Artificial Intelligence and Social Computing, Vol. 122, 2024, 167–174
https://doi.org/10.54941/ahfe1004653

Unveiling Mental Health Insights: A Novel NLP Tool for Stress Detection through Writing and Speaking Analysis to Prevent Burnout

Matteo Mendula[1], Silvia Gabrielli[2], Francesco Finazzi[1], Cecilia Dompé[1], and Mauro Delucis[1]

[1]MYNDOOR SRL, 20088 – Rosate (MI) – Italy
[2]FBK | FONDAZIONE BRUNO KESSLER, 38123 – Povo (TN) – Italy

ABSTRACT

Nowadays, innovative approaches that precisely identify and treat health-related problems are becoming more and more necessary in a time of rapid technological advancement and growing mental health awareness. Given the prevalence of mental health issues, different tools that employ Artificial Intelligence to support rapid and effective interventions have been developed. This study focuses on the relationship between language expression and mental health, recognizing subtle nuances in both written and spoken communication as potential stress indicators and presenting a novel AI enhanced tool for autonomous and passive stress detection. Thanks to this strong psychometric framework for correlating language manifestation of stress with clinical diagnosis, in this article we present the first, to our knowledge, NLP (Natural Language Processing) tool for autonomous and passive stress detection. This includes a variety of emotional and cognitive stress indicators to provide a deeper understanding of stress that takes into account both subjective experiences and objective manifestations. Initial results show a strong relationship between the biomedical markers and the stress scores obtained from language analysis. By combining data science techniques with psychometric insights, our stress detection achieves 81.7% and 83.4% F1 score for the english and italian language, respectively.

Keywords: Exemplary paper, Human systems integration, Systems engineering, Systems modeling language

INTRODUCTION

The ongoing discourse surrounding the intricate interplay between natural intelligence and artificial intelligence continues to captivate the attention of researchers and intellectuals. Recent strides in the field of electronics have bestowed upon us an ever-expanding computational prowess. This newfound computational might, when harmoniously coupled with existing mathematical paradigms, has not only empowered AI systems to rival but, in some cases, even surpass our conventional benchmarks for human intelligence.

Regrettably, this approach yields remarkable outcomes primarily in scenarios where problem requirements are amenable to structured modeling, such as games, pipeline management, anomaly detection, and the like. However, the same techniques do not seamlessly translate to addressing human-centric issues like mental health. Here, the domain knowledge required to help a real patient from a physiological point of view makes the adoption of data science impracticable. In fact, the deep and profound understanding of the human brain cannot be easily coded into a well defined modeling which can be interpreted by a machine.

While stress may not be formally classified as a mental disorder, a wealth of research studies has unequivocally illustrated its profound impact on both the emergence and exacerbation of mental disorders (Cockerham 2020, Hyman 2010, Widiger 2005). According to the American Psychological Association (APA, 2018), stress is a typical response to the demands of daily life, but it can become detrimental when it disrupts our day-to-day activities. In such instances, stress can make it challenging to unwind and can be accompanied by a spectrum of emotions, such as anxiety and irritability. Stressful situations can also trigger or worsen mental health conditions, particularly anxiety and depression, necessitating access to healthcare (Marin et al., 2011).

With this notion at its core, we have adopted a human-centered approach, leveraging AI as a supportive tool for the proactive identification and assessment of individuals' stress levels, using the most ubiquitous mode of contemporary communication: written text. Specifically, in our study data scientists and psychologists collaborate to create and validate a ground-breaking knowledge base. This innovative database combines psychometrics, biometrics, and linguistic analysis to provide a comprehensive evaluation of stress levels. We used biomedical indicators, such as blood pressure, heart rate variability (HRV), and cortisol levels correlations to validate the results. The multidisciplinary team brought together expertise from data science and psychology to create a novel database with a wide range of sentences that have been annotated with matching stress levels.

Figure 1: Master Sgt. Khadija Ross interviews a patient seeking care Feb. 20, 2013, at the Mental Health Clinic on Dover Air Force Base (adapted from Militarytimes, 2015).

The work presented here encapsulates the culmination of seven years of painstaking data collection across a spectrum of clinical trials and exploration expeditions spanning the globe. In these endeavors, carefully selected individuals diligently recorded their daily experiences, complemented by a battery of stress questionnaires. The objective was to unravel the intricate relationship between word choice and stress levels.

The results obtained in our studies have been analyzed looking for correlations among the data and highlighting the most representative lemmas when identifying stress level of observed individuals.

