DSM-based criteria (primarily DSM-5 in recent studies, with a smaller number using DSM-IV), several studies referred to ICD-10/11 classifications, and others employed validated symptom scales such as PCL or CAPS [58]. Although these approaches are broadly aligned, they are not identical, and this heterogeneity may introduce bias in the comparison of results. To mitigate this, this study explicitly reports the diagnostic frameworks used and emphasizes that future research should adopt standardized definitions, preferably DSM-5 or ICD-11, to ensure comparability across studies.

4.2 Datasets

A variety of datasets have been employed in research on NLP methods used to assess and treat PTSD. An in-depth examination of these datasets is essential to ascertain the efficacy of the methods, their potential applications, and the challenges encountered. The section below presents a summary of the key points regarding the datasets utilized in the articles referenced in Table 11.

Table 11. Datasets

Reference	Datasets Used	Data Type	Dataset Size	Dataset Accessibility
[43]	Verbal features in selfnarratives	Text	159 narratives	Closed
[44]	Automated classification of psychotherapy note text	Text	4310 notes	Closed
[45]	High-level conceptualization of medical notes	Text	60000+ notes	Closed

Table 11. Datasets

Reference	Datasets Used	Data Type	Dataset Size	Dataset Accessibility
[46]	Measuring use of evidencebased psychotherapy for PTSD	Text	3.191 notes	Closed
[49]	Differential diagnosis of PTSD using cognitive computing methods	Text and clinical records	Several thousand records (exact number not reported)	Closed
[47]	Crisis hotline conversations	Speech	50 hours	Closed
[48]	Evidence-based psychotherapies in VA specialty clinics	Text	1.000+ notes	Closed
[52]	"History of present illness" in psychiatric notes	Text	53000+ notes	Closed
[50]	Patients' self-narratives	Text	200+ narratives	Closed
[51]	VA electronic medical record texts	Text	5.000+ records	Closed
[53]	Evidence-based psychotherapy trends in a national healthcare system	Text	19.000+ records	Closed
[56]	Evidence-based psychotherapy trends for PTSD in a national healthcare system	Text	19.000+ records	Closed
[55]	Twitter analysis model for PTSD assessment	Social media data	57.000+ tweets	Open
[54]	NLP tools for assessing veteran populations	Text	8,183 notes	Closed
[57]	Psychiatric assessments and patient records	Text and clinical records	2.000+ records	Closed
[58]	AI language predictors for twoyear trauma-related outcomes	Text and clinical records	4,000+ records	Closed
[60]	Social media data from online forums and platforms	Social media data	Not reported	Open
[59]	VA clinical notes	Text	Not reported	Closed
[61]	EHR data from the Manitoba Primary Care Research Network (MaPCReN) of primary care clinicians in Manitoba, Canada.	Text	Not reported	closed
[62]	Child abuse mental symptom cases in India	Survey data	1,000+ cases	Closed
[64]	Pan-Canadian EHR data were used.	Short diagnostic text fields and free text encounter notes obtained from some regional networks.	56.795	Processed with anonymization algorithms and accessible by certain researchers
[63]	AVEC-19 semi-structured clinical interview transcripts	Text	275	Closed

Table 11. Datasets

Reference	Datasets Used	Data Type	Dataset Size	Dataset Accessibility
[65]	Used data from 1,127 postpartum women who participated in a research survey conducted during the COVID-19 pandemic	Participants ’ written accounts of their personal birth experiences , PTSD scales	1127	Available by agreement
[66]	Included US VA patients diagnosed with PTSD between 2004 and 2013	Psychother apy notes, free text encounter notes.	568	Is limited

Some studies [43, 44, 48, 50] were conducted with patients' own narratives and clinical notes. These data include patients' PTSD symptoms and the way they express their condition. Such data can be difficult to analyze due to the complexity of language and individual differences. However, NLP and text mining techniques can be used effectively to extract important features from these narratives. Shiner et al. [46] similarly measured the use of evidence-based treatment methods by analyzing clinical notes and psychotherapy texts. Some studies [48] were conducted with medical records and EHR data. These datasets contain large amounts of unstructured data and provide extensive information about patients' medical history. Processing such data may present challenges due to the volume and diversity of data, but NLP techniques can be used to extract meaningful information from the data. Studies were conducted with social media and other open-source data. Twitter data was used in the study by Alam and Kapadia [55]. Social media data consists of information that users voluntarily share, and the data can reflect real-time emotional states. However, social media data has a high potential to be noisy and misleading, so it is important to use reliable and explainable models in analyzing the data. Some studies were conducted with psychotherapy notes and interviews. For example, psychotherapy notes and crisis line conversations were used [44, 47, 48]. Such data are valuable for understanding patients' treatment processes and responses because they involve direct patient-therapist interactions. However, the emotional content and context in the data must be interpreted correctly.

The predominance of closed datasets (e.g., EMR and VA clinical notes) in PTSD research raises challenges for reproducibility and comparability across studies. While such datasets ensure clinical validity, they are often inaccessible to independent researchers due to strict privacy regulations, limiting external validation and generalization of findings. In contrast, open data sources such as social media provide broader accessibility and support replicability, but they introduce challenges related to noise, context validity, and ethical concerns. A balanced approach is required: advanced anonymization techniques, federated learning, and secure data-sharing frameworks can help preserve patient privacy while enabling academic transparency. The establishment of controlled-access repositories with de-identified clinical data may also support reproducibility without compromising confidentiality.

