Stress Detection in College Students Using Machine Learning Algorithm

Ms. Ancy Paul[1], Ms. Resija P R[2]

[1]Department of Computer Science Vimala College (Autonomous), Thrissur, Kerala ancypaul9999@gmail.com
[2]Assistant Professor, Department of Computer Science Vimala College (Autonomous), Thrissur, Kerala resijapr1995@gmail.com

ABSTRACT—

Mental stress is a major issue nowadays, especially among youngsters. The age that was considered once most carefree is now under a large amount of stress. Today's increased stress causes a variety of issues, including depression, suicide, heart attacks, and stroke. Our goal is to examine stress in college students at various stages of their life. Some of the factors that effect on the students which often goes unnoticed. We will perform an analysis on how these factors affect the mind of a student and will also correlate this stress with the time spent on the internet. In this model the main goal is to use machine learning algorithms to estimate the levels of stress. Data is collected from Vimala College (Autonomous), Thrissur students through surveys in online mode and it consisted of 954 student's data. Our model is a classification type in supervised ML. There are 3 classes a)chronic b) episodic c)acute. Our objective is to detect the different level of stress in students. We were asked basic questions about their feelings in situations and analyze their answers using Machine Learning techniques and make predictions.

Index Terms—Machine Learning, SVM, Random forest, Naive bayes, Logistic regression, Randomforest, KNN, Adaboost, Deci- sion tree, SGD, Linear regression

I. INTRODUCTION

The current educational system and the tough competition inclines the anxiety and stress amongst students. Other factors which contribute towards the mental disparity amongst stu- dents include parental pressure, peer pressure, health issues, financial conditions. [1]The additive has been the pandemic of the corona virus, dysfunctioning the normalcy of student's lives and suscepting them to more pressure thus leading to ill performance.

The automation for student stress prediction in institutes and educational organizations has been very minimal. Observing each student and his or her profile is a huge task. This responsibility lies under human interaction and that is why our work paves way for the automatic stress prediction of each student succumbing under various parameters and proposes the solution to each student rightly. This is done with the help of Machine learning and data science techniques. Keeping a check of each students stress levels, and monitoring it closely, helps to heighten their performance in an organization.

Stress is the body's reaction to pressure from a particular situation or event. It can be any- physical, mental, or emotional reaction. Job, family illness, or money troubles are some of the common triggers. [2] When a human experiences stress, it develops a physical and mental response; this is because the body is designed to experience and react to it. Any stress responses assist the body in a new environment. It can be positive by keeping us alert, motivated and ready to avoid danger. But it's important to know that stress becomes an issue when stressors continue without relief or periods of relaxation.

Acute Stress: Your body's response to a novel or difficult environment causes acute stress. It's that sensation you ex- perience when a deadline is drawing near or when you just avoid being hit by a car. We may even encounter it as a result of an enjoyable activity. like a thrilling roller coaster ride or a remarkable personal accomplishment.Short-term stress is categorised as acute stress. Emotions and the body typically return to normal after a short period of time.

Episodic Acute Stress: Continual acute pressures are re- ferred to as episodic acute stresses. The cause of this can be persistently tight job deadlines. It might also be a result of the regular high-stress situations that some professions, including those involving healthcare, experience. [3] When we are under this kind of stress, we don't have time to return to a calm and relaxed state. Furthermore, the effects of frequent, acute stresses accumulate. As a result, we usually feel as though we are experiencing one crisis after another.

Chronic Stress: Stressors that last for an extended amount of time lead to chronic stress. Living in a neighbourhood with a high crime rate is one example, as is frequently quarrelling with your life partner. Stress like this seems to never stop. We frequently struggle to see any way to alter or better the circumstance that is the source of our ongoing worry.

The objective of this project is to detect the level of stress in students. This model evaluate and analyze the level of stress and find a solution according to the students opinion. Use various Machine Learning algorithms to build detection mod- els, evaluate the accuracy and performance of these models. Finding out the best model and providing the best accurate result.

