The Journal of Desk Research Review and Analysis, Vol. 1, Issue II, 2023, 103-109.

A NARRATIVE REVIEW ON HUMAN ANXIETY DETECTION IN SOCIAL MEDIA USING NATURAL LANGUAGE PROCESSING

Ashala Senanayake[1] and Sidath Liyanage[2]

Abstract

Currently, with the boundless proliferation of social media platforms, it is undoubted that the world is in the middle of a significant phase of social media evolution. These virtually created user environments allow users to create, share, and interact with content. Most people who interact with social media make a pitch to share their mood, tension, feelings, and behaviour unhesitatingly with the community. Mostly, these data can be taken as a mirror that reflects the mental health status of a person, such as stress, anxiety, and suffering. With increased social media penetration, people tend to share their disturbances rather than suffer alone. Anxiety can be taken as an ordinary human emotion which prepares the human body for potentially vulnerable situations. It would be beneficial if there were any system that could predict if a person is anxious before going through a critical situation that would have to be clinically treated. This study is a narrative review of anxiety detection over the past ten years. Furthermore, this qualitative study discusses an overview of anxiety detection using Natural Language Processing (NLP) and key concepts highlighting emerging trends. We conclude that the existing state-of-the-art social media anxiety detection mechanisms can be outperformed using deep learning based on large pre-trained language models. Furthermore, it can be safely stated that a specific pre-trained large language model for social media anxiety detection is a crucial necessity yet to be fulfilled.

Keywords: Anxiety Detection, Natural Language Processing, Social Media, Social Networking

---

[1]Faculty of Computing & Technology, University of Kelaniya

Email: ashusenanayake@gmail.com https://orcid.org/0009-0005-4634-1657

[2] Senior Lecturer, Faculty of Computing & Technology, University of Kelaniya

Email: sidath@kln.ac.lk https://orcid.org/0000-0003-0778-0874

Accepted the revised version: 02 December 2023

Introduction

Social media platforms have evolved with various technological and innovative advancements over several decades, creating a rich history. From early online Bulletin Board Systems (BBS), which launched around the 1970s and 1980, up to TikTok, which launched in 2016, the social media paradigm has expanded gradually, passing several milestones like the launching of Friendster in 2002, LinkedIn in 2003, Facebook in 2004, YouTube in 2005, Twitter in 2006, WhatsApp in 2009, Instagram in 2010 and Snapchat in 2011. This evolution showcases the change in technological advancements, user interactions, the desire to share content and the desire to consume content in this interconnected world. The key feature of social media platforms is the ability to generate content in multiple forms. They could be texts, images, audio, videos and many more. Another key feature of social media is networking, or the ability to interconnect with people regardless of demographic barriers. This results in the opportunity to interact with each other to share, comment, like and join discussions via shared content. Furthermore, there are many other vital features like real-time communication, virality of content or rapid sharing and this opens various paths for business and marketing opportunities.

According to Statistica[1] more than half of the world's population actively consume social media. After the outbreak of the Covid-19 pandemic, people's day-to-day engagements have been disrupted by the sudden spread of the plague and the lockdown enforced by the governments. This has caused a profound impact on the exponential growth in world social media penetration. Social media penetration is the proportion of social media active users in a country or region. This resulted in an exponential growth of people sharing their social media content. For instance, personal updates like birthdays, anniversaries, travel experiences, professional updates and their photos, videos, invitations, awareness and more over personal opinions like inspirational quotes, views, recommendations and most importantly, feelings. Most of these data can be taken as a mirror that reflects the mental health status of a person, such as stress, anxiety, and suffering.

With the current socio-economic issues, people undergo many emotional disturbances. Those might be because of a lack of access to quality health care, relationship insecurities, stressful backgrounds like unemployment, financial instabilities, issues with education, issues with nutrition supplies, job insecurity, poor living conditions and many more. With increased social media penetration, people tend to share their disturbances rather than suffer alone.

Among the emotional disturbances that prevail in the community, depression, anxiety, Post-Traumatic Stress Disorder (PTSD), and Bipolar Disorder are prevalent emotional disturbances. According to a statistical survey which was conducted in the United Arab Emirates, with the collected data in 2018, which was before the outbreak of the Covid-19 pandemic, the prevalence rate of emotional disorders like anxiety, depression and stress disorders has been identified as 55%, 38% and 29% respectively. According to, these numbers have been 47.1%, 27.1% and 27%, respectively, within a group of Turkish students and according to, respectively 21%, 41% and 27% with a group of tertiary educational students of Hong Kong. These results indicate the high prevalence of anxiety among other emotional disturbances, which showcases the importance of researching a robust computational machine learning model for detecting anxiety.

