Mental Health of Computer Science Students Factors Influencing and Statistical Analysis

1st Abdullah Ashfaq
Department of Computer Science
Faculty of Computing and Information Technology
University of the Punjab
Lahore, Pakistan
bsdsf21m034@pucit.edu.pk

2nd Muhammad Zubair
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab
Lahore, Pakistan
bsdsf21m004@pucit.edu.pk

3rd Anas Waleed Tahir
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab
Lahore, Pakistan
bsdsf21m028@pucit.edu.pk

4th Muhammad Abdullah Ghazi
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab
Lahore, Pakistan
bsdsf21m040@pucit.edu.pk

5th Etisam Ul Haq
Department of Data Science
Faculty of Computing and Information Technology
University of the Punjab
Lahore, Pakistan
bsdsf21m042@pucit.edu.pk

***Abstract*—The purpose of the research paper is to identify the factors that affect the mental health of university students. We are only interested in IT sector students. Many research papers are available regarding the mental health of medical students but relatively fewer reaches are present for IT students. We gathered data from the different universities in Lahore, Pakistan. Most students were undergraduates while some were from post-graduate level as well. We used statistical analysis to identify the factors influencing student mental health. To identify the factors,we developed a scale of stress. This scale was built by using 6 most important features from our dataset and applying advanced learning techniques. Then we tried to find which factors are affecting this scale and if there is any significant affect. We also tried to find correlation between multiple factors such as finding if high stress affect the CGPA which resulted in negation that high stress does not affect CGPA in any manner. Also we tried to find stress management of male and females and if there is significant difference in their response to stressors. We also tried to find which activities can help students to release stress.**

***Index Terms*—Neuroendoscopy, Unity3D, VR, tactile feedback, ETV.**

I. Introduction

A mental health problem is a health condition that affects a person's thoughts, behaviours, emotions, and interpersonal communication [8]. The most common symptoms in mental health are depression which is followed by anxiety and stress levels. The mental health factors can also be influenced by the psychological and physical problems [8]. Over the past few years, many research studies have been conducted on public health, diagnosis, and clinical administration regarding mental health problems [9].

Students' mental health problems are a growing concern in higher education. In recent times, these concerns have doubled due to the significant socioeconomic, political, and technological changes in the world. With the emergence of the ongoing novel COVID-19 pandemic, students in higher education are faced with increased mental health challenges. [7]

In COVID-19 pandemic many students faced imminent threats to their personal health, financial strain and uncertainty, risk of unemployment or redundancy, social distancing measures and isolation, loneliness and limited access to basic necessities, such as food and medicine [1]. In today's era, college students face mental health issues frequently [3]. In the past few years we have seen an increase in the symptoms of depression and anxiety among students, especially in college. Moreover, not enough research is done which is needed to understand the causes of these problems [4]. College life is stressful for many students due to the academic workload and sudden shift towards adulthood responsibilities without having a certain level of maturity. For instance, some students have to work for a living and support their families [2]. Race-related stress impacts entrepreneurship and mental health of students. [5] Our prime objective is to identify the factors impacting the student's mental health using advanced statistical analysis and machine learning techniques.

II. Demographics Section

We conducted a study using a dataset consisting of, around 100 individuals to evaluate the mental well being of computer science students. Through analyzing characteristics we gained a deeper understanding of their overall health. Notably, our findings indicated that male and female students accounted for 73.3% and 26.7% of the participants respectively. In terms of age, 32% of the population centered around the age group of 20 years old. Furthermore, our research showed that university enrollment patterns showcased the prevalence of the University of Punjab capturing over 63% of participants academic paths. Examining data from the year revealed a diverse landscape with first-year students comprising over 37% and third-year students making up more than 32%.

By evaluating Cumulative Grade Point Averages (CGPA) which reflect performance we were able to gain insights into participants achievements in their studies. Around thirty-one percent of the cohort achieved CGPA scores ranging from three to five highlighting a portion with academic abilities. Additionally, our investigation into sports engagement and lifestyle factors uncovered that over 46% of students were not actively involved in sports activities raising concerns, about behaviors. People have expressed worries, about their sleep patterns with more than half of them saying they typically get 4 to 6 hours of sleep per night which could potentially impact their well being.