The reliance on closed datasets such as VA clinical records may lead to selection bias, as these sources typically represent specific populations and limit generalization to broader contexts. Conversely, open datasets like Twitter and Reddit provide broader coverage and replicability, but raise ethical concerns (e.g., lack of informed consent) and methodological challenges due to noisy, unstructured, and culturally variable content. Another critical issue is data imbalance, since PTSD-positive cases are often underrepresented compared to negative or control samples. This imbalance can bias models toward majority classes and reduce sensitivity to true PTSD cases. Addressing this challenge requires the use of techniques such as data augmentation, resampling strategies such as Synthetic Minority Over-sampling Technique (SMOTE), cost-sensitive learning, and the integration of synthetic or multimodal data. Such strategies help mitigate bias, improve generalizability, and ensure more ethically robust PTSD detection systems.

4.3 Preprocessing Techniques

The use of NLP techniques and ML algorithms in PTSD research plays an important role in understanding patients' conditions and improving treatment processes. Table 12 shows the datasets, pre-processing techniques, classification algorithms used, and the results obtained.

The preprocessing techniques employed in these studies are essential for transforming raw text data into a format that is conducive to analysis. These techniques enhance the quality of the data, reducing noise and extracting meaningful information, which is essential for NLP and ML models to perform effectively. The selection of preprocessing techniques is contingent upon the characteristics of the data and the analytical objectives. For instance, in the context of clinical and medical studies, entity recognition and standardization are essential for the extraction and classification of medical terms. In the domain of social media analysis, sentiment analysis and topic modelling are crucial for understanding public opinion and detecting mental health conditions. In general, the judicious choice and implementation of preprocessing techniques are pivotal for the success of text analysis and ML projects.

Table 12. Preprocessing techniques

Reference	Used Datasets	Preprocessing Techniques	Classification Algorithms	Impact on PTSD Detection / Sensitivity
[44]	Automated classification of psychotherapy note text	Text cleaning, tokenization, entity recognition, and sentiment analysis	SVM and NB	Entity recognition and sentiment analysis improved affective feature extraction, leading to higher sensitivity in therapy notes.
[43]	Verbal features in self-narratives	Text cleaning, tokenization, lemmatization, and stop word removal	SVM and LR	Lemmatization preserved linguistic meaning; excessive stop-word removal risked losing affective cues (e.g., negations).
[46]	Measuring use of evidence-based psychotherapy for PTSD	Text cleaning, NLP, and data standardization	LR and RF	Standardization improved comparability across records, critical for longitudinal consistency.
[45]	High-level conceptualization of medical notes	Text cleaning, entity recognition, standardization, and text normalization	NB and DT	Entity recognition of traumarelated concepts improved detection; normalization reduced noise.
[48]	Evidence-based psychotherapies in VA specialty clinics	Text cleaning, data standardization, and missing data analysis	LR and NB	Handling missing data ensured robust outcomes; preprocessing minimized false negatives.
[47]	Crisis hotline conversations	Text cleaning, speech recognition, language modeling, and data coding	SVM and NB	ASR quality is critical; errors in speech recognition directly reduced PTSD detection sensitivity.
[49]	Differential diagnosis of PTSD using cognitive computing methods	Text cleaning, data normalization, statistical methods, and data integration	NN and SVM	Normalization and integration improved comparability; NN showed better robustness to preprocessing noise.
[51]	VA electronic medical record texts	Text cleaning, template extraction, entity recognition, and missing data analysis	Heuristic algorithms and DT	Template extraction improved structured feature capture; entity recognition boosted affective feature detection.
[50]	Patients' selfnarratives	Text cleaning, tokenization, lemmatization, and stop word removal	SVM and LR	lemmatization aided consistency, stop-word removal sometimes removed emotional words.
[52]	"History of present illness" in psychiatric notes	Text cleaning, NLP, data coding, and data normalization	DNN and CNN	Normalization essential for DNN/CNN; improved signal-tonoise ratio for affective features.
[53]	Evidence-based psychotherapy trends in a national healthcare system	Text cleaning, NL.P and data standardization	LR and NB	Standardization critical for multi-center data; ensured stable feature extraction.

Table 12. Preprocessing techniques

Reference	Used Datasets	Preprocessing Techniques	Classification Algorithms	Impact on PTSD Detection / Sensitivity
[54]	NLP tools for assessing veteran populations	Text cleaning, NLP, and data standardization	None reported (NLP extraction pipeline, not a classifier)	preprocessing stability improved longitudinal assessments.
[55]	Twitter analysis model for PTSD assessment	Text cleaning, sentiment analysis, and topic modeling	LR and SVM	Sentiment analysis directly enhanced affective signal detection; noise in tweets remained a challenge.
[56]	Evidence-based psychotherapy trends for PTSD in a national healthcare system	Text cleaning, NLP, and data standardization	LR and NB	Standardization key for largescale VA datasets; reduced variability across facilities.
[61]	EHR data from the MaPCRenen	Bollb lemmatization, and TF-IDF	Multi-layered NN, and RF	Lemmatization and TF-IDF improved text vectorization; critical for sparse EHR data.
[59]	VA clinical notes	Text cleaning, NLP, entity recognition, missing data analysis, data standardization, and text normalization	LR and RF	Entity recognition and normalization improved affective feature extraction and reduced false negatives.
[60]	text cleaning, linguistic markers, emoji analysis	None (conceptual paper)	SVM and NB	Highlights feasibility of online PTSD detection; no empirical model tested.
[58]	AI language predictors for twoyear trauma-related outcomes	Text cleaning, NLP, and sentiment analysis	NN and LR	Sentiment analysis crucial for affective feature extraction; improved long-term PTSD prediction.
[57]	Psychiatric assessments and patient records	text cleaning, tokenizer, subwerd. embeddings	transformerbased classifiers (BERT/RoBERTa), multi-aspect transfer	Transfer learning significantly improves performance for lowresource mental health datasets.
[64]	Pan-Canadian EHR data	basic text normalization and cleaning	None (not a classification study)	Preprocessing enables extraction of structured diagnostic patterns; study focuses on EMR completeness, not model performance.
[62]	Child abuse mental symptom cases in India	Survey data cleaning, missing data imputation, text normalization, and data labeling	RF, SVM, LR, KNN, DT	RF performed best among classic ML models after preprocessing.
[63]	AVEC-19 semistructured clinical interview transcripts	text cleaning, sentiment extraction, lexical features	LR, SVM, RF, XGBeeat	sentiment features improved discrimination between PTSD and non-PTSD