Anxiety is a general human emotion that prepares the human body for situations that could potentially pose a risk. They can be either threats or stressors. Simply, the emotion of anxiety helps to cope with today's challenges, such as getting ready for an exam. Feelings like worry, stress and nervousness can arise while preparing for an exam. These feelings could be because of the pressure of expecting good

[1] https://www.statista.com/

grades at exams, fear of failing exams, load of content to cover, uncertainty of questions and many more. So, it is clear that some level of anxiety is normal and even helpful for getting prepared for challenging situations like exams. However, an extensive amount of anxiety could be counterproductive when it is excessive, lasting, and disturbing to one's daily life.

Social media anxiety detection is the process of identifying an individual with anxiety considering social media engagements like post sharing. This has become a crucial research area in the social media research arena, even though most of the research has been carried out about emotional disturbances like depression and suicidal tendencies. Two research questions will be addressed through our narrative review:

RQ1: What are the established NLP-based computational mechanisms for social media anxiety detection?

RQ2: How can the detection of social media anxiety be further improved using modern NLP trends?

Methodology

Social media anxiety detection becomes crucial for many reasons. Early identification of anxiety leads to timely intervention and getting support from relevant parties. By the identification of individuals at risk of anxiety, mental support consultants could give personalized support. Social media anxiety detection provides insight into the mental well-being of a certain population. This helps public health officials and policymakers to be aware of the prevalence and distribution of anxiety-related issues in order to develop the mental well-being of the community. Normalizing and discussing the issues related to anxiety helps to break the taboo in society. Furthermore, social media anxiety detection is a growing research area that highly helps digital mental health. According to a narrative review carried out by (Zhang et al., 2022) on NLP-applied mental disturbance detection 2022, a wide range of research has been conducted on emotional disturbances like depression (45%), suicide (20%) and stress (6%). Considerably, a pinch of research work has been conducted on anxiety detection, which is around 2%. Social media anxiety detection can be recognized using different mechanisms as follows.

Anxiety Detection Using Sentiment Analysis

Sentiment analysis is one of the main Natural Language Processing (NLP) approaches which is used to determine the emotional tone of a certain text. Sentiment analysis is widely used to discover a person's feelings towards a product, person, event, or topic. In this approach, anxiety is identified based on the negative, positive, and neutral comments of a person in which the negative data is identified as anxiety. In contrast, positive and neutral data are identified as non-anxiety. This mechanism has been applied to identify the anxiety of people during the COVID-19 period. (Saifullah et al., 2021) We have used social media comments regarding government campaigns to deal with the pandemic. One of the main limitations of this approach is the lack of contextual understanding.

Anxiety Detection using Emotion Detection

Beyond the sentiment analysis approach, emotion detection focuses more on the emotion submerged within the texts. Emotions which are associated with anxiety, like fear and sadness, provide more insights into the text with regards to anxiety. One of the main challenges of this approach is emotions being subjective concepts with fuzzy boundaries. EmotexStream is an application that classifies emotions in real-time to identify emotional disturbances like anxiety (Hasan et al., 2019). The difficulty of handling contextual ambiguity is one of the drawbacks of this approach.

Anxiety Detection Using Language Analysis

Language and vocabulary analysis is another technique applied in anxiety detection. The usage of specific terms, count of repetitive words, frequency of negative words, and excessive use of certain pronouns are considered in this technique. Multiple research studies have been carried out to find out the linguistic correlation of social anxiety. Hofmann et al. (2012) researched linguistic correlates of Social-media Anxiety Disorder (SAD). Demographic linguistic variations are one of the main challenges in this study.

Anxiety Detection using Topic Modeling

Topic modelling is also a commonly used mechanism in anxiety detection. In this approach, common themes or topics used in the social media posts are considered. Here, the common issues or subjects that individuals are discussing might include emotional distress or factors related to anxiety. Algorithms like Latent Dirichlet Allocation (LDA) by (Blei et al., 2003) and Non-Negative Matrix Factorization (NMF) (Lee & Seung, 2000) were used as the topic modelling algorithms. LDA is a popular generative probabilistic modelling algorithm identifying the topics distributed within a text. Topic modelling is an unsupervised text mining approach which discovers topics over a corpus of texts. (Tyshchenko, 2018) has used LDA for topic modelling in anxiety and depression detection using blog post data. Shen & Rudzicz (2017) Have detected anxiety on top of topic modelling using Reddit as their data source. Furthermore, they have identified that combining Linguistic Inquiry and Word Count (LIWC) (Pennebaker et al., 2001) and N-gram elevates classification accuracy to a greater extent. Besides contextual ambiguity, the lack of ground truth labels becomes challenging in this approach.

