Scope
Volume 13 Number 02 June 2023

Stress Evaluation in Veteran Population Using BRFSS Data with Machine Learning

Radhakrishnan[1], Dr Priyanka Vashistha[2], Dr Ashima Narang[3]

[1,2,3]Department of Computer Science, Amity University,Haryana,India

Abstract

Using data from the Behavioral Risk Factor Surveillance System (BRFSS) and automated learning methods, this research intends to assess veteran stress. Veteran stress might be attributed in part to the specific difficulties many veterans encounter as a result of their military service. In order to create effective therapies as well as assistance networks, it is essential to get an in-depth knowledge of and approach to stress in this group [1]. The BRFSS dataset, compiled by the CDC, is a precious asset for this investigation because of the breadth and depth of the data it contains on a wide range of habits associated with health and disorders. This study analyzes the BRFSS data in order to learn more about the causes of veterans' stress via a variety of statistical methods, including descriptive statistics, inferential statistics, regression analysis, and classification techniques. Demographic characteristics including age, gender, ethnicity, degree of education, and length of military service are used in statistical analysis to provide a picture of the veteran community[2]. In order to get a better knowledge of the factors that may contribute to stress amongst veterans, inductive statistics are used to discover major relationships between parameters and stress levels. Algorithms that forecast stress levels are developed using the technique of regression analysis. Socioeconomic status, lifestyle decisions, health status, and access to medical care are just few of the factors that go into the models of multiple regression that are created. The study's primary objective is to determine what characteristics of veterans' lives are most associated with elevated stress. In order to further forecast stress levels according to multiple indicators, approaches to classification such as logistic regression, decision trees, random forests, and support vector machines are used [3]. These automated learning methods create algorithms that can predict an individual service member's stress level, which might help pinpoint at-risk service members and guide more precise treatment plans. Suitable parameters are used to measure the algorithms' effectiveness terms of their accuracy and dependability. The results of this research add to our knowledge of the causes of veterans' stress and may help in the design of more effective treatments and services for this population. The well-being and quality of life of veterans may be enhanced by determining effective techniques to treat stress in this group. Finally, the thesis suggests how machine learning approaches may be used to the assessment of veterans' stress.

Keywords: Stress evaluation, veteran population, BRFSS data, machine learning, descriptive statistics, inferential statistics, regression analysis, classification algorithms

Introduction

The living standard of an individual is affected significantly by different emotional states, like anxiety and stress [4]. Stress can be described as the complex behavioral and psychological state resulting from any perception of an imbalance between all demands placed upon an individual and their perceived abilities in meeting such demands. Different ML techniques detect stress and identify an individual as unstressed or stressed. To achieve such an objective, several steps are followed, like understanding the format and structure of the dataset, transforming and cleaning data, exploring them, and comparing them. This detection could help monitor stress to prevent dangerous diseases related to stress. Stress detection is normally accomplished by performing a series of actions in sequential order[5]. To begin, it is necessary to understand the structure and format of the dataset that will be utilized for stress detection. This comprises acquiring insights into the variables and their meanings, as well as doing any required preparation activities to ensure that the data are in a format that is suitable for analysis. After a thorough understanding of the dataset has been achieved, data transformation and cleaning methods are done to prepare the data for analysis.

After that, exploratory data analysis, or EDA, is carried out to acquire a more in-depth comprehension of the dataset. This entails conducting statistical studies, looking at the patterns or correlations that emerge from the distribution of variables, determining whether or not there are any significant associations between the variables and the stress levels, and studying the distribution of the variables [6]. EDA is an extremely important component in the process of locating pertinent traits or variables that might lead to the identification of stress.

After examining the dataset, various machine-learning algorithms are utilized to train models for stress detection. Using the labeled data, these algorithms understand patterns and relationships between the features and the target variable (stressed or unstressed). The models are then trained to utilize the dataset, and their performance is tested utilizing proper assessment criteria to guarantee that they can accurately forecast stress levels.