Table 12. Preprocessing techniques

Reference	Used Datasets	Preprocessing Techniques	Classification Algorithms	Impact on PTSD Detection / Sensitivity
[65]	They provided brief written accounts of their birth experiences.	all-mpnet-base-v2, Deep Feed Forward Neural Network (DFNN), Sentence-Transformers	DFNN	Transformer embeddings reduced reliance on preprocessing; better affective capture.
[66]	US VA patients diagnosed with PTSD between 2004 and 2013.	Tokenization, removing stop words, removing symbols and punctuation, and converting all words to lowercase	Dynamic topic modelling and LDA	Tokenization essential for topic modeling; stop-word removal sometimes removed negations critical for PTSD features.

5 Results and Discussion

While this review underscores the potential of NLP techniques in identifying PTSD-related symptoms, it is crucial to delineate the distinction between symptom detection and diagnostic assessment. The studies reviewed herein employed NLP primarily for the identification of linguistic markers and symptomatology associated with PTSD. However, direct diagnostic classification using NLP remains a complex and evolving domain, requiring further validation through clinical datasets and standardized diagnostic tools (e.g., PCL-5, CAPS, and ICD-11). Consequently, this review does not claim that NLP techniques can independently diagnose PTSD but rather highlights their supplementary role in symptom detection and clinical assessment. PTSD is widely recognized as a significant mental health concern. This study was designed to examine the relationship between PTSD research and AI methods, with a particular focus on NLP. Systematic searches were conducted in four major databases: Web of Science, IEEE Xplore, ScienceDirect, and PubMed. The databases were subjected to a comprehensive examination through the application of systematic literature review and bibliometric analysis techniques. With regard to the bibliometric analysis, the Web of Science database was queried with the pertinent keywords, resulting in the retrieval of 4,138 documents. In the context of the systematic literature review, 328 documents were accessed through the use of relevant keywords across four databases, namely Web of Science, IEEE, ScienceDirect and PubMed. The systematic literature review was structured around four research questions, which are outlined in Table 8 (Chapter 3). The first of these, "What are the challenges associated with the utilization of NLP for the classification of PTSD?", was addressed under the following thematic headings. The corresponding findings are presented in Table 13.

Table 13. Challenges encountered

Challenge Category	Description
Data quality and diversity	Clinical data is often of high quality, whereas data from unstructured sources like social media may have lower quality. This makes it challenging to accurately understand emotional and linguistic contexts.
Linguistic and contextual complexity	Variations in language and contextual differences between data sources can make it difficult for NLP techniques to accurately capture emotional and traumatic expressions.
Data security and privacy	Handling health data involves significant concerns regarding data security and privacy. Protecting data requires compliance with legal and ethical standards.
Lexical and semantic parsing	Highlights the importance of semantic parsing and linguistic features in text. Semantic parsing can affect the accuracy of NLP models, especially as individuals use different linguistic expressions to describe trauma experiences, making the parsing process complex.

Challenges in detecting PTSD using NLP involve several factors, from data quality to the selection of algorithms. Major factors include data cleaning and standardization, linguistic and contextual complexities, dataset size and diversity, choice of classification algorithms, and semantic parsing. Future studies should focus on investigating various strategies and improvements for overcoming these challenges and developing more accurate and reliable NLP-based PTSD detection methods. As for the second research question, "To what extent is AI used in detecting PTSD within the healthcare system?", the answers can be seen in Table 14.

Table 14. Health system integration

Criterion	Description
Early detection and intervention	AI systems can quickly analyze large datasets to detect PTSD symptoms early, which can accelerate early intervention and treatment processes.
Big data analysis	AI can process large datasets and identify emotional patterns. It can analyze data from sources such as social media and EHR.
Personalized treatment approaches	AI can provide customized treatment recommendations based on individual data and past experiences, making treatment processes more effective.
Data quality and security	The quality and security of health data are crucial. Concerns about data privacy and security can complicate the protection of personal information.

This holds a number of advantages with regard to health systems for the detection of PTSD using AI. The benefits it confers due to early diagnosis, scalability, personalized treatment, and holistic data analysis can help manage the condition more effectively. However, challenges such as data quality, accuracy of algorithms, cultural and linguistic barriers, and ethical issues also exist. Overcoming these challenges will grant new frontiers of leveraging AI systems into healthcare applications safely and more reliably. As for another question, "What are the datasets used in PTSD detection? What are the challenges of these datasets?", the answers are shown in Table 15.

The datasets for diagnosing PTSD are full of challenges. EHR consists of patients' medical histories and treatment processes. However, while analyzing these datasets, the main challenges found were those related to a lack of data standards, incomplete data, and data security. Crisis hotline conversations have other issues, which involve personal information protection and diversity in language accommodation. In contrast, social media data is marked by linguistic variations and noisy data that hamper proper classification and analysis. Similarly, psychiatric notes are also linguistically and contextually diverse, which may impact the effectiveness of NLP techniques. All these datasets have other challenges, such as data anonymization and protection of personal information. These challenges may cause problems in most ways of carrying out AI, especially NLP-based methods, for PTSD and highly underline careful processing and standardization of data. Beyond general considerations, the ethical challenges of applying NLP to PTSD data are unique and multi-layered. First, trauma narratives expose both researchers and clinicians to the risk of secondary trauma, underlining the need for careful curation, anonymization, and support mechanisms. Second, algorithmic transparency is critical: PTSD detection tools must provide explainable outputs so that clinicians and patients understand the basis of risk assessments. Third, mining social media data for PTSD research raises issues of informed consent, user privacy, and the potential re-exposure of individuals to trauma content. Although open data enhance replicability, ethical safeguards such as anonymization, adherence to platform policies, and privacy-preserving methods (e.g., differential privacy and federated learning) are essential. Finally, clinical data such as EMR and VA records demand strict data governance and advanced security protocols, as breaches may have severe consequences for vulnerable populations. These considerations highlight that ethical design is not peripheral but central to NLP applications in PTSD. As for the fourth and final question, "What are the AI techniques employed in the treatment of post-PTSD? What are the opportunities and challenges presented by these techniques?", the answers are shown in Table 16.

