Depression and Anxiety Detection with Machine Learning Techniques using Data from Wearable Sensors

Mani Barathi SP S[1], Rekha V S[1], Dr.M. Sujithra[2], Dr.K. Rajarajeshwari[2]

[1] *M.Sc. Data Science, Department of Computing- Data Science, Coimbatore Institute of Technology, Coimbatore, Tamil Nādu, India*
[2] *Assistant Professor, Department of Computing - Data Science, Coimbatore Institute of Technology, Coimbatore*

Abstract:

stress is a physical, mental, or emotional factor that causes bodily or mental tension. Stresses can be external (from the environment, psychological, or social situations) or internal (illness, or from a medical procedure). This paper provides an effective method for the detection of cognitive stress levels using data retrieved from the wearable device. The main motive of this system was to use a machine learning approach in stress detection. Individually, the effect of each stressor was evaluated using logistic regression and then a combined model was built and assessed using variants of ordinal logistic regression models including logit, probit, and complementary log-log. This system was used and evaluated in a real-time environment by collecting a data from adults working in IT and other sectors in India. The novelty of this work lies in the fact that a stress detection system should be as non- invasive as possible for the user.

keywords--cognitive, detection, real-time, non-invasive.

INTRODUCTION

Stress has become an embedded part of our daily life and is a noticeable concept in public health. Recently, stress has become an integral part of professional life, especially in today's fiercely competitive economy. In the workplace, an individual has to continuously face several situations, such as work overload, job insecurity, lack of job satisfaction, and the pressure to stay up-to-date. The continuous presence of stress can lead to several negative health effects, such as high blood pressure, lack of sleep, susceptibility to infections, and cardiovascular disease. All these situations result in mental stress, which has become the leading cause of many diseases. Such adverse effects not only affect the employees' health and well-being, but also affects workplace productivity and overall profit. Conventionally, psychological and physiological specialists decide stress condition of an individual using questionnaire-based stress analysis. This approach carries lot of uncertainty and is unreliable as it depends entirely on the individuals' responses and the people will be timorous to answer the questionnaire. The objective of the proposed work is to automatically detect the stress condition of an individual by using the physiological data recorded during the stressful situations. Such a detection can help in monitoring stress to prevent dangerous stress-related diseases.

Fig. 1 shows the conceptual model of our research work on stress. The significance of sleep, physical activity, number of working hours and change in heart rate regarding stress levels are analyzed.

Fig 1: The conceptual model of stress

Many researchers are working on investigating stress levels among different professionals. We are currently working on studying the stress levels (low, medium, high) among professionals using the data collected from people.

HYPOTHESIS:

H1: Is there any significant relationship between their health issues and stress levels?

H2: Is there any significant relationship between their family issues and stress levels?

H3: Is there any significant relationship between number of working hours and their stress levels?

H4: Is there any significant relationship between any changes in resting hour and their stress levels?

LITERATURE SURVEY

Stress detection using natural language processing

Tanya Nijhawan proposed a paper on stress analysis using natural language processing. They extend sentiment and emotion analysis for detecting the stress of an individual based on the posts and comments shared by him/her on social networking platforms. They leverage large-scale datasets with tweets to accomplish sentiment analysis with the aid of machine learning algorithms and a deep learning model, BERT for sentiment classification. Here they also adopted Latent Dirichlet Allocation which is an unsupervised machine learning method for scanning a group of documents, recognizing the word and phrase patterns within them, and gathering word groups and alike expressions that most precisely illustrate a set of documents. They help them to predict which topic is linked to the textual data. With the aid of these models, they will be able to detect the emotion of users online. Further, these emotions can be used to analyze stress or depression. In conclusion, the ML models and a BERT model have a very good detection rate. This research is useful for the well-being of one's mental health. The results are evaluated using various metrics at the macro and micro levels and indicate that the trained model detects the status of emotions based on social interactions.

Mental Stress Detection Using Wearable Sensors

A comprehensive review has been presented, which focuses on stress detection using wearable sensors and applied machine learning techniques. This paper investigates the stress detection approaches adopted in accordance with the sensory devices such as wearable sensors, Electrocardiogram (ECG), Electroencephalography (EEG), and Photoplethysmography (PPG), and depending on various environments like during driving, studying, and working. The stressors, techniques, results, advantages, limitations, and issues for each study are highlighted and expected to provide a path for future research studies. Also, a multimodal stress detection system using a wearable sensor-based deep learning technique has been proposed at the end.