The ability to detect stress through machine learning algorithms has the potential to have important ramifications for monitoring stress and preventing diseases associated with stress. It is possible to adopt suitable interventions and support systems for persons experiencing high levels of stress by first identifying those individuals experiencing high levels of stress. This will assist in alleviating the negative impact that the stress is having on their health and well-being [7]. By identifying and managing stress in its early stages, it is possible to lessen the likelihood of acquiring serious stress-related problems, such as cardiovascular diseases, mental health disorders, and other chronic illnesses.

Literature Review

The literature review will consider different studies on stress evaluation and machine learning. In the past few years, several efforts have been made to automate the detection and prediction of stress using machine learning models that are trained with the use of physiological responses to emotional stimuli and stress [8]. WESAD dataset has been introduced for wearable effect as well as stress detection, along with making this available to all people. This dataset uses wearable devices to include physiological data, such as body temperature, electrodermal activity, respiration, blood volume pulse, and electromyogram activity. The BRFSS is a network of health-related surveys that also receive technical assistance. This BRFSS involves one core standardized questionnaire having a few set questions which might adopt as per the needs. Users of BRFSS data aggregate different state samples from those core questionnaires for use as the database [9]. Understanding every primary cause of stress and how individuals positively or negatively manage all stress is crucial. Stress has been measured traditionally by a few indicative parameters, like pupil diameter, heart rates, and galvanic skin responses.

A questionnaire is another method that helps find an individual prone to stress. In addition, a few life events also help in detecting stress. However, such traditional methods need expensive sensors, continuous assessments, observation, and a need to believe that an individual shares the right mindset answers.

Individuals in this upscaling technological generation try to share their thoughts as well as ideas on platforms of social media always by posting specific content [10]. The correlation between social interactions and the psychological stress states of users has been studied by a hybrid model which investigates the factor graph model using a convolutional neural network (CNN). There is the capability of CNN to learn unified latent features from different modalities [11]. Daily demands in a job also lead to different physiological and psychological strains between people. All individuals experience strain with time, which is dependent on the real and perceived consequences of how each copes with stress [12]. A stress detection system has been designed that uses sleep patterns, social interactions, and physical activities to detect stress.

The study conducted in this work aims to comprehensively investigate various behavioral symptoms of stress by utilizing data from the Behavioral Risk Factor Surveillance System (BRFSS) in combination with machine learning techniques. The primary objective is to analyze the relationship between stress levels and specific factors, namely changes in heart rate and increased body mass index (BMI), to gain a deeper understanding of how these factors contribute to individuals' experience of stress. One notable finding of the study is the direct correlation between changes in heart rate and increased stress levels. Heightened stress often leads to physiological arousal and an elevated heart rate [13]. Monitoring heart rate fluctuations can provide valuable insights into an individual's stress levels and indicate their overall stress response.

Additionally, the study identifies a positive relationship between increased BMI and higher stress levels. This observation suggests that physical health, as indicated by BMI, may influence an individual's susceptibility to stress. Individuals with a higher BMI may experience more stress due to factors such as body image concerns, potential health issues, or societal pressures. Understanding the association between BMI and stress can help inform interventions addressing physical and psychological well-being. The study also highlights the significance of physical activities as potential de-stress agents. Regular engagement in physical exercise has been widely recognized as an effective means of reducing stress levels. Individuals can actively manage and alleviate their stress by incorporating physical activities into daily routines [14]. This finding underscores the importance of promoting and encouraging physical activity as an integral component of stress management strategies.

Machine learning techniques are employed to analyze the BRFSS data and evaluate stress levels within the community. Machine learning offers powerful tools for processing and interpreting large datasets, enabling researchers to identify hidden patterns and relationships that may not be apparent through traditional statistical analyses. By leveraging machine learning algorithms, the study aims to gain deeper insights into the complex dynamics of stress and its impact on individuals' overall well-being. Furthermore, the integration of machine learning techniques in stress research holds the potential to complement psychological and social studies on stress management. By incorporating data-driven approaches, researchers can enhance their understanding of stress by uncovering novel associations and developing more effective interventions and support systems [15]. Machine learning algorithms can identify crucial features and patterns within the data, enabling the development of predictive models and personalized stress management strategies tailored to individual needs.