II. LITERATURE SURVEY

The existing works in the area have been discussed here.

A. Mental Stress Detection Using Machine Learning Algo- rithm

[4] Stress is a subjective phenomenon that is difficult to measure comprehensively. However, we can classify and quantify stress and how it affects one's personal health, including various biological and psychological vulnerabilities. If we only hear what an individual says and ignore what the face of that person is telling us, then we just have half the story. This paper represents appearance based facial expres- sion recognition system using Convolutional Neural Network (CNN). The Local Binary Pattern (LBP) was used to extract the appearance features. The CNN is trained to categorize into four basic facial expressions (anger, fear, unhappy and non-stressed expressions). This system was experimented with Indian and Cohn-Kanade database.

B. Stress Detection With Machine Learning And Deep Learn- ing Using Multimodal Physiological Data

[5] This paper proposes different machine learning and deep learning techniques for stress detection on individuals using multimodal dataset recorded from wearable physiolog- ical and motion sensors, which can prevent a person from various stress-related health problems. Data of sensor modal- ities like three-axis acceleration (ACC), electrocardiogram (ECG), blood volume pulse (BVP), body temperature (TEMP), respiration (RESP), electromyogram (EMG) and electroder- mal 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 Neighbour, Linear Discriminant Analysis, Random Forest, Decision Tree, AdaBoost and Ker- nel Support Vector Machine. Besides, a 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 percentage and 93.2 percentage are achieved for three-class and binary classification problems respectively, and by using deep learning, the achieved accuracy is up to 84.3 percentage and 95.21 percentage respectively.

C. Stress Detection From Sensor Data Using Machine Learn- ing Algorithms

[6]The paper's main goal is to use machine learning algorithms to estimate the levels of stress that can be detected by grouping together many measurements such as pulse rate, body temperature, heart rate, and systolic blood oxygen satu- ration (spo2). When a person is under stress, their bio-signals, such as thermally, electrical, impedance, acoustic, and optical, change noticeably. These bio-signals can be used to measure stress levels. Accelerometer, body temperature, respiration, blood volume pulse electrocardiogram, (BVP), [7]electro der- mal activity, and other sensor modalities (EDA). Machine Learning Classification methods such as Kernel Support Vector Machine, K-Nearest Neighbour, AdaBoost, Random Forest

NO	TITLE	AUTHOUR	METHOD	ACCURACY
1	Mental stress detection using machine learning algorithm	Essra hassan	CNN (emotions like happy,sad,fear,a ngry)	85%
2	Stress detection with ML and deep learning using psychological model data	Ryan Haider	KNN (wearable sensors)	81%
3	Stress detection from Sensor data using ML.	Fajardo	KNN	93%
4	Stress prediction and Detection in IOT	Jaycel Garlit	KNN	91%
5	Stress detection of human wearing mask	Mai Hazem	KNN	90%

Fig. 1. comparison table

and Decision Tree methods used to evaluate and compare the classifications. The Random Forest model beat the other approaches with F1-scores of 93.77 and 70.03 for classification model and three-class classification, respectively.

D. Stress Prediction And Detection In Internet Of Things Using Learning Methods

[8] The article focuses on the impact of stress detection of a person with analysis of Internet of Things (IoT) sensed data along with machine learning algorithms. The physiological data can be detected by performing the stressor tests on the person, and they include arithmetic calculations. Some of those algorithms are Support Vector Machine (SVM), Logistic- regression, K-nearest neighbor, adaptive neuro-fuzzy inference system (ANFIS) decision-tree

and so on. By analyzing the re- sults from the literature, machine learning algorithms provide high detection accuracy of 95 percentage with 90 percentage F1 score and reduced prediction errors.