Application of AI techniques shows great promise in detecting and treating PTSD, though it is not an obstacle-free territory. Techniques like NLP, ML, DL, and probabilistic models can bring about improvement in the management of PTSD. However, data quality issues, concerns pertaining to privacy, and model explainability might hamper the effectiveness of such techniques. Therefore, the integration of AI applications requires careful assessment and planning in order to be effectively realized. Articles under review indicated a different level of influence that NLP had on the detection and analysis of PTSD. Some common methods followed in general involve text cleaning, tokenization, lemmatization, and many other types of classification algorithms. The most common algorithms adopted include SVM and LR. Other research also went deep to conduct more advanced DNN and CNN techniques. Various strengths can be found in the high-performance models. Most of the articles use a high-performance model for the detection of PTSD. In particular, methods such as DNN and CNN have been demonstrated to be effective in understanding complex patterns and relationships. For instance, Tran and Kavuluru [52] attained favorable outcomes through the utilization of DL techniques. The utilization of extensive data sets is a further strength. The utilization of

Table 15. Dataset challenges

Dataset	Description	Challenges
EHR	Records containing patients' medical histories, treatment processes, and symptoms.	Non-standardized data, missing data, data security and privacy issues.
Crisis hotline conversations	Calls and conversations during crisis situations.	Protection of personal information, data anonymization, and diversity in spoken language.
Social media data	Messages and posts collected from social media platforms.	Variability in data quality, linguistic variation, and lack of context.
Psychiatric notes	Notes written by psychiatrists about patients.	Linguistic and contextual diversity, and lack of standardized data.
Medical notes and reports	Various medical reports and notes about patients.	Data heterogeneity, and missing or ambiguous labels.
Twitter and Reddit data	Messages and comments from users on social media.	Data noise, linguistic complexity, and mislabeling.

extensive data sets enhances the model's capacity for generalization. The findings of Maguen et al. [53], which were obtained using large-scale health data, serve to emphasize the importance of large data sets. A very critical cleaning and preprocessing of data must be performed in order for the quality and reliability of a data set to be present. Effective cleaning and preprocessing techniques employed in articles enhanced the performance of models. Among those, especially, text preprocessing techniques like cleaning, tokenization, and removal of stop words proved very effective to enhance the accuracy of results.

Table 16. Opportunities and challenges of AI techniques used for PTSD

AI Technique	Usage	Opportunities	Challenges
NLP	Analysis of patient reports, psychiatry notes, and social media data.	Extracting meaningful information from large datasets, and developing early warning systems	Data heterogeneity, linguistic variations, contextual differences, and protection of personal data.
ML	Symptom prediction and classification.	Learning capacity from large datasets, and accurate symptom classification.	Data quality and quantity, model interpretability, and generalization ability.
DL	Extracting complex features from text and speech data.	Capturing complex symptom models and high performance on large datasets.	High computational power and data requirements and model interpretability issues.
Probabilistic models	Identifying latent topics and semantic structures.	Understanding hidden structures and topics in data, and meaningful representation of texts.	Data preprocessing, hyperparameter tuning, and interpretability of results.

As summarized in Table 10, classical ML models such as SVM and LR are effective for small, structured datasets, while DL models (CNN, LSTM, and DNN) achieve higher accuracy in complex or sequential data but require extensive resources. Transformer-based models (e.g., MentalBERT) currently demonstrate the best performance in EMR-based studies, though they are dependent on large annotated corpora. Thus, model selection in PTSD detection should be guided by dataset size, structure, and computational resources. The aspects below also need more emphasis. Another source of worry is the quality and diversity of data. Quite a good number of studies have reported constraints regarding the quality and diversity of data. In particular, the limited size or scope of the data sets may affect the model's ability to generalize. For example, He et al. [43] obtained limited results using a small data set. The deployment of DL methodologies necessitates the availability of substantial computational resources and data processing capabilities. This may be a limitation in small-scale studies or projects with limited resources [52]. While the benefits of using NLP techniques in PTSD analysis are apparent, a host of challenges includes data quality issues, generalization ability, and computational requirements. Future performance enhancement to result in higher accuracy may only be realized once larger data sets and more sources of data have been utilized. Moreover, new computational power enables the methods to be applied on a wider array of data and enhances the NLP techniques themselves.

The impact of preprocessing steps on PTSD detection performance varies across data types. For example, entity recognition plays a crucial role in extracting trauma-specific concepts (e.g., "combat" and "assault"), thereby improving the sensitivity of affective feature extraction. Lemmatization and normalization reduce linguistic variability and are particularly beneficial in morphologically rich languages or noisy social media data. Conversely, aggressive stop-word removal may inadvertently eliminate emotionally charged words (e.g., "never" and "alone"), thus reducing classification sensitivity. This suggests that preprocessing pipelines should be carefully tailored. While general cleaning improves robustness, certain steps such as entity recognition and normalization are critical for effective feature capture in PTSD detection.