Furthermore, the survey examined how content students were with their chosen field of study and discovered that over 37% of them felt satisfied. Finally in terms of ways to alleviate stress an interesting pattern emerged; over 53 percent of students mentioned engaging in activities as a means to relieve stress. This detailed account of the demographics provides a foundation for our study. Paves the way for further exploration into the intricate connections, between these factors and the mental well being of computer science students.

III. Results

1. There is no significant difference in stress scores among male and female students.
2. There is a significant difference in stress scores across different years of university.
3. There is no significant difference in stress scores of on-campus and off-campus students. 4. There is a significant difference in stress scores across different majors. 7. There is no significant correlation between Sports and Stress Scores.
8. There are significant differences in stress scores among different levels of sports engagement.
9. There is a significant correlation between Sleep and Stress Score.
10. There are significant differences in stress scores among different levels of sleep.
11. No significant differences in stress scores between groups with and without campus mistreatment.
12. There is a significant association between Gender and Stress Relief Activities.

IV. Discussion

Stress Score:

In our research paper, we aimed to find stress levels using 6 distinct features and advanced machine-learning techniques. These features were selected based on their relevance to student's stress. These features include academic workload, financial pressure, depression, anxiety, isolation(lack of socialization), and satisfaction with the field of education. We know that stress can come from many different sources, so we tried to capture all possible different factors that could impact students most. We used advanced machine learning algorithms to analyze a vast amount of data and develop a stress score system. We further categorized the stress scores into three distinct levels: low, moderate, and high. The following figure shows the normally distributed graph of stress score.

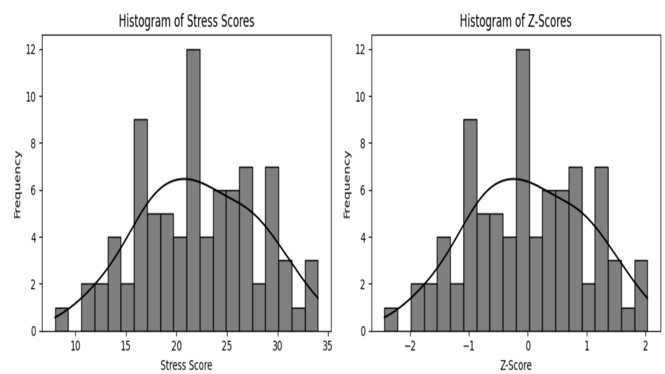

Fig. 1: Histogram of stress score

Statistical Analysis:

By applying statistical analysis, we were able to find valuable insights and factors affecting student's mental health.

Stress score comparison between genders:

There is a common perception among people that women are more likely to be stressed than men. Some researchers support this claim in their research domain. [10]. We can also find researches that claim both genders equally respond to stressors. [6]. So we also tried to compare stress among male and female students. To check if there is any significant difference in stress levels between male and female students, we plotted the stress score across both genders. The following figure shows stress scores across male and female students.

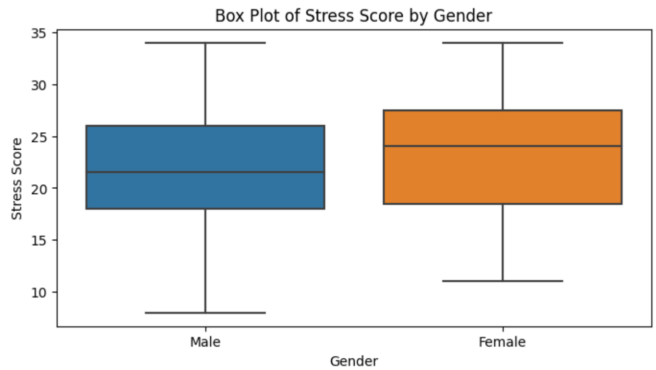

Fig. 2: Box plot of stress score across male and female students

As you can see the mean stress score of both genders is almost the same which suggests that there is no significant difference in stress scores among male and female students. To verify this claim, we also performed hypothesis testing with the following null and alternative hypotheses:
H_0: There is no significant difference in stress scores of male and female students.
H_1: There is a significant difference in stress scores of male and female participants
The T-statistic value is -1.1015 and the calculated p-value is 0.2777. Assuming a 95% confidence level and using the p-value test, we failed to reject the null hypothesis that there is no significant difference between the stress scores of male and female students.
Our research suggests that both male and female students are moderately stressed in the IT domain.

Stress score comparison across different years:
We collected data from students of all years of university. We tried to check whether there is any difference in how students perceive stress based on their stage in the degree program - whether they're just starting, in the middle, or nearing the end. The following figure shows a box plot of stress scores across 4 years of university.