Stress Detection with Machine Learning and Deep Learning using Multimodal Physiological Data

This paper proposes different machine learning and deep learning techniques for stress detection on individuals using multimodal dataset recorded from wearable physiological and motion sensors, which can prevent a person from various stress related health problems. Data of sensor modalities like three axis acceleration (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG) and electrodermal activity (EDA) are for three physiological conditions - amusement, neutral and stress states, are taken from WESAD dataset. The accuracies for three-class (amusement vs. baseline vs. stress) and binary (stress vs. non-stress) classifications were evaluated and compared by using machine learning techniques like K-Nearest Neighbor, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost and Kernel Support Vector Machine. Besides, simple feed forward deep learning artificial neural network is introduced for these three-class and binary classifications. During the study, by using machine learning techniques, accuracies of up to 81.65% and 93.20% are achieved for three-class and binary classification problems respectively, and by using deep learning, the achieved accuracy is up to 84.32% and 95.21% respectively.

METHODOLOGY

The steps followed in the stress detection process, which is divided into various steps including data collection, data pre-processing, feature selection, hypothesis building, stress detection model, hypothesis testing, and interpretation of results.

Data collection

The methods used in this study to measure stress are the data collected from the various people and an online perceived stress scale questionnaire (Form-based). In this study, 12 users working in IT and other fields participated in the experiment for more than 12 months in the period between February 2019 and the present day. The ages of the individuals vary from 25 to 45 years. The participants were equipped with a FITBIT watch and the app was installed on their smartphone. The data gathering process from individuals was carried out on a monthly basis. The following are the data details:

1. Body related information – Date, weight, BMI, fat; 2. Physical activity information – Date, calories burned, steps, distance, floors, minutes sedentary, minutes lightly active, minutes fairly active, minutes very active, activity calories; 3. Sleep information – Date, minutes asleep, minutes awake, number of awakenings, time in bed; and 4. Heart rate information during physical activity – Date, calories, average_bpm, maximum and minimum heart rate, active minutes, zone, resting heart rate, type_of_activity, heart rate zone. Table II shows the data statistics for the stress detection system.

THE DATA STATISTIC FOR THE STRESS DETECTION SYSTEM.

Number of users 12, Number of days 300, Number of attributes 29, Number of records 3000.

The perceived stress scale (PSS) is the most preferred psychological instrument used to measure stress. The scale includes a questionnaire, which consists of two parts: A demographic questionnaire and a stress questionnaire. The PSS consists of 10 questions in several categories and the scores are obtained by reversing the responses (e.g., 0 = 4, 1 = 3, 2 = 2, 3 = 1 and 4 = 2). These scores indicate the level of stress with a score of 27–40 indicating very high stress, 14–26 indicating moderate stress and 0–13 indicating low stress. Fig. 3 shows the online PSS questionnaire form used to gather data from the participants.

In addition to the FITBIT data and PSS score, we have used various demographic attributes including name, age, gender, and working hours as well as the maximum heart rate and heart rate level of the participants. The maximum heart rate shows the change in the heart rate of an individual and is computed as follows:

Maximum heart rate = 217 - (0.85 * valid$Age)

The heart rate level indicates how closely our computed maximum heart_rate is with the recorded maximum heart rate (in percentage) and is computed as follows:

Heart rate level = (Recorded maximum heart rate / computed maximum heart rate) * 100

FEATURE SELECTION

The feature selection aims to choose a set of highly discriminant attributes i.e., those that are capable of discriminating the samples that belong to a completely different class from the data set. A feature fi belongs to a class Cj if fi and Cj are highly correlated. In this paper, the first step involved a multicollinearity checking between the attributes, where variable clustering was carried out using the correlation metric and the variables with high correlation clustered together. In the second step, agglomerative hierarchical clustering was applied on the data set, which builds a cluster hierarchy using a tree diagram called a dendrogram. This initially builds each object in a separate cluster and then at each step, the two clusters that are most similar are joined into a single new cluster. In the third step, oblique principal component analysis for feature selection was carried out, which is superior to the orthogonal results given by PCA and produces cleaner and easily interpreted results based on the cluster distances. In the final step, variable importance plots were built to identify the prominent features required for model building.