The fusion of expert domain knowledge with machine learning techniques makes our project a pioneering innovation within the realm of psychology, allowing the adoption of AI to stress identification and expanding the capabilities of data science to non STEM related issues.

MOTIVATIONS

Self-reported questionnaires have long been critiqued for their inherent subjectivity (Wakefield, 2016), leading to a notable gap between the subjective experience of individuals and the underlying neurobiological processes. However, in recent decades, the remarkable advancement of computing capabilities, coupled with the accumulation of extensive neuroimaging datasets, has ushered in a new era. This era allows researchers to bridge the divide between subjective experience and neurobiology by harnessing AI to identify, model, and potentially address developmental and psychiatric disorders.

In this context, while significant progress has been made in assessing stress through facial expressions, tone of voice, and biometric data, there remains a pressing need to explore these disorders from the vantage point of lexical and syntactic choices in language.

It is entirely plausible to envision a future where AI assumes a pivotal role in the treatment of psychiatric disorders and cognitive development. Computer-assisted therapy (CAT), featuring AI-driven chatbots offering cognitive behavioral therapies, is already undergoing testing for the treatment of various psychiatric conditions, including depression and anxiety (Carroll and Rounsaville, 2010; Fitzpatrick et al., 2017; Fulmer et al., 2018). This emerging paradigm holds the promise of revolutionizing the way we understand and manage mental health and cognitive well-being. By using natural language processing and AI, chatbots can provide timely personalized therapeutic conversations to deliver clinically validated treatment which can enhance the therapeutic alliance and improve clinical outcomes. An advantage of digital interventions delivered by chatbots is that they are available 24/7, can provide an anonymous and safe environment, and can reduce stigma associated with seeking therapy. Moreover, they can be used through stepped-care approaches (Richards et al., 2012) to empower patients in the self-management of their diseases and to allow clinical professionals to focus on high-risk cases by freeing their time on low-risk cases and routinary tasks. In this regard, AI-powered digital intervention chatbots can be beneficial to support the achievement of therapeutic goals and decision-making during treatment.

Stress and Related Factors

Although we are used to considering stress as a symptom unrelated to others, recent studies have shown that in reality from the patient's point of view, a state of stress is instead the sum of a series of highly correlated factors such as anxiety, depression and physical tiredness (Narasappa Kumaraswamy, 2013, Markus Heilig, 2004). The sleep-wake cycle, nutrition, a regular lifestyle and a serene working climate are just some of the parameters that a support algorithm for the classification of the state of psychological stress has the task of considering in order to reach accuracy and confidence intervals such as to be useful to a qualified operator.

Among the indicators taken into consideration by a psychologist are a series of questionnaires which, on the basis of the results obtained by the patient, facilitate the determination of the state of psychological stress and support the psychologist during the diagnosis process. Among these we find:

- Perceived Stress Scale (PSS): It is a classic tool for assessing stress. The tool, though originally developed in 1983, remains a popular choice for helping us understand how different situations affect our feelings and perceived stress. The questions on this scale concern the patient's feelings and thoughts during the last month. You are also asked to indicate how often the patient felt or thought in a certain way. Even though some of the questions are similar, there are strong differences between them which treat each as a separate question. There are 14 questions and the answers are in the 5-point Likert-Scale format (Strongly disagree; Disagree; Neither agree nor disagree; Agree; Strongly agree).
- Hamilton Depression Rating Scale (HDRS): is the depression rating scale most widely used by clinicians. The original version includes 17 items (HDRS17) relating to symptoms of depression experienced in the last week. While the scale was designed to be completed after an unstructured clinical interview, semi-structured interview guides are now available. HDRS was originally developed for hospital patients, so the emphasis is on the melancholic and physical symptoms of depression. A later version with 21 items (HDRS21) included 4 items intended to subtype depression, but are sometimes erroneously used to rate depression severity. A limitation of HDRS is that atypical symptoms of depression (eg, hypersomnia, hyperphagia) are not evaluated.
- Hamilton Anxiety Rating Scale (HAM-A): It was one of the first rating tools developed to measure the severity of anxiety symptoms and is still widely used today in both clinical and research settings. The scale consists of 14 questions, each defined by a set of symptoms, and measures both psychic anxiety (mental turmoil and psychological distress) and somatic anxiety (physical complaints related to anxiety). Although HAM-A remains widely used as an outcome measure in clinical trials, it has been criticized for its poor ability to discriminate between anxiolytic and antidepressant effects, as well as between somatic anxiety and somatic side effects. The HAM-A does not provide any standardized test questions.