Anxiety Detection using Word Embeddings

The usage of word embeddings for anxiety detection could be found in earlier research. Word embeddings capture the semantic relationship between words, which could be used to detect word clusters related to anxiety. Word2Vec (Mikolov et al., 2013) and GloVe (Pennington et al., 2014) are word embedding mechanisms. A specific research study was conducted on detecting anxiety in political texts extracted from Canadian parliament Hansard records conducted with GloVe. This unsupervised algorithm generates word embeddings (Rheault, 2016). The capability of contextual understanding is less in traditional word embeddings, which has become a primary challenge.

Anxiety Detection Using Deep Learning

Currently, deep learning plays a specific role in anxiety detection from texts. Recurrent Neural Networks (RNN) perform well with sequential data, such as a sequence of words in a sentence (Rumelhart et al., 1986). Long Short-Term Memory (LSTM) is a unique form of RNN specifically used to detect sequential dependencies in data (Hochreiter, 1997). Bidirectional LSTM models have been designed to identify better emotional distress (Winata et al., 2018). Word embeddings are commonly used as inputs for deep learning models as they enhance the capability of deep learning models to understand and process textual data by transfer learning.

Furthermore, Transformers have become the latest breakthrough deep learning model in the field of NLP. Unlike the traditional RNN models and much more advanced deep learning models like LSTM and Gated Recurrent Units (GRU), the Transformers do not rely on sequential processing. Instead, these pre-trained language models (PLM) use a special mechanism called 'self-attention', which allows transformers to identify dependencies between words regardless of the position of the sentence (Vaswani et al., 2017). Transformer-based Large Language Models (LLM) have become the state-of-the-art (SOTA) in NLP due to their exceptional ability to understand, classify, and generate texts. Bidirectional encoder representations from transformers (BERT) and generative pre-trained transformers (GPT) are such as SOTA PLM. Being an encoder type, BERT supports the text classification task (Devlin et al., 2018), while GPT supports the text generation task like summarization as its decoder type (Radford et al., 2018). When it comes to anxiety detection, since it is a classification-based task, BERT-based approaches have become the prime focus among researchers in the modern world. Vanilla BERT has become the foundation for many LLM builds intending to handle different tasks like Bio BERT for biological concept analysis (Lee et al., 2020), Clinical-BERT for biomedical text analysis (Alsentzer et al., 2019), SciBERT for scientific literature analysis (Beltagy et al., 2019). MentalBERT is such an LLM which was introduced for mental health care (Ji et al., 2021). Although many task-specific BERT-based LLMs have been introduced, a specific LLM for anxiety detection has yet to be discovered.

Based on the literature conducted above, the following predictive models in Table 1 have shown state-of-the-art performance in each anxiety detection approach separately.

Table1: NLP Techniques and reported experiment results.

NLP Technique	Methodology	Results	
Sentiment Analysis	Random Forest (RF), Term Frequency–Inverse Document Frequency (TF-IDF), Vector Count (VC)		Accuracy
		TF-IDF + RF	82.63 %
		VC + RF	84.99%
Emotion Detection	EmotexStream	Accuracy - 90%	
Language Analysis	LIWC computer program	Individuals with generalized SAD had higher adverse effects than the non-anxious controls.	
Topic Modeling	LDA and NMA	LDA has increased classification accuracy up to 98%.	
Word Embeddings	SVM	Accuracy - 74.6%	
Deep Learning	BERT	Accuracy - 68.44%	

Discussion

Social media anxiety detection using NLP has been a dynamic and rapidly evolving field over the past decade. Drastically, the attention of researchers has been drawn towards this field as they have identified the importance of addressing mental health concerns, including anxiety, within the social media context. This review is a narrative review consisting of a qualitative study. Through this review, five main anxiety detection approaches are identified, and their research studies are presented. They are anxiety detection using sentiment analysis, emotion detection, language analysis, Topic modelling, machine learning techniques like word embeddings and deep learning mechanisms. According to the analysis, the current social media anxiety detection trend focuses considerably on deep learning-based approaches. Although pre-trained large language models have shown state-of-the-art performance in various domains, limited work has been conducted in anxiety detection. Enhancing the performance of social media anxiety detection using PLM would be a promising future research avenue.

Conclusion

Considering the overall analysis, it can be concluded that the *RQ1: What are the established NLP-based computational mechanisms for social media anxiety detection?* is answered. To our knowledge, this narrative review can be considered the most significant work conducted so far for analyzing the NLP-based computational approaches for social media anxiety detection. This analysis can be extended to explore further investigations concerning other review mechanisms such as systematic reviews, scoping reviews, etc.