This work emphasizes the importance of employing machine learning techniques to analyze the behavioral symptoms of stress. Using BRFSS data, researchers can gain valuable insights into the factors influencing stress levels within the community. These findings contribute to the growing body of knowledge on stress management and serve as a foundation for developing evidence-based interventions to promote overall well-being and mitigate stress-related health issues.

Methodology

Stress's negative impacts on public health play a vital role in behavioral disorders like anxiety and depression. There are two tasks of classification in this study based on the emotional states of individuals for detecting stress. All classifiers are to be employed, and there needs to be a comparison of these classifiers. Binary classification and three-class classification are to be used for this study. Binary classification is to classify individuals as either unstressed or stressed [16]. Three-class classification is classifying individuals as normal, stressed, or amused. The BRFSS dataset is the dataset to be utilized for the study. The working group of CDC staff and BRFSS state coordinators have designed the questionnaire for BRFSS.

This questionnaire needs to be approved by every state coordinator prior to utilizing it in the research. This questionnaire has three parts: the core component, state-added questions, and optimal modules. The core component includes the emerging, fixed, and rotating core. Every health department should ask every core component question without modifying wordings; however, every module is optional. This fixed core is the standard question set every state asks during this research, including questions on demographic characteristics [17]. This rotating core comprises two unique question sets, each asked alternatively by every state. This emerging core is the five questions added to these rotating and fixed cores.

Self-reported data acquired via phone conversations are the basis for the BRFSS dataset. Recall that prejudices and social desirability bias are potential problems with self-reported information. The researchers ensured that, since there may also be linguistic obstacles for respondents who speak both English and Spanish since the survey is administered in both languages. Despite these constraints, The BRFSS dataset is valuable for researching population-level health behaviors and illnesses. However, investigators imposed appropriate data cleaning and preparation methods to guarantee data quality and reduce biasness. The researchers are making use of these categorization tasks, in addition to the exploitation of the BRFSS dataset, in an effort to gather insights regarding the prevalence of stress and its impact on emotional states. This research makes an important contribution to the understanding of diseases that are linked to stress, and it gives helpful information that can be used in the development of suitable interventions and support systems for individuals who are currently undergoing stress. An exhaustive clearing procedure is performed on the questionnaire that is used in the Behavioral Risk Factor Surveillance System (BRFSS). before carry out the survey, this method required the participation of each state coordinator. The inquiry is broken up into three distinct portions, each of which is referred to as the basic component, state-added questions, and ideal modules in descending order of importance. Each of these components has a unique purpose, and as a result, the collecting of comprehensive data on public health and risk factors is dependent on each of these components.

After that, the researchers further segmented the core component into the emerging core, the fixed core, and the rotating core. The emerging core consists of five questions that were added to the rotating and fixed cores in order to capture emerging health trends and satisfy unique research objectives. These questions were added to the cores because the rotating core already contained these questions. On the other side, the fixed core is a standardized group of questions that are asked by every state that is a participant in the BRFSS. These questions are asked in order to ensure consistency throughout the states. The BRFSS includes these inquiries in its questionnaire. The information that the respondents offer in response to these questions will not only cover demographic aspects but will also provide essential background information about the respondents. Each state would ask questions taken from either of the two separate sets of questions that make up the rotating core, and they will do so in an alternating pattern. During the survey session, the core is formed using two different sets of questions. This technique ensures that a wide variety of topics and areas of interest are covered over the course of time, so making it possible to form a comprehensive assessment of the current condition of public health. By altering the question sets, the BRFSS is able to collect information on a larger variety of health-related issues, which enables it to keep the length of the questionnaire at a more appropriate level.