It's interesting to see how stress levels can vary at different points in a student's academic journey. It's interesting to note that students tend to experience moderate stress during their first year of college. However, as they become better adjusted to their new environment during their second year, stress tends to decrease to a more manageable level. Unfortunately, the following years seem to bring a resurgence of stress. It is possible that the uncertainty of the future may contribute to these increased stress levels. We also used ANOVA analysis to verify this claim of difference in stress across 4 years of university. The level of confidence is assumed to be 95%. The null and alternative hypotheses are the following: H_0:

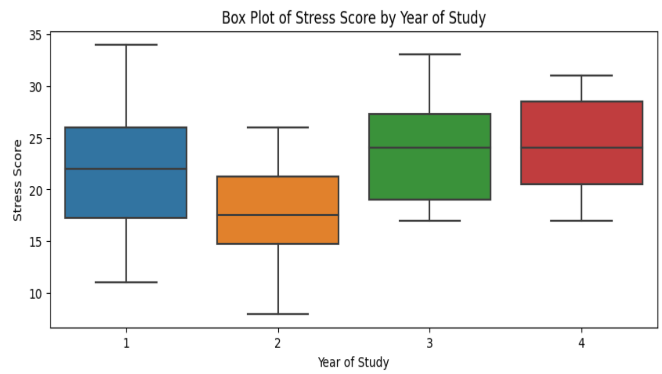

Fig. 3: Box plot of stress score across 4 years of university

There is no significant difference in stress across different years of university.
H_1: There is a significant difference in stress across different years of university.
The f-statistic value is 3.8954 while the corresponding p-value is 0.0118 which is smaller than 0.05 which is the alpha value showing the level of confidence. So we will reject the null hypothesis. This also verifies that there is a significant difference in stress across different years of university.

Stress score comparison between on-campus & off-campus students:
As we collected data from both hostelers and commuter students so we were intrigued to find whether the struggles of hostel life impact students' mental health or not. To find this we performed an ANOVA analysis on the following null and alternative hypotheses. H_0: There is no significant difference in stress scores among n-Campus and Off-Campus residents..
H_1: There is a significant difference in stress scores among On-Campus and Off-Campus residents.
The values of the f-statistic and p-value are 0.0043 and 0.9476 respectively. Assuming a confidence level of 95%, we fail to reject the null hypothesis. With the rejection, we can presume that there is no significant difference in stress scores among on-campus and off-campus residents. The following figure of stress scores among on-campus and off-campus resident students also tells the same story.

Stress score comparison across different majors: As there exist multiple majors in the IT domain, so we tried to compare the 4 most common majors of the IT domain. As seen in the figure, the mean of stress scores across all majors varies which suggests that there is a significant difference in stress scores in different majors. To prove this, we performed an ANOVA analysis here too with following null and alternative hypothesis. H_0: There is no significant difference in stress scores across different majors..
H_1: There is a significant difference in stress scores across different majors..

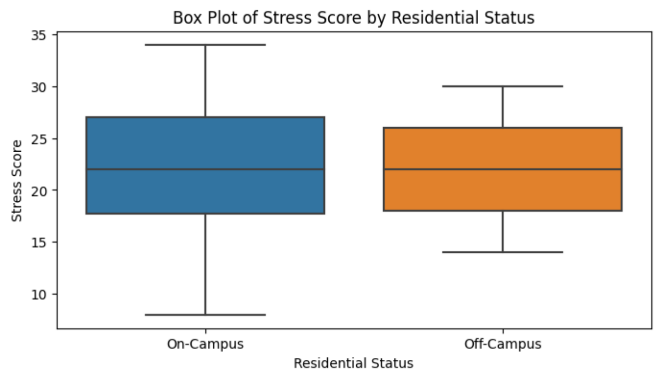

Fig. 4: Box Plot of Stress Score among on-campus and off-campus students

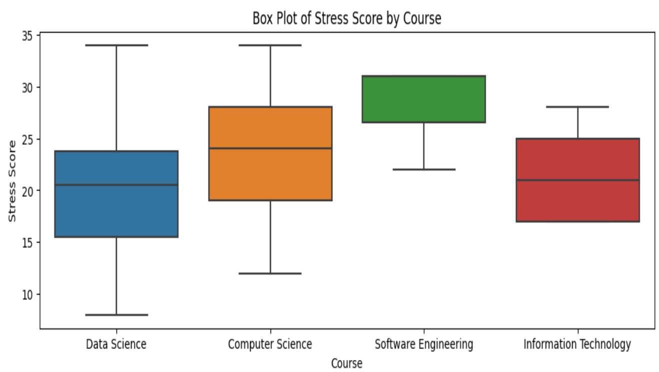

Fig. 5: Box Plot of Stress Score across different majors

With an F-statistic value of 3.7003 and p-value of 0.015, we reject the null hypothesis which proves that there is a significant difference in stress score across different majors.