E. Detection Of Stress In Humans Wearing Face Masks Using Machine Learning And Image Processing

[9] The primary purpose of this work is to employ vivid Image Processing and Machine Learning Techniques to recog- nise stress in the human body. Our innovative system is an improved prior iteration system that includes live detection, pe- riodic employees analysis, detecting physical and mental stress level and providing appropriate stress management remedies via a survey form. Our strategy focuses on stress management and providing employees with a healthy and spontaneous work environment so that they may get the most out of their time at work.

III. PROPOSED SYSTEM

Stress detection, in this context is framed as a multivariate analysis. There are three possible output labels: chronic (high), episodic ((medium) and acute (low). So using nine methods
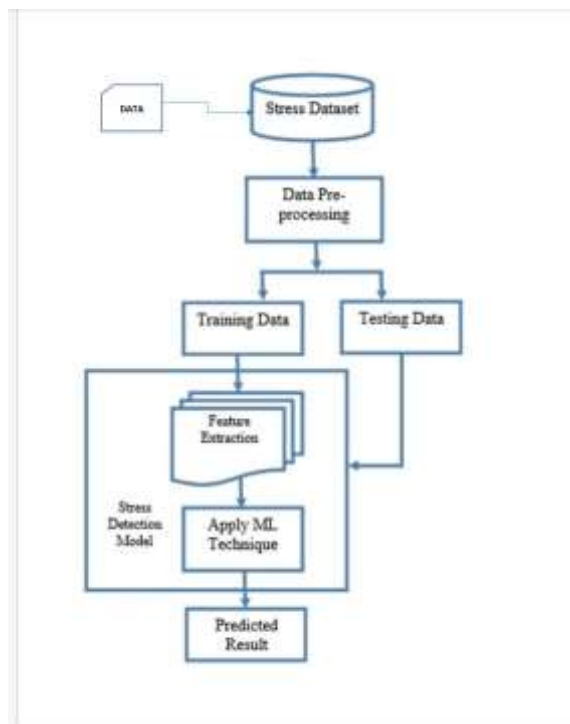

Fig. 2. System Architecture

to detect the stress. Comparing these nine algorithms to take highly accurately detected algorithm out of them. The system architecture of our model is shown in figure 2.

A. Data Collection

The most important thing in my project is the dataset. Data for my project is collected by myself from the students in Vimala College(Autonomous), Thrissur. For data collection I use google form. [10]With Google Forms, We can create surveys or quizzes right in your mobile or web browser with no special software required. We get instant results as they come in. And, you can summarize results at a glance with charts and graphs. Before that, I want to prepare my questionaire for my google form. I study stress with the help of the psychology department of Vimala College (Autonomous), Thrissur for further progress. And after discussion I made my questionnaire for the google form. Real time dataset , Collected by google form from the students. 35 columns are included , Out of

35 columns 33 are questions for the students . Questions are in multiple choice , short answer, paragraphs, checkbox, Ranking. Totally 954 Responses are collected.

B. Data Pre-processing

It is a process of detecting and correcting (or removing) inaccurate entries from the data set. It involves duplicate entries, missing values, fix types, spell check, lower case and upper-case letters, punctuation, formatting, blank spaces, removing null values etc.

We know that ML algorithms only work in numeric values. Our dataset is a combination of both strings and numeric

Fig. 3. Excel Data From Google Form

Fig. 4. After Vectorization

values. So we first need to convert our string format dataset to numeric format using some vectorization techniques in text an- alytics. Text Vectorization is the process of converting text into numerical representation. [11]Here are some popular methods to accomplish text vectorization: Binary Term Frequency. Bag of Words (BoW) Term Frequency. (L1) Normalized Term Frequency. we use word to vectorization techniques for the vectorization of our dataset.

For example we take question from the questionnaire : How many hours do you spend in a day to do work given from college?

1) Less than one hour

2) 1-2 hours

3) 3-4 hours

4) 4-5 hours

5) More than 5 hours Actually answer of this question was in string format .zWe need to convert this in to vector. So we can arrange it to Option 1, option 2—Less stress(Acute): Represented as 0 in dataset Option 3,option 4—Medium stress(Episodic):Represented as 1 in dataset Option 5—Over stress(Chronic):Represented as 2 in dataset. In this manner I convert all my remaining questions according to its options. After vectorization the dataset is shown in figure 4.