Although all the tools available to a psychologist encourage rapid completion, there is no rating scale that takes them into account.

ARTIFICIAL INTELLIGENCE AND NATURAL LANGUAGE PROCESSING (NLP)

Natural Language Processing (NLP) is a vital field within computing, specifically under the umbrella of artificial intelligence (AI). Its primary focus lies in empowering computers to comprehend and interpret text and spoken language, akin to the way humans do.

NLP draws on a fusion of disciplines, including computational linguistics, rule-based modeling of human language, and various statistical, machine learning, and deep learning techniques. Collectively, these technologies equip computers to process human language, encompassing both textual and spoken data, and to grasp its full meaning, including nuances of intent and emotion conveyed by the speaker or author. The applications of NLP are vast, powering computer programs capable of translating text across languages, responding to voice commands, and rapidly summarizing extensive textual content, even in real-time scenarios. Moreover, NLP is playing an increasingly pivotal role in enterprise solutions, facilitating more efficient business operations, enhancing employee productivity, and streamlining critical business processes.

However, applying such an approach to stress assessment poses several challenges. Firstly, it's uncertain whether a single questionnaire can adequately represent a person's psychological stress state. Secondly, guaranteeing a robust correspondence between interviews, questionnaires, and the stress indicator values, to serve as a training database for classification algorithms distinguishing stress from non-stress states, is not assured. These complexities underscore the need for innovative solutions in the field of stress assessment.

StressSense: NLP-Based Stress Detection

StressSense (SS) is a sophisticated machine learning model meticulously crafted for the precise detection of stress levels in individuals. Leveraging an advanced and proprietary algorithm, it analyzes behavioral patterns and physiological signals in written text to provide accurate insights into the emotional well-being of users.

Our endeavors in the field of stress assessment have effectively addressed the challenges outlined in the previous section on multiple fronts.

Firstly, in response to the inherent uncertainty stemming from the array of available questionnaires, we have devised an "ensemble" approach. This ensemble method aggregates the outcomes of various questionnaires, effectively mitigating the errors and inaccuracies inherent in individual scores.

The diverse array of question types, varying scales employed by patients in their responses, and disparities in completion times led to the development of an initial Artificial Intelligence model. This model intelligently weighs the scores derived from individual questionnaires, enhancing the overall accuracy and precision of our stress assessment.

The integration of scores from these questionnaires is just one facet of the innovations our team has introduced. This achievement was made possible through a collaborative effort with a team of specialized psychologists. These experts meticulously analyzed user interviews conducted on our platform, assigning stress levels to each sentence from a corpus of approximately 6,000 sentences collected over six years of web application operation. This panel of psychologists, comprising three seasoned professionals, employed a voting system to collectively determine stress levels for each sentence.

This comprehensive process, aligning interviews, questionnaire results, and psychological stress values, enabled us to extend our work through a technique known in the Data Science field as "distant labeling." Instead of exclusively and manually labeling data via psychologist intervention, we harnessed the questionnaire results to infer stress values, yielding an additional 40,000 sentences. This data generation process, termed "synthetic," circumvented the need for further psychologist involvement in subsequent algorithm development stages. The introduction of synthetic data, generated based on psychologist-supported questionnaire outcomes, represents yet another pioneering contribution from our team.

In addition, we assess the validity of our distant labeling approach by validating a statistical subset of the synthetic dataset with our team of psychological experts. This validation process involved sampling 5% of the synthetic dataset, with the team comprising five distinct domain experts. To ensure robust cross-validation, 10% of the subset was shared among all experts, while the remainder was equally divided among team members. This approach aims to mitigate potential personal biases that might arise from relying solely on the perspective of an individual psychologist.

DETERMINING MODEL PERFORMANCE

Since the data collected for the algorithm's predictive purposes do not enjoy the property of balancing between the classes, a classic algorithm accuracy evaluation approach would lead to an overly optimistic evaluation.

In other words, since the majority of available data deals with cases in which the stress level is medium-low, a traditional Machine Learning algorithm, trained with the aim of minimizing the distance in terms of calculated error between the prediction and the real value, he would limit himself to almost always predicting a medium-low stress level in order to cover almost all of the cases subjected to him.

For this reason we opted for a binary rating system where the low-medium stress value is compared with the high-high stress level based on the ratio between correctly classified and misclassified cases.