Correlation between CGPA and Stress Score:
The common sense suggests that if a student is stressed by studies then it will affect his/her CGPA but during our r&d process we found that stress does not affect cgpa in any way so we decided to find out if stress significantly affects CGPA of students. To find this, the hypotheses are as follows:
H_0: There is a significant correlation between CGPA and Stress Score.
H_1: There is no significant correlation between CGPA and Stress Score.
After correlation test on CGPA and Stress score we found that there is no significant relationship between CGPA and stress score which is a quite interesting fact.

Correlation between sports and stress score:
Hypothesis testing for finding the correlation between sports engagement and stress scores shows that there is no significant correlation between these two entities. The hypotheses are as follows:

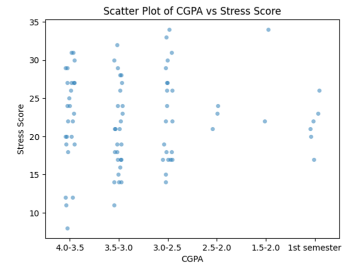

Fig. 6: Scatter Plot of CGPA & Stress Score

H_0: There is no significant correlation between Sports and Stress Score.
H_1: There is a significant correlation between Sports and Stress Score.
The value of Pearson's correlation coefficient is -0.1944 and the calculated p-value is 0.06636. Assuming the level of confidence to be 95%, and through the p-value test, we fail to reject the null hypothesis that claims there is a significant correlation between Sports and Stress Score.

Stress score comparison across different levels of sports:
Although we found that there is no significant correlation between Sports and Stress Scores, we are also interested in finding whether there are significant differences in stress scores among different levels of sports engagement or not. We categorise students into four categories based on sports activities. These are no sports, low, moderate and high. Those students who play sports more than 7 times a month are categorized as high, 4 to 6 as moderate, 1 to 3 as low and no sports means they do not play even once a month. The following figure shows the plotting of these levels based on stress scores.

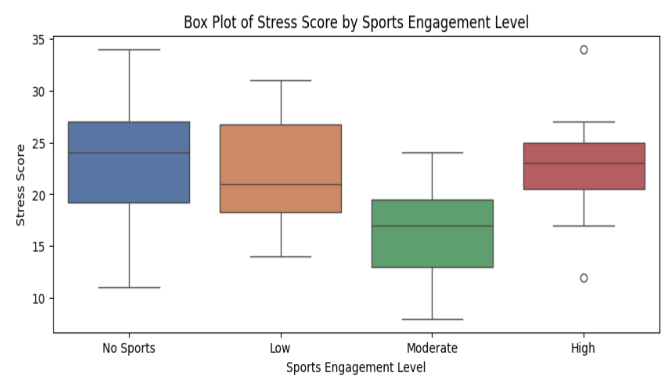

Fig. 7: Box Plot of Stress Score by Sports Engagement Level

We used ANOVA analysis to find whether significant differences occurred among different levels of sports and stress scores. The level of confidence is assumed to be 95%. The hypotheses are as follows:
H_0: There are no significant differences in stress scores

among different levels of sports engagement.
H_1: There are significant differences in stress scores among different levels of sports engagement.
The F-statistic value is 4.6984 while the p-value is 0.004369 which is less than 0.05 of alpha value. Through this hypothesis testing, we reject the null hypothesis. This gives us the result that there are significant differences in stress scores among different levels of sports engagement. We can see a general trend in the above graph as well for the students, who play moderate levels of sports have lower stress scores.

Correlation between sleep and stress score:
The next thing we want to figure out is whether sleep and stress scores have a significant correlation between them. To test this, we implied hypothesis testing for correlation. We assumed the level of confidence is 95%. The hypotheses are as follows:
H_0: There is no significant correlation between Sleep and Stress Score.
H_1: There is a significant correlation between Sleep and Stress Score.
The calculated value of the correlation coefficient is -0.2859 while the p-value is 0.006291. This rejects our null hypothesis. So we are 95% confident that there exists a significant correlation between sleep and stress score. The following Scatter Plot shows some level of correlation as well.