Although NLP technologies show promise in supporting personalized PTSD treatment, their clinical applicability remains limited. Most current models are at the research stage and have not undergone prospective clinical validation. In practice, NLP can complement existing diagnostic tools such as PCL-5 and CAPS by automatically screening EMR notes or therapy transcripts for symptom patterns, thereby reducing clinician burden. However, integration into clinical workflows requires addressing key barriers, including physician acceptance, regulatory approval, and trust issues arising from inadequate model interpretability. Pilot projects, such as NLP-assisted EMR screening initiatives in VA hospitals, illustrate potential pathways, yet broader deployment will depend on ensuring transparency, explainability, and secure integration into electronic health record systems.

6 Future Work

Results obtained from this study serve as a very good insight into the state of analyzing PTSD with NLP techniques. Nevertheless, the following recommendations could further develop the research of the subject in an area of its limitation to explore future studies.

• Diversity of data and its coverage: Larger and more diverse data should be used in the future. Studies have focused their attention on data sources such as social media and medical records. Data from diverse cultural and geographical regions can enable models to provide wider generalization, and therefore model validity assessments can be realized in a much larger population.

• Advanced NLP methods: The technique is ever-evolving in the field of NLP. With the introduction of DL methodologies and the rise of large language models, such as GPT-4 and BERT, the possibility for more advanced forms of analysis becomes possible. It is recommended that future studies investigate the potential of these advanced NLP techniques in the context of PTSD detection and further explore how such methods can be operationalized in practical applications.

• Data quality and ethics: While improving data quality, ethical aspects also require stronger standards in collecting and treating data. Most importantly, this involves ensuring that ethical legislation and privacy standards are guaranteed during the processing of personal and sensitive information. Moreover, standardized cleaning and preprocessing methods of data can sometimes help in enhancing the performance of the models as well.

• Multimodal data usage: Future research can make use of multimodal datasets that include audio, images, and other forms of data in addition to text data. Available data in this regard can help gain deep insights into the psychological state of subjects and perhaps will be more useful for the analytics involved.

Future work should focus on the integration of multimodal data (e.g., text, voice, and physiological signals) to improve the robustness of PTSD detection. However, multimodal integration faces specific challenges. Heterogeneity across modalities requires synchronization and representation learning that can bridge linguistic, acoustic, and physiological features. Moreover, labeling multimodal datasets is resource-intensive, as expert annotation is needed for both clinical and affective cues. Possible technical pathways include early, late, and hybrid fusion strategies, as well as the use of self-supervised multimodal models that reduce annotation costs. Pilot projects could involve collecting text and voice data from therapy sessions or crisis hotlines, supplemented with wearable sensor data, to test the feasibility of multimodal PTSD detection in real-world settings.

Author Contributions

Conceptualization, E.S. and E.Y.; Methodology, E.S. and E.Y.; Software, E.S.; Validation, E.Y.; Formal Analysis, E.S.; Investigation, E.S. and E.Y.; Data Curation, E.S.; Writing—Original Draft Preparation, E.S.; Writing—Review and Editing: E.S. and E.Y.; Visualization, E.S.; Supervision, E.Y.; Project Administration, E.S.; Funding Acquisition, E.S. All authors have read and agreed to the published version of the manuscript and consented to its submission to ATAIML (Advances in Technology and Artificial Intelligence in Machine Learning).

Data Availability

The data used to support the research findings are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflict of interest.

References

[1] World Health Organization, "World mental health report: Transforming mental health for all," 2022. https://www.who.int/publications/i/item/9789240049338

[2] S. Hillis, J. Mercy, A. Amobi, and H. Kress, "Global prevalence of past-year violence against children: A systematic review and minimum estimates," *Pediatrics*, vol. 137, no. 3, p. e20154079, 2016. https://doi.org/10.1542/peds.2015-4079

[3] C. A. Nelson, Z. A. Bhutta, N. B. Harris, A. Danese, and M. Samara, "Adversity in childhood is linked to mental and physical health throughout life," *BMJ*, vol. 371, 2020. https://doi.org/10.1136/bmj.m3048

[4] A. Fiorillo and P. Gorwood, "The consequences of the COVID-19 pandemic on mental health and implications for clinical practice," *Eur. Psychiatry*, vol. 63, no. 1, p. e32, 2020. https://doi.org/10.1192/j.eurpsy.2020.35

[5] R. Orozco, D. Vigo, C. Benjet, G. Borges, S. Aguilar-Gaxiola, L. H. Andrade, A. Cia, I. Hwang, R. C. Kessler, M. Piazza *et al.*, "Barriers to treatment for mental disorders in six countries of the Americas: A regional report from the world mental health surveys," *J. Affect. Disord.*, vol. 303, pp. 273–285, 2022. https://doi.org/10.1016/j.jad.2022.02.031

[6] M. Sloover, L. A. C. van Est, P. G. J. Janssen, M. Hilbink, and E. van Ee, "A meta-analysis of mentalizing in anxiety disorders, obsessive-compulsive and related disorders, and trauma and stressor related disorders," *J. Anxiety Disord.*, vol. 92, p. 102641, 2022. https://doi.org/10.1016/j.janxdis.2022.102641

[7] R. B. Goldstein, S. M. Smith, S. P. Chou, T. D. Saha, J. Jung, H. T. Zhang, R. P. Pickering, B. J. Huang, and B. F. Grant, "The epidemiology of DSM-5 posttraumatic stress disorder in the United States: Results from the National Epidemiologic Survey on Alcohol and Related Conditions-III," *Soc. Psychiatry Psychiatr. Epidemiol.*, vol. 51, no. 8, pp. 1137–1148, 2016. https://doi.org/10.1007/s00127-016-1208-5

[8] J. Weizenbaum, "ELIZA—A computer program for the study of natural language communication between man and machine," *Commun. ACM*, vol. 9, no. 1, pp. 36–45, 1966. https://doi.org/10.1145/365153.365168