Individual states have the option of including questions that are specific to that state in addition to the core components that are required to be included. These questions are specific to each state, and their purpose is to elicit information regarding lifestyle characteristics and health concerns that are particularly relevant to the population of that state. Because of this flexibility, the program may be adapted to meet the needs of the community by taking into account the priorities that have been identified as well as the emerging problems that have been identified. In order to finish the approval procedure for the questionnaire, collaboration between the staff at the CDC and the state coordinators for the BRFSS is required. This ensures that the questions are appropriate, relevant, and reflect the objectives of the data gathering process that is being carried out by the BRFSS [18]. The fundamental component, which is standardized, coupled with the optional modules provide a framework that is both comprehensive and adaptable for the collecting of essential health data on a nationwide scale. This framework is designed to support electronic health records (EHRs). Researchers ensured that the BRFSS uses this format for its questionnaire so that it will be simpler to collect trustworthy data from a variety of different states during the survey. Moreover, this was done so that the BRFSS could better serve its mission and obtain accurate results. Researchers and policymakers now have the chance to gain insightful knowledge on public health trends, risk factors, and the effectiveness of treatments as a result of this. To obtain the accurate results during the survey, the data used were generated using excel.

Additional information regarding some of the questions that were chosen to be included in the survey questionnaire is as follows:

Are there any signs of stress that you can identify in your current state? The purpose of this inquiry is to identify people who are experiencing stress-related symptoms at the moment. It is a preliminary screening question that helps evaluate whether or not there is stress present.

Rate your degree of stress on a scale from one to ten, with one being mild stress and ten representing excessive stress: This question attempts to quantify how stressed-out respondents actually feel about themselves. Researchers are able to gauge the perceived severity of stress thanks to the fact that it provides a subjective estimate of the amount of stress.

Have you ever been evaluated and given a diagnosis for a stress-related condition, such as anxiety or depression? Through the use of this question, we hope to determine whether or not respondents have been provided with a clinical diagnosis of stress-related disorders. It contributes to a better knowledge of the prevalence of illnesses that have been diagnosed and may be associated to stress.

Have you observed any recent variations in the amount of stress that you are experiencing? The purpose of this inquiry is to determine whether respondents have become aware of any changes in their stress levels in the recent past. It offers insights into temporal fluctuations as well as probable stress sources.

How do you normally deal with stressful situations? The purpose of this inquiry is to investigate the respondents' strategies for dealing with stress. It gives light on the ways that individuals adopt to minimize stress and adapt to tough circumstances.

Is there a particular event or set of circumstances that, on a recurring basis, causes you to feel stressed? The purpose of this inquiry is to determine the precise sources of stress that individuals routinely experience. It is helpful in recognizing the sources or conditions that continuously lead to the individuals' lives being filled with stress.

The primary questions, which rotate throughout the survey, are meant to capture various facets of stress and the management of it. For instance, in Set 1, questions center on engaging in physical activity, favored means of unwinding, and consulting with trained professionals for assistance with managing stress. Set 2, on the other hand, the researchers investigated the relationship between stress levels and the impact of work-related tasks, maintaining a healthy work-life balance, and receiving support at work.

The purpose of conducting the survey is to obtain a full picture of stress experiences, triggers, coping strategies, and potential therapies. In order to do this, the survey will combine demographic questions with core and rotating core questions. It is possible to further modify the specific questions in light of the study goals and the population that will be the focus of the survey. The data used in this research generated from excel.

Explanation of the survey

The survey mentioned in the context is conducted to gather information and data related to stress levels and its impact on individuals. The purpose of the survey is to understand the prevalence and factors associated with stress, as well as to assess the effectiveness of stress management strategies.

The survey typically consists of a set of structured questions that cover various aspects related to stress. These questions are designed to capture information about individuals' perceptions, experiences, and behaviors in relation to stress. The survey may include questions about demographic characteristics (e.g., age, gender, occupation), stressors experienced in different domains of life (e.g., work, relationships), coping mechanisms used to manage stress, and the impact of stress on physical and mental well-being.

The survey may utilize a combination of closed-ended and open-ended questions. Closed-ended questions provide respondents with predefined response options, such as multiple-choice questions or Likert scale ratings. These types of questions help in quantifying and categorizing responses for statistical analysis. Open-ended questions, on the other hand, allow respondents to provide detailed written responses, allowing for a more in-depth understanding of their experiences and perspectives.