C. Feature Extraction

It includes only the required amount of data fields for the detection process. This dataset had some unwanted Columns like Name,Date,college name, gender. And take all other columns as features for this detection process. Next we want to calculate the amount of stress in each student data and correlation between each dependent column to the independent column. After we get: 10-40 : Acute , 40-70: Episodic , 70 above: chronic. Then I find the stress level in each student by taking this criteria. And make a column called STRESS

Fig. 5. Final Dataset After Feature Extraction

MODEL	ACCURACY
Logistic regression	99%
Histgradient boosting	95%
Naive bayes	94%
Random forest	99%
KNN	96%
Adaboost	96%
Decision tree	97%
SGD	97%
Linear regression	81%

Fig. 6. Model Accuracy Comparison Table

LEVEL and classify these calculated values to 3 classes. After feature extraction the result is shown in figure 5.

D. Model Evaluation

Model Evaluation is a step after training the model. Model Evaluation is an integral part of the model development process. It helps to find the best model that represents our data and how well the chosen model will work in the future. Evaluating model performance with the data used for training is not acceptable in data science because it can easily generate overoptimistic and over-fitted models.

This phase involves the training phase; there are several techniques [12] that are available in machine learning for this purpose here random forest classifier, logistic regression, K-Nearest Neighbor, SVM, Naive bayes, linear regression, Adaboost, Hist gradient boosting, Decision tree models are being used. Training accuracy comparison of different models shown in figure 6.

IV. RESULT

After feature extraction our dataset is split into 80 per- centage is for training and remaining 20 percentage is for testing. Then we trained it with 9 algorithms and get satisfying accuracy. In figure 6 we can see the accuracy comparison of models. Out of 9 algorithms we get better accuracy in 2 models, one is random forest classifier and another one is logistic regression. The better accuracy getting model is used
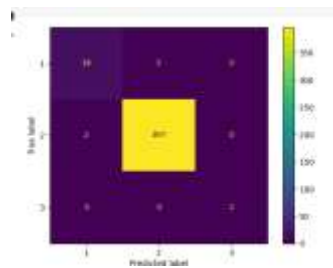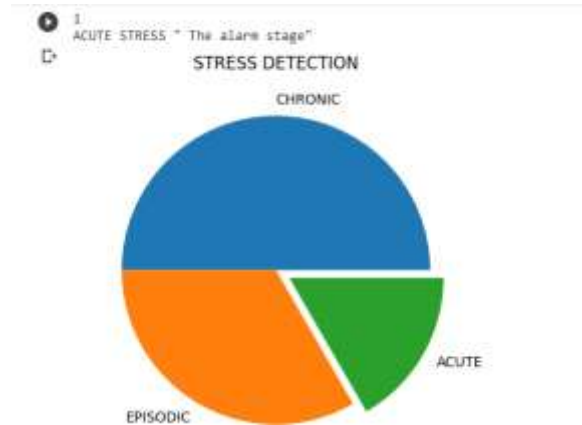

Fig. 7. Confusion Matrix

Fig. 8. Result After Testing (Example of Acute Stress)

for testing Here we get better accuracy in both Random forest classifier and logistic regression. Our model is a classification type so I select random forest for testing. And make a [13]confusion matrix for to get the represent the prediction summary in matrix form, it shows how many prediction are correct per class and it was shown in figure 7. Then we apply random forest classifier in testing phase and clearly detect the level of stress and display the class of stress and its level , also plot the level of stress by using a pie chart and we get 99 percentage of testing accuracy. The sample of one of our result is shown in figure 8.