In particular, reference is made to a metric called F1 score, which combines the Precision and Recall of a classifier into a single metric by taking their harmonic mean.

F1 is calculated as:

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} \tag{1}$$

Where:

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

TP: Represents cases classified as high-high stress that are actually high-high stress cases.

FP: Represents cases classified as low-medium stress that are actually low-medium stress cases.

FN: Represents cases classified as low-medium stress that are actually high-high stress cases.

The F1 score, as assured by SS, stands at **81.2%** for the English dataset and **83.1%** for the Italian dataset.

CONCLUSION

In conclusion, the evolving discourse on the relationship between natural intelligence and artificial intelligence underscores the remarkable progress made in harnessing computational capabilities, allowing AI to approach and sometimes exceed human intelligence benchmarks. However, this success is often confined to scenarios with well-defined problem structures, leaving the challenge of addressing complex, human-related issues such as mental health.

While stress may not bear the formal classification of a mental disorder, extensive research demonstrates its profound influence on the onset and exacerbation of mental health conditions. Leveraging AI as a supportive tool, particularly through Natural Language Processing (NLP) techniques, provides a promising avenue for proactively identifying and evaluating stress levels in individuals, using written text as a valuable window into their well-being. Moreover, chatbot technology, using AI augmented tools including machine learning and NLP has been introduced into the health sector to address current healthcare challenges, such as shortage of healthcare providers and lack of healthcare access, showing several positive effects in supporting tailored intervention which is better able to address users' needs over a digital treatment (Alpaydin, 2020).

The fusion of domain expertise with machine learning has ushered in an exciting era of innovation in the field of psychology. This interdisciplinary approach enables AI to play a pivotal role in stress identification and offers data science the opportunity to contribute meaningfully to non-STEM domains.

The culmination of a five-year effort, drawing upon diverse data sources from clinical trials and global exploration missions, has illuminated the intricate relationship between word choice and stress levels. Our resulting algorithm, boasting an impressive 81.2% F1 accuracy score, represents a unique and promising interdisciplinary solution. It bridges the gap between AI's computational power and the nuanced understanding of expert psychologists, offering a valuable tool for addressing one of the most pressing challenges of our time: the assessment and management of stress in the modern world.

ACKNOWLEDGMENT

The authors express their gratitude to Fondazione Bruno Kessler (FBK) for their valuable support and contributions.

REFERENCES

Alpaydin E. (2020) Introduction to Machine Learning, fourth edition: MIT Press.

APA, (2018), "Stress," APA Dictionary of Psychology. https://dictionary.apa.org/stress

Carroll, Kathleen M., and Bruce J. Rounsaville. "Computer-assisted therapy in psychiatry: Be brave—it's a new world." Current psychiatry reports 12 (2010): 426–432.

Cockerham, William C. Sociology of mental disorder. Routledge, 2020.

Fitzpatrick, Kathleen Kara, Alison Darcy, and Molly Vierhile. "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial." JMIR mental health 4.2 (2017): e7785.

Fulmer, Russell, et al. "Using psychological artificial intelligence (Tess) to relieve symptoms of depression and anxiety: Randomized controlled trial." JMIR mental health 5.4 (2018): e9782.

Heilig, Markus. "The NPY system in stress, anxiety and depression." Neuropeptides 38.4 (2004): 213–224.

Hyman, Steven E. "The diagnosis of mental disorders: the problem of reification." Annual review of clinical psychology 6 (2010): 155–179.

Kumaraswamy, Narasappa. "Academic stress, anxiety and depression among college students: A brief review." International review of social sciences and humanities 5.1 (2013): 135–143.

Marin M. F. et al., "Chronic stress, cognitive functioning and mental health," Neurobiol. Learn. Mem., vol. 96, no. 4, pp. 583–595, Nov. 2011.

Richards, D. A., Bower, P., Pagel, C. et al. Delivering stepped care: an analysis of implementation in routine practice. Implementation Sci 7, 3 (2012). https://doi.org/10.1186/1748-5908-7-3

Wakefield, Jerome C. "Diagnostic issues and controversies in DSM-5: return of the false positives problem." Annual review of clinical psychology 12 (2016): 105–132.

Widiger, Thomas A., and Douglas B. Samuel. "Diagnostic categories or dimensions? A question for the Diagnostic and statistical manual of mental disorders--." Journal of abnormal psychology 114.4 (2005): 494.