To ensure the survey's reliability and validity, the questionnaire is typically developed by a team of experts or researchers with knowledge and experience in stress research. It may also undergo a pilot testing phase to evaluate its clarity, comprehensibility, and relevance to the target population. This helps in refining the survey instrument and ensuring that it effectively captures the intended information.

The survey administration can be done through various methods, including online surveys, face-to-face interviews, or paper-based questionnaires. The selection of the administration method depends on the target population and the resources available for data collection. Researchers may aim for a representative sample of the population to ensure the survey results can be generalized to a larger population. Once the survey data is collected, it undergoes a process of data cleaning and analysis [19]. This involves checking for errors, inconsistencies, and missing data, and performing statistical analyses to examine relationships, patterns, and associations between variables related to stress.

Analysis

Veterans are described using statistical information that break them down by age, gender, race/ethnicity, education level, and length of military service. Means, frequencies, and percentages were be calculated during the survey in order to offer a thorough summary of the data collected. It may show, for instance, how many people in a certain age range are veterans, what percentage of the population is made up of men and women, how many people of varying ethnic backgrounds live there, and how many people have completed college.

Significant connections between factors and veterans' stress levels will be identified using inferential statistics. Stress and other variables including socioeconomic status, lifestyle decisions, health status, and proximity to medicine may be evaluated using hypothesis testing like chi-square analyses or t-tests. By estimating confidence intervals, we may learn more about the strength of these correlations and the accuracy of our estimations. Factors strongly related with stress in the veteran population may be determined using the findings of inferential statistics.

Models that estimate veterans' stress levels will be developed with the use of regression analysis. To determine the relative importance of various factors in contributing to stress, multiple regression models will be created. Regression models may, for example, include in factors like income, employment, physical exercise, and mental health status. The results of the research will show which variables have a meaningful impact on stress levels, allowing us to zero in on the most important variables.

Segmentation Algorithms: In this research investigators used techniques from machine learning, including logistic regression, decision trees, random forests, and assistance neural networks to estimate veterans' stress levels on the basis of a number of different variables. The found indicators will be used by these categorization techniques to create models that can assess veterans' stress levels. Assessment standards such as accuracy, precision, recall, and F1 score will be used to evaluate the models' efficacy. This study's findings will shed light on whether or not artificial intelligence techniques are useful for forecasting veterans' stress levels.
In general, we evaluated veteran stress levels using the BRFSS information by combining descriptive and inferential statistics, as well as regression analysis and classification methods. Understanding the causes of veterans' stress, zeroing in on the most important indicators, and creating accurate prediction models are all outcomes of this research [20]. These results may aid in the design of more effective programs and services for veterans to reduce stress and enhance their quality of life.

Conclusion

In the work, a study must be carried out on every behavioral symptom of stress using BRFSS data with machine learning. Changes in heart rate, along with increased BMI, increase individuals' stress levels. Physical activities act as the de-stress agent upon human stress. By increasing physical activities daily, individuals could reduce stress levels. Machine learning techniques could be an efficient approach for analyzing BRFSS data for evaluating stress levels within the community and possessing the potential to provide valuable supplements to psychological and social studies of stress and stress management.

References

1. Virnig, B. A., &McBean, M. (2001). Administrative data for public health surveillance and planning. *Annual review of public health*, *22*(1), 213-230.
2. Dai, H., & Richter, K. P. (2019). A national survey of marijuana uses among US adults with medical conditions, 2016-2017. *JAMA network open*, *2*(9), e1911936-e1911936.
3. Colosimo, M. M. (2020). *The Chronic Illness Risk and Resiliency Profile Project: Utilizing Machine Learning to Build a Multivariate Correlational Model of Disease* (Doctoral dissertation, Drexel University).
4. Betancourt, J. A., Stigler Granados, P., Pacheco, G. J., Shanmugam, R., Kruse, C. S., & Fulton, L. V. (2020, June). Obesity and morbidity risk in the US veteran. In *Healthcare* (Vol. 8, No. 3, p. 191).MDPI.
5. St. Hill, C. A., Swanoski, M. T., Lipsky, M. S., &Lutfiyya, M. N. (2017, July). A population-based, cross-sectional study examining health services deficits of US veterans using 2014 behavioral risk factor surveillance system data: Is rural residency an independent risk factor after controlling for multiple covariates? In *Healthcare* (Vol. 5, No. 3, p. 39).MDPI.
6. Galloway, R. D. (2003). Health promotion: causes, beliefs, and measurements. *Clinical medicine & research*, *1*(3), 249–258.