V. CONCLUSION AND FUTURE SCOPE

In this model, we are calculating the mental stress of students in college. Our objective is to analyze stress in college students at different points in his lives. Some of the factors that effect on the students which often goes unnoticed. We will perform an analysis on how these factors affect the mind of a student. The dataset was taken from Vimala College (Au- tonomous), Thrissur and it consisted of 954 student's data. Our model is a classification type in supervised ML. There are 3 classes a)chronic b) episodic c)acute. Our objective is to detect the different level of stress in students. After implementing 9 machine learning techniques . In the randomized experimental runs, we concluded that Random Forest Classifier is a better algorithm get 99 percentage accuracy.

Like this project we can make predictions of stress in other areas like IT professionals, House wife, Teachers and any other category of person by just changing the questionnaire in the survey. And also predict the stress by real time images, videos and the wearable sensors also. We are the persons who use different social media platforms like twitter and [14] facebook, So we can make use of the post , videos, status , stories that were posted to detect the stress.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] A. Jain and M. Kumari, "Prediction of stress using machine learning and iot," in 2022 11th International Conference on System Modeling Advancement in Research Trends (SMART), pp. 282–285, 2022.

[2] A. Kene and S. Thakare, "Mental stress level prediction and clas- sification based on machine learning," in 2021 Smart Technologies, Communication and Robotics (STCR), pp. 1–7, 2021.

[3] L. Mohan and G. Panuganti, "Perceived stress prediction among em- ployees using machine learning techniques," in 2022 International Con- ference on Communication, Computing and Internet of Things (IC3IoT), pp. 1–6, 2022.

[4] K. Sengupta, "Stress detection: A predictive analysis," in 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1–6, 2021.

[5] A. Bannore, T. Gore, A. Raut, and K. Talele, "Mental stress detection using machine learning algorithm," in 2021 International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME), pp. 1–4, 2021.

[6] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in 2020 Second In- ternational Conference on Inventive Research in Computing Applications (ICIRCA), pp. 51–57, 2020.

[7] S. R. Anthay, V. Nagarjuna, T. V. Mahesh, T. Aravind Royal, and

S. V. Varma, "Detection of stress in humans wearing face masks using machine learning and image processing," in 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), pp. 1104–1110, 2022.

[8] S. K. Kanaparthi, S. P, L. P. Bellamkonda, B. Kadiam, and B. Mungara, "Detection of stress in it employees using machine learning technique," in 2022 International Conference on Applied Artificial Intelligence and Computing (ICAAIC), pp. 486–493, 2022.

[9] S. Elzeiny and M. Qaraqe, "Machine learning approaches to automatic stress detection: A review," in 2018 IEEE/ACS 15th International Conference on Computer Systems and Applications (AICCSA), pp. 1–6, 2018.

[10] P. B. Pankajavalli, G. S. Karthick, and R. Sakthivel, "An efficient machine learning framework for stress prediction via sensor integrated keyboard data," IEEE Access, vol. 9, pp. 95023–95035, 2021.

[11] V. Sasikala, T. Rajeswari, S. Naseema Begum, C. Divya Sri, and

M. Sravya, "Stress detection from sensor data using machine learning algorithms," in 2022 International Conference on Electronics and Re- newable Systems (ICEARS), pp. 1335–1340, 2022.

[12] R. J. Pramodhani, P. S. S. Vineela, V. S. Aseesh, K. Kumar, and B. K. Devi, "Stress prediction and detection in internet of things using learning methods," in 2022 4th International Conference on Inventive Research in Computing Applications (ICIRCA), pp. 303–309, 2022.

[13] A. de Santos Sierra, C. Sa´nchez A´vila, G. Bailador del Pozo, and

J. Guerra Casanova, "Stress detection by means of stress physiological template," in 2011 Third World Congress on Nature and Biologically Inspired Computing, pp. 131–136, 2011.

[14] C. Vuppalapati, M. S. khan, N. Raghu, P. Veluru, and S. Khursheed, "A system to detect mental stress using machine learning and mobile development," in 2018 International Conference on Machine Learning and Cybernetics (ICMLC), vol. 1, pp. 161–166, 2018.