It can be concluded that a PLM-based large language model is one of the most prominent ways to enhance the performance of social media anxiety detection computationally. Based on this narrative review, the most prominent mechanisms for social media anxiety detection are summarized. Our idea is to enhance the performance of those systems by adapting a transfer learning mechanism on top of an existing encoder-type large language model. Transferring knowledge of a PLM from one domain to another has shown some significant results, so we hope to use such a mechanism in future (Zhang et al., 2022). Considering these novel mechanisms, it can be concluded that the *RQ2: How can the detection of social media anxiety be further improved using modern NLP trends?* is also answered.

Our system will be used only as a screening tool for early identification of human anxiety, and this should not be used for any clinical-based diagnosis purposes. Also, we recommend that the proposed model should only be tested with million (1M) parametric-based PLMs and not be used for the experimental purposes of billion (1B) parametric-based PLMs.

References

Alsentzer, E., Murphy, J. R., Boag, W., Weng, W.-H., Jin, D., Naumann, T., & McDermott, M. (2019).

Publicly available clinical BERT embeddings. arXiv preprint arXiv:1904.03323.

Bayram, N., & Bilgel, N. (2008). The prevalence and socio-demographic correlations of Depression, anxiety and stress among a group of university students. Social psychiatry and psychiatric epidemiology.

Beltagy, I., Lo, K., & Cohan, A. (2019). SciBERT: A pre-trained language model for Scientific text. arXiv preprint arXiv:1903.10676.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. Journal of machine Learning Research, 3, 993--1022.

Cohan, A., Desmet, B., Yates, A., Soldaini, L., MacAvaney, S., & Goharian, N. (2018). SMHD: a large-scale resource for exploring online language usage for multiple mental health conditions. arXiv preprint arXiv:1806.05258.

Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep Bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Hasan, M., Rundensteiner, E., & Agu, E. (2019). Automatic emotion detection in text streams by analyzing Twitter data. International Journal of Data Science and Analytics, 7(Springer), 35--51.

Hochreiter, S. a. (1997). Long short-term memory. Neural Computation, 9, 1735--1780.

Hofmann, S. G., Moore, P. M., Gutner, C., & Weeks, J. W. (2012). Linguistic correlates of Social anxiety disorder. Cognition & emotion, 26(Taylor & Francis), 720--726.

Ji, S. A., Ansari, L., Fu, J., Tiwari, P., & Cambria, E. (2021). Mentalbert: Publicly available Pre-trained language models for mental healthcare. arXiv preprint arXiv:2110.15621.

Lee, D., & Seung, H. S. (2000). Algorithms for non-negative matrix factorization. Advances in neural information processing systems.

Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-Trained biomedical language representation model for biomedical text mining. Bioinformatics, 36.

Marzouqi, A. M., Otim, M. E., Alblooshi, A., Marzooqi, S., Talal, M., & Wassim, F. (2022). State of Emotional Health Disorders of Undergraduate Students in the United Arab Emirates: A Cross-Sectional Survey. Psychology Research and Behavior Management.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word Representations in vector space. arXiv preprint arXiv:1301.3781.

Pennebaker, J. W., Francis, M. E., & Booth, R. J. (2001). Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word Representation. Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP).

Radford, A., Narasimhan, K., Salimans, T., Sutskever, I., & others. (2018). Improving Language understanding by generative pre-training. OpenAI.

Rheault, L. (2016). Expressions of anxiety in political texts. In Proceedings of the first workshop on NLP and computational social science.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by Back-Propagating errors. Nature, 533--536.

Saifullah, S., Fauziah, Y., & Aribowo, A. S. (2021). Comparison of machine learning for Sentiment analysis in detecting anxiety based on social media data. arXiv preprint arXiv:2101.06353.

Shen, J. H., & Rudzicz, F. (2017). Detecting anxiety through Reddit. Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology—From Linguistic Signal to Clinical Reality.

Tyshchenko, Y. (2018). Depression and anxiety detection from blog posts data. Nature Precis. Sci., Inst. Comput. Sci., Univ. Tartu, Tartu, Estonia.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., . . . Polosukhin, (2017). Attention is all you need. Advances in neural information processing systems, p. 30.

Winata, G. I., Kampman, O. P., & Fung, P. (2018). Attention-based lstm for psychological Stress detection from spoken language using distant supervision. 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

Wong, J. G., Cheung, E. P., Chan, K. K., Ma, K. K., & Wa Tang, S. (2006). Web-based Survey of depression, anxiety and stress in first-year tertiary education students in Hong Kong. Australian & New Zealand Journal of Psychiatry.

Zhang, T., Schoene, A. M., Ji, S., & Ananiadou, S. (2022). Natural language processing Applied to mental illness detection: a narrative review. NPJ digital medicine, 5(Nature Publishing Group UK London).