7. Wennberg, D. E., Kellett, M. A., Dickens, J. D., Malenka, D. J., Keilson, L. M., & Keller, R. B. (1996). The association between local diagnostic testing intensity and invasive cardiac procedures. *JAmA*, *275*(15), 1161-1164.
8. Arora, A. (2022). *Association Between Arthritis and Yoga Among Veterans* (Doctoral dissertation, University of Nevada, Las Vegas).
9. Xie, Z., Nikolayeva, O., Luo, J., & Li, D. (2019). Peer reviewed: Building risk prediction models for type 2 diabetes using machine learning techniques. *Preventing chronic disease*, *16*.
10. Green, M. (2021). Evaluating the performance of personal, social, health-related, biomarker and genetic data for predicting an individual's future health using machine learning: A longitudinal analysis. *arXiv preprint arXiv:2104.12516*.
11. Navarro, C. L. A., Damen, J. A., Takada, T., Nijman, S. W., Dhiman, P., Ma, J., ...&Hooft, L. (2021). Risk of bias in studies on prediction models developed using supervised machine learning techniques: a systematic review. BMJ, *375*.
12. Fregoso-Aparicio, L., Noguez, J., Montesinos, L., &García-García, J. A. (2021). Machine learning and deep learning predictive models for type 2 diabetes: a systematic review. *Diabetology& Metabolic Syndrome*, *13*(1), 1-22.
13. Betancourt, J. A., Stigler Granados, P., Pacheco, G. J., Shanmugam, R., Kruse, C. S., & Fulton, L. V. (2020, June). Obesity and morbidity risk in the US veteran. In *Healthcare* (Vol. 8, No. 3, p. 191).MDPI.
14. Findley, P. A., Wiener, R. C., Mitra, S., Wang, H., Shen, C., &Sambamoorthi, U. (2023). Whole Health in Parts: Omissions from National Data Sets. *Population Health Management*, *26*(1), 22-28.
15. Cooper, C. (2017). *The Influence of Cardiovascular Disease Risk Factors on Exercise Participation* (Doctoral dissertation, Walden University).
16. Raifman, J., Moscoe, E., Austin, S. B., Hatzenbuehler, M. L., &Galea, S. (2018). Association of state laws permitting denial of services to same-sex couples with mental distress in sexual minority adults: A difference-in-difference-in-differences analysis. *JAMA psychiatry*, *75*(7), 671-677.
17. Wang, M., Qiu, J., & Ong, P. (2018, November). Machine learning in spatial study of asthma rate distribution in losangeles county. In *2018 Chinese Automation Congress (CAC)* (pp. 3696-3700).IEEE.
18. Hodges, C. B. (2020). *Heterogenity in Brain Injury: An Investigation of the Efficacy of Qualitative Comparative Analysis in Diffusion Tensor Imaging* (Doctoral dissertation, Brigham Young University).
19. Obisesan, O. H., Mirbolouk, M., Osei, A. D., Orimoloye, O. A., Uddin, S. I., Dzaye, O., ... &Blaha, M. J. (2019). Association between e-cigarette use and depression in the behavioral risk factor surveillance system, 2016-2017. *JAMA network open*, *2*(12), e1916800-e1916800.
20. Dean, J. P. (2014). *Emotional intelligence, stress, and cardiovascular disease in law enforcement officers* (Doctoral dissertation, Walden University).