MODELING PROCESS

This paper aims to detect the stress levels of an individual. Normally, ordinal logistic regression models are used when the dependent variable has more than two categories and can have nominal and/or continuous independent variables. In the data set, the response variable stress_level is categorical and ordered (Low < Medium < High) in nature, so we attempted to fit the cumulative probabilities of the response variable with the cumulative logit model, probit model, and complementary log-log model in our study.

HYPOTHESIS BUILDING

To assess the impact of physical activity, sleep patterns, working hours, and heart rate on an individual’s stress levels, the following hypotheses were formulated. H: Hypothesis, H0: Null Hypothesis

H1: Health issues is considered as a strategy for managing stress.

H0: Health issues is not considered as a stress management strategy.

H2: No of resting hour is an indicator of human stress.

H0: No of resting hour doesn't lead to human stress.
H3: Long working hours have a negative impact on human stress.
H0: Long working hours doesn't have any impact on human stress.
H4: Family issues is an indicator of human stress.
H0: Family issues has no role in defining human stress.

A. Logistic Model to Support Hypothesis 1

Physical activity works as a de-stress agent for an individual. Regular indulgence in physical activity such as walking, climbing stairs, and minutes active will help to reduce stress levels. To test this hypothesis, a logistic model was built using various the physical activity related attributes including calories burned, steps, floors, distance, minutes sedentary, minutes lightly active, minutes very active, and minutes fairly active. The final model based on an ANOVA test shows the number of floors climbed, minutes sedentary, minutes fairly active, and minutes very active are influential factors in determining the stress level of an individual. In addition, the multiplicative term minutes sedentary minutes very active reflects the conditional relationships and can influence the stress levels of an individual.

B. Logistic Model to Support Hypothesis 2

NO of resting hour are also indicators of an individual's health and stress levels. Usually, people who have low stress will be able to get good sleep. To test this hypothesis with the data set, a logistic model was applied to the sleep data set and the coefficients were computed. In the data set, various attributes including minutes asleep, minutes awake, time in bed and number of awakenings were related to sleep and chosen for model building to check their influence on the stress levels. The final model based on an ANOVA test shows the minutes awake was the only influential factor in determining the stress levels in an individual.

C. Logistic Model to Support Hypothesis 3

A competitive work environment contributes to maximum stress. Jobs that demand high working hours are those that cause more stress in an individual. Long working hours are considered as a good indicator for determining an individual's stress levels. In the data set, the working hours of the participants have been recorded. The final model shows an individual's working hours are a significant influential factor (based on p-value) in determining their stress levels.

D. Logistic Model to Support Hypothesis 4

To measure the variability of the health issues, no medical equipment was used to capture the ECG. Instead, it was modeled using the following formula:

Maximum heart rate = 217 - (0.85 * valid$Age)

The maximum heart rate showcases the fluctuation in the heart rate on an individual. The closest the individual achieves during physical activity, the better it helps in reducing their stress levels. To test this hypothesis, a logistic model was built using various heart rate related attributes including the average bpm, maximum, minimum, and resting heart rate, activity (WALK, SPORTS, RUN) and heart rate zones (Fatburn, Cardio, Peak). The final model shows the average bpm, maximum and minimum heart rate are positive indicators of stress and calories, resting heart rate and heart rate zone are a negative indicator (de-stress agents) of stress. In addition, the activity type (RUN, SPORT, WALK) doesn't contribute to stress.

E. Combined Model

The basic form of the multiple regression model is shown below:

$Y = \alpha + \beta1\ X1 + \beta2\ X2 + \beta3\ X3 + \beta4\ X4 + \beta5\ X5 + \ldots\ldots + \varepsilon$

Table III shows the list of attributes used for stress detection.

1) Logit model

The formula used for the ordinal logistic regression is shown below:

g(p) = log (p / (1 - p) The initial model was built with the following attributes and then an odds ratio was computed for the model.

Stress level (Y) = $\alpha + \beta1\ X1 + \beta2\ X2 + \beta3\ X3 + \beta4\ X4 + \beta5\ X5 + \beta6\ X6 + \beta7\ X7 + \beta8\ X8 + \beta9\ X9 + \beta10\ X10 + \beta11\ X11 + \varepsilon$

a) For each unit increase in the BMI, the odds of moving from low stress to medium stress or from medium stress to high stress are 1.023.

b) In the same way, if the person increases one floor more to his daily activity, the odds of moving from low stress to medium stress or from medium stress to high stress decreases, and was in fact reduced by a factor of 0.975 [Physical activity is a de-stress agent].

c) For each one hour increase in working hours, the odds of moving from low stress to medium stress or from medium stress to high stress decreases, and was in fact reduced by a factor of 0.915. This depends on how an individual handles his/her workload.