[9] J. A. Wallin, "Bibliometric methods: Pitfalls and possibilities," *Basic Clin. Pharmacol. Toxicol.*, vol. 97, no. 5, pp. 261–275, 2005. https://doi.org/10.1111/J.1742-7843.2005.PTO_139.X

[10] N. J. van Eck and L. Waltman, "Software survey: VOSviewer, a computer program for bibliometric mapping," *Scientometrics*, vol. 84, no. 2, pp. 523–538, 2010. https://doi.org/10.1007/S11192-009-0146-3

[11] M. J. Page, J. E. McKenzie, P. M. Bossuyt, I. Boutron, T. C. Hoffmann, C. D. Mulrow, L. Shamseer, J. M. Tetzlaff, E. A. Akl, S. E. Brennan *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," *BMJ*, vol. 372, 2021. https://doi.org/10.1136/BMJ.N71

[12] A. Liberati, D. G. Altman, J. Tetzlaff, C. Mulrow, P. C. Gøtzsche, J. P. A. Ioannidis, M. Clarke, P. J. Devereaux, J. Kleijnen, and D. Moher, "The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration," *BMJ*, vol. 339, 2009. https://doi.org/10.1136/bmj.b2700

[13] A. Montazeri, S. Mohammadi, P. M. Hesari, H. R. Ghaemi, Z. Riazi, and S. Sheikhi-Mobarakeh, "Preliminary guideline for reporting bibliometric reviews of the biomedical literature (BIBLIO): A minimum requirements," *Syst. Rev.*, vol. 12, p. 239, 2023. https://doi.org/10.1186/s13643-023-02410-2

[14] M. E. J. Newman, "Coauthorship networks and patterns of scientific collaboration," *Proc. Natl. Acad. Sci. U.S.A.*, vol. 101, pp. 5200–5205, 2004. https://doi.org/10.1073/PNAS.0307545100

[15] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *J. Am. Soc. Inf. Sci.*, vol. 24, pp. 265–269, 1973. https://doi.org/10.1002/asi.4630240406

[16] A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nat. Med.*, vol. 25, pp. 24–29, 2019. https://doi.org/10.1038/s41591-018-0316-z

[17] F. Jiang, Y. Jiang, H. Zhi, Y. Dong, H. Li, S. Ma, Y. Wang, H. Dong, H. Shen, and Y. Wang, "Artificial intelligence in healthcare: Past, present and future," *Stroke Vasc. Neurol.*, vol. 2, no. 4, 2017. https://doi.org/10.1136/svn-2017-000101

[18] A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI," *Z. Med. Phys.*, vol. 29, no. 2, pp. 102–127, 2019. https://doi.org/10.1016/j.zemedi.2018.11.002

[19] D. S. W. Ting, L. Peng, A. V. Varadarajan, P. A. Keane, P. M. Burlina, M. F. Chiang, L. Schmetterer, , L. R. Pasquale, N. M. Bressler, D. R. Webster *et al.*, "Deep learning in ophthalmology: The technical and clinical considerations," *Prog. Retin. Eye Res.*, vol. 72, p. 100759, 2019. https://doi.org/10.1016/j.preteyeres.2019.04.003

[20] S. C. Dharmage, J. L. Perret, and A. Custovic, "Epidemiology of asthma in children and adults," *Front. Pediatr.*, vol. 7, p. 246, 2019. https://doi.org/10.3389/fped.2019.00246

[21] H. Zhang, J. Plutzky, S. Skentzos, P. Morrison, M. Mar, M. Shubina, and A. Turchin, "Discontinuation of statins in routine care settings: A cohort study," *Ann. Intern. Med.*, vol. 158, no. 7, pp. 526–534, 2013. https://doi.org/10.7326/0003-4819-158-7-201304020-00004

[22] C. H. Lee and H. J. Yoon, "Medical big data: Promise and challenges," *Kidney Res. Clin. Pract.*, vol. 36, no. 1, p. 3, 2017. https://doi.org/10.23876/j.krcp.2017.36.1.3

[23] D. Price, R. Kaplan, D. Jones, A. Freeman, S. Burden, J. Gould, J. von Ziegenweidt, M. Ali, C. King, and M. Thomas, "Long-acting muscarinic antagonist use in adults with asthma: Real-life prescribing and outcomes of add-on therapy with tiotropium bromide," *J. Asthma Allergy*, vol. 8, pp. 1–13, 2015. https://doi.org/10.2147/JAA.S76639

[24] Y. Wang, L. Wang, M. Rastegar-Mojarad, S. Moon, F. Shen, N. Afzal, S. Liu, Y. Zeng, S. Mehrabi, S. Sohn, and H. Liu, "Clinical information extraction applications: A literature review," *J. Biomed. Inform.*, vol. 77, pp. 34–49, 2018. https://doi.org/10.1016/j.jbi.2017.12.002

[25] G. Perera, M. Broadbent, F. Callard, C. K. Chang, J. Downs, R. Dutta, A. Fernandes, R. D. Hayes, M. Henderson, R. Jackson *et al.*, "Cohort profile of the South London and Maudsley NHS Foundation Trust Biomedical Research Centre (SLaM BRC) case register: Current status and recent enhancement of an electronic mental health record-derived data resource," *BMJ Open*, vol. 6, no. 3, p. e008721, 2016. https://doi.org/10.1136/bmjopen-2015-008721

[26] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Inf. Process. Manag.*, vol. 24, no. 5, pp. 513–523, 1988. https://doi.org/10.1016/0306-4573(88)90021-0

[27] W. B. Cavnar and J. M. Trenkle, "N-gram-based text categorization," *Proc. 3rd Annu. SDAIR Symp. Doc. Anal. Inf. Retr.*, vol. 161175, p. 14, 1994. https://doi.org/10.1007/978-3-642-87059-9_11

[28] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to Information Retrieval*. Cambridge, UK: Cambridge University Press, 2008. https://nlp.stanford.edu/IR-book/

[29] K. W. Church, "Word2Vec," *Nat. Lang. Eng.*, vol. 23, no. 1, pp. 155–162, 2017. https://doi.org/10.1017/S1351324916000334

[30] T. Yao, Z. Zhai, and B. Gao, "Text classification model based on fastText," in *2020 IEEE International Conference on Artificial Intelligence and Information Systems (ICAIIS), Dalian, China*, 2020, pp. 154–157. https://doi.org/10.1109/ICAIIS49377.2020.9194939

[31] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," *Adv. Neural Inf. Process. Syst.*, vol. 26, pp. 3111–3119, 2013.