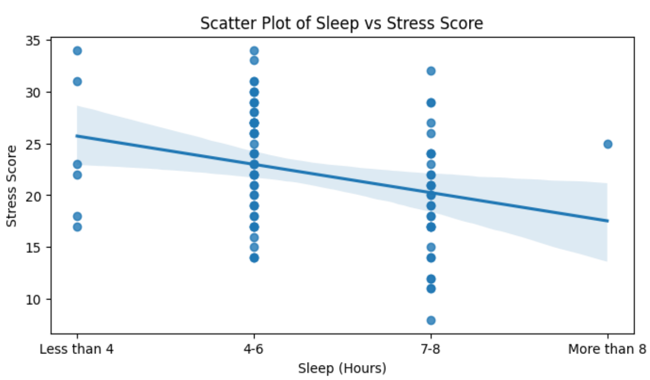

Fig. 8: Scatter Plot of Sleep vs Stress Score

Through the above-mentioned analysis, we found that there is a significant correlation between sleep and stress scores. The value of correlation is negative which means by increasing one entity, other thing will decrease. The following plot also shows that as the number of sleep hours increases, stress scores overall decrease.

Stress score comparison across different levels of sleep:
To test whether the difference in stress score across different levels of sleep is significant or not, we used one-way ANOVA testing. Following are the hypotheses:
H_0: No significant differences in stress scores among different levels of sleep.

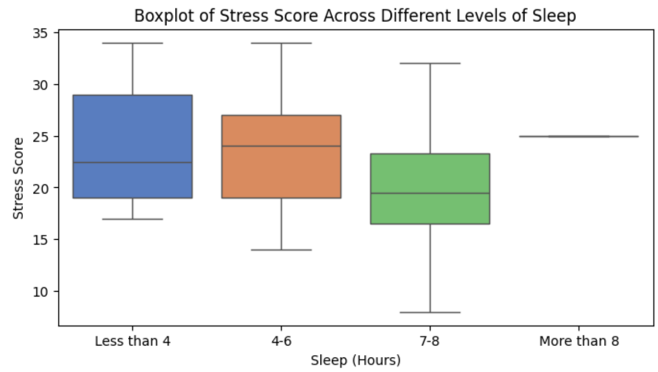

Fig. 9: Boxplot of Stress Score Across Different Levels of Sleep

H_1: There are significant differences in stress scores among different levels of sleep.
The f-statistic value is 3.7373 while the corresponding p-value is 0.01410 which is smaller than 0.05 which is the alpha value showing level of confidence. So we will reject the null hypothesis. This shows that there is a significant difference between stress score across different levels of sleep.

Stress score comparison across groups of mistreatment on campus:
Some students may have higher stress score if they experience mistreatment on campus. The box plot also shows that those students who experience mistreatment have higher stress scores.

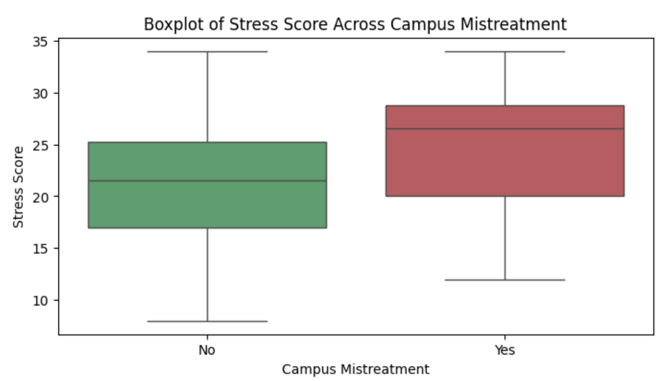

Fig. 10: Boxplot of Stress Score Across Different Levels of Sleep

But the question is whether this difference is significant or not. To verify this, we used hypothesis testing. Let us say the level of confidence is 95%. The hypotheses are as follows:
H_0: No significant differences in stress scores between groups with and without campus mistreatment.
H_1: There is a significant difference in stress scores between groups with and without campus mistreatment.

Through the p-value test, the value of p comes to 0.06156 which is slightly greater than alpha = 0.05. So we fail to reject the null hypothesis. Although the above figure shows that there is a difference between stress scores between the two groups, but hypothesis