2) Probit Model

The formula used for the probit model is shown below:

g(p) = ϕ -1 (p)

The odds ratio was computed for the probit model and was combined with the confidence intervals.

a) For each fluctuation in the average_bpm, the odds of moving from low stress to medium stress or from medium stress to high stress are 1.01.

b) In the same way, if the person increases the amount of cardio in their daily activity, the odds of moving from low stress to medium stress or from medium stress to high stress decreases, and is in fact reduced by factor 0.951.

c)

3) Complementary Log-Log model:

The formula for the complementary log-log model is shown below:

g(p) = log(-log(1-p))

The odds ratio was computed for the complementary log-log model and it was combined with the confidence intervals.

a) For each unit increase in working_hours, the odds of moving from low stress to medium stress or from medium stress to high stress are 1.061.

b) In the same way, if the person increases their BMI, the odds of moving from low stress to medium stress or from medium stress to high stress increases by a factor 1.025.

II. RESULTS

The Akaike information criterion (AIC) is used to quantify the relative quality of logistic models for a given data set. Given a set of models for the data, the AIC estimates the quality of each model, relative to each of the other models. The lower the AIC value, the better the model. Our study clearly indicates that Probit was the most suitable model for the data set and our final model based on an ANOVA test. Fig. 8 shows the efficiency of the models based on the AIC value. Final Model:

Stress.Level ~ X1 + X2 + X3 + X4 + X5+ X6 +X7+ X8 + X9 + X10 + X11 + X2:X9 + X1:X5 + X9:X11 + X3:X5 + X1:X2 + X1:X11 + X1:X7 + X4:X5 + X3:X10 + X4:X9

Conclusion:

A medium percentage of the people did have high stress. Person facing stress at the middle age level leads to lot of psychological problems in the form of decreased motivation, absenteeism for class and examinations, incompletion of all works etc. The stress management is a leading fact that each and every management should concentrate so that they can keep an eye on their academic and personal life. All the people regardless of his / her age, gender, income level or any other priority should be treated equally and should manage without any dissatisfaction is necessary. Academic factors were one of the most important stressors. The introduction of stress management education into the curriculum could prove useful in combating this problem. person themselves should become trainers of managing stress. This trend will lead to empower the people and to get succeed in their academic and personal life. person facing stress are advised to attend stress management courses which will help them to build coping strategies and cause out their stress. The stress management cause comprises of a package program consisting of

- **Tranquility**
- **Positive outlook towards works**
- **Self-analysis through personality type test**
- **Inter personal skill development**
- **yoga cum meditation**
- **Time management.**

Effective communication between person and the parents, faculties should be promoted. This could help people to find the appropriate stress reduction methods and to improve their academic performance.

REFERENCES

[1] Agrawal, R.K. and Chahar, S.S. (2007), “Examining role stress among technical students in India”,Social Psychology of Education, Vol. 10 No. 1, pp. 77–91.

[2] D'Zurilla, T. J., and Sheedy, C. F. (1991), “The relation between social problem-solving ability and subsequent level of psychological stress in college students”, Journal of Personality and Social Psychology, Vol 61, pp. 841-846.

[3] Hellhammer, H. (2007), “Solutions on Stress (SOS): programmes, packages and products for helping Teenagers”, Young Consumers, Vol. 8 No. 1, pp. 29-35.

[4] Keinan,G. & Perlberg, A.(1986). Sources of stress in Academe: The Israeli Case. Higher Eduv´cation, Vol.15,No.1/2,73-88.

[5] Moore,K.A, Burrows, G.D. & Dalziel, J.(1992). Stress: How to define and Challenge it. Mental Health in Australia, 32-40.

[6] Radcliff, C. & Lester,H.(2003). Undergraduate medical Education. Perceived Stress during undergraduate medical training: A qualitative study. Medical Education 37(1), 32-38.

[7] Rakesh Kumar Agrawal & Shailendra Singh Chahar , Examining role stress among technical students in IndiaSocial Psychology of Education March 2007, Volume 10, Issue 1, pp 77–91.

[8] Ross, S.E., Niebling, B.C. and Heckert, T.M. (1999), “Sources of Stress among College Students”, College Student Journal, Vol. 33.