[32] J. Pennington, R. Socher, and C. D. Manning, "GloVe: Global vectors for word representation," in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2014, pp. 1532–1543. https://aclanthology.org/D14-1162.pdf

[33] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *J. Artif. Intell. Res.*, vol. 37, pp. 141–188, 2010. https://doi.org/10.1613/jair.2934

[34] J. C. Campbell, A. Hindle, and E. Stroulia, "Latent Dirichlet allocation: Extracting topics from software engineering data," in *The Art and Science of Analyzing Software Data*, 2015, pp. 139–159. https://doi.org/10.1016/B978-0-12-411519-4.00006-9

[35] P. M. Gavali and S. K. Shiragave, "Text representation for sentiment analysis: From static to dynamic," in *2023 3rd International Conference on Smart Data Intelligence (ICSMDI), Trichy, India*, 2023, pp. 99–105. https://doi.org/10.1109/ICSMDI57622.2023.00025

[36] B. McCann, J. Bradbury, C. Xiong, and R. Socher, "Learned in translation: Contextualized word vectors," in *Advances in Neural Information Processing Systems (NIPS 2017)*, vol. 30, 2017. https://proceedings.neurips.cc/paper_files/paper/2017/file/20c86a628232a67e7bd46f76fba7ce12-Paper.pdf

[37] O. Melamud, J. Goldberger, and I. Dagan, "context2vec: Learning generic context embedding with bidirectional LSTM," in *Proceedings of the 20th SIGNLL Conference on Computational Natural Language Learning (CoNLL 2016)*, 2016, pp. 51–61. https://doi.org/10.18653/v1/k16-1006

[38] K. Ashihara, T. Kajiwara, Y. Arase, and S. Uchida, "Contextualized context2vec," in *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT@EMNLP 2019)*, 2019, pp. 397–406. https://doi.org/10.18653/V1/D19-5552

[39] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling with Recurrent Neural Networks*.

Berlin, Heidelberg: Springer, 2012. https://doi.org/10.1007/978-3-642-24797-2_4
[40] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997. https://doi.org/10.1109/78.650093
[41] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017. https://papers.nips.cc/paper/7181-attention-is-all-you-need
[42] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” *Proc. 2019 Conf. North Am. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (NAACL-HLT 2019)*, vol. 1, pp. 4171–4186, 2019. https://doi.org/10.18653/v1/N19-1423
[43] Q. He, B. P. Veldkamp, and T. de Vries, “Screening for posttraumatic stress disorder using verbal features in self narratives: A text mining approach,” *Psychiatry Res.*, vol. 198, no. 3, pp. 441–447, 2012. https://doi.org/10.1016/j.psychres.2012.01.032
[44] B. Shiner, L. W. D’Avolio, T. M. Nguyen, M. H. Zayed, B. V. Watts, and L. Fiore, “Automated classification of psychotherapy note text: Implications for quality assessment in PTSD care,” *J. Eval. Clin. Pract.*, vol. 18, no. 3, pp. 698–701, 2011. https://doi.org/10.1111/j.1365-2753.2011.01634.x
[45] S. J. Fodeh, M. Zirkle, D. Finch, R. Reeves, J. Erdos, and C. Brandt, “MedCat: A framework for high level conceptualization of medical notes,” in *2013 IEEE 13th International Conference on Data Mining Workshops, Dallas, TX, USA*, 2013, pp. 274–280. https://doi.org/10.1109/ICDMW.2013.89
[46] B. Shiner, L. W. D’Avolio, T. M. Nguyen, M. H. Zayed, Y. Young-Xu, R. A. Desai, P. P. Schnurr, and L. Fiore, “Measuring use of evidence based psychotherapy for posttraumatic stress disorder in a large national healthcare system,” *Adm. Policy Ment. Health*, vol. 40, no. 4, pp. 311–318, 2013. https://doi.org/10.1007/s10488-012-0421-0
[47] M. Pacula, T. Meltzer, M. Crystal, A. Srivastava, and B. Marx, “Automatic detection of psychological distress indicators and severity assessment in crisis hotline conversations,” in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy*, 2014, pp. 4863–4867. https://doi.org/10.1109/ICASSP.2014.6854526
[48] B. V. Watts, B. Shiner, L. Zubkoff, E. Carpenter-Song, J. M. Ronconi, and C. M. Coldwell, “Implementation of evidence-based psychotherapies for posttraumatic stress disorder in VA specialty clinics,” *Psychiatr. Serv.*, vol. 65, no. 5, pp. 648–653, 2014. https://doi.org/10.1176/appi.ps.201300176
[49] N. Howard, L. Jehel, and R. Arnal, “Towards a differential diagnostic of PTSD using cognitive computing methods,” in *2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing, London, UK*, 2014, pp. 9–20. https://doi.org/10.1109/ICCI-CC.2014.6921435
[50] Q. He, B. P. Veldkamp, C. A. W. Glas, and T. de Vries, “Automated assessment of patients’ self-narratives for posttraumatic stress disorder screening using natural language processing and text mining,” *Assessment*, vol. 24, no. 2, pp. 157–172, 2015. https://doi.org/10.1177/1073191115602551
[51] A. M. Redd, A. V. Gundlapalli, G. Divita, M. E. Carter, L. T. Tran, and M. H. Samore, “A pilot study of a heuristic algorithm for novel template identification from VA electronic medical record text,” *J. Biomed. Inform.*, vol. 71, pp. S68–S76, 2017. https://doi.org/10.1016/j.jbi.2016.07.019
[52] T. Tran and R. Kavuluru, “Predicting mental conditions based on “history of present illness” in psychiatric notes with deep neural networks,” in *Proceedings of the 5th Workshop on Noisy User-Generated Text (W-NUT 2019)*, 2017, pp. S138–S148. https://doi.org/10.18653/V1/D19-5552
[53] S. Maguen, E. Madden, O. V. Patterson, S. L. DuVall, L. A. Goldstein, K. Burkman, and B. Shiner, “Measuring use of evidence based psychotherapy for posttraumatic stress disorder in a large national healthcare system,” *Adm. Policy Ment. Health*, vol. 45, no. 4, pp. 519–529, 2018. https://doi.org/10.1007/s10488-018-0850-5
[54] K. P. Norman, A. Govindjee, S. R. Norman, M. Godoy, K. L. Cerrone, D. W. Kieschnick, and W. Kassler, “Natural language processing tools for assessing progress and outcome of two veteran populations: Cohort study from a novel online intervention for posttraumatic growth,” *JMIR Form. Res.*, vol. 4, no. 9, p. e17424, 2020. https://doi.org/10.2196/17424
[55] M. A. U. Alam and D. Kapadia, “LAXARY: A trustworthy explainable twitter analysis model for post-traumatic stress disorder assessment,” in *Proceedings of the 2020 IEEE International Conference on Smart Computing (SMARTCOMP 2020)*, 2020, pp. 308–313. https://doi.org/10.1109/SMARTCOMP50058.2020.00069
[56] S. Maguen, N. Holder, E. Madden, Y. Li, K. H. Seal, T. C. Neylan, C. Lujan, O. V. Patterson, S. L. DuVall, and B. Shiner, “Evidence-based psychotherapy trends among posttraumatic stress disorder patients in a national healthcare system, 2001–2014,” *Depress. Anxiety*, vol. 37, no. 4, pp. 356–364, 2020. https://doi.org/10.1002/da.22983
[57] A. S. Uban, B. Chulvi, and P. Rosso, “Multi-aspect transfer learning for detecting low resource mental disorders on social media,” in *Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC*

2022), 2022, pp. 3202–3219.
[58] J. R. Oltmanns, H. A. Schwartz, C. Ruggero, Y. Son, J. Miao, M. Waszczuk, S. A. P. Clouston, E. J. Bromet, B. J. Luft, and R. Kotov, "Artificial intelligence language predictors of two-year trauma-related outcomes," *J. Psychiatr. Res.*, vol. 143, pp. 239–245, 2021. https://doi.org/10.1016/j.jpsychires.2021.09.015
[59] B. Shiner, M. Levis, V. M. Dufort, O. V. Patterson, B. V. Watts, S. L. DuVall, S. J. Russ, and S. Maguen, "Improvements to PTSD quality metrics with natural language processing," *J. Eval. Clin. Pract.*, vol. 28, no. 4, pp. 520–530, 2022. https://doi.org/10.1111/jep.13587
[60] C. Shurafa, "Detecting and inoculating against PTSD in online forums and platforms," in *2021 8th International Conference on Behavioral and Social Computing (BESC), Doha, Qatar*, 2021, pp. 1–3. https://doi.org/10.1109/BESC53957.2021.9635325
[61] H. Zafari, L. Kosowan, F. Zulkernine, and A. Signer, "Diagnosing post-traumatic stress disorder using electronic medical record data," *Health Inform. J.*, vol. 27, no. 4, 2021. https://doi.org/10.1177/14604582211053259
[62] R. Elankeerthana, D. Hanshiya, A. Pranavajothi, A. Shanmugapriya, and S. Sujitha, "Child abuse mental symptom prediction model using machine learning techniques," in *2022 6th International Conference on Trends in Electronics and Informatics (ICOEI 2022), Tirunelveli, India*, 2022, pp. 1459–1465. https://doi.org/10.1109/ICOEI53556.2022.9777138
[63] J. Sawalha, M. Yousefnezhad, Z. Shah, M. R. G. Brown, A. J. Greenshaw, and R. Greiner, "Detecting presence of PTSD using sentiment analysis from text data," *Front. Psychiatry*, vol. 12, p. 811392, 2022. https://doi.org/10.3389/fpsyt.2021.811392
[64] L. Kosowan, A. Singer, F. Zulkernine, H. Zafari, M. Nesca, and D. Muthumuni, "Pan-Canadian electronic medical record diagnostic and unstructured text data for capturing PTSD: Retrospective observational study," *JMIR Med. Inform.*, vol. 10, no. 12, p. e41312, 2022. https://doi.org/10.2196/41312
[65] A. Bartal, K. M. Jagodnik, S. J. Chan, M. S. Babu, and S. Dekel, "Identifying women with postdelivery posttraumatic stress disorder using natural language processing of personal childbirth narratives," *Am. J. Obstet. Gynecol. MFM*, vol. 5, no. 3, p. 100834, 2023. https://doi.org/10.1016/j.ajogmf.2022.100834
[66] M. Levis, J. Levy, V. Dufort, C. J. Russ, and B. Shiner, "Dynamic suicide topic modelling: Deriving population-specific, psychosocial and time-sensitive suicide risk variables from electronic health record psychotherapy notes," *Clin. Psychol. Psychother.*, vol. 30, no. 4, pp. 795–810, 2023. https://doi.org/10.1002/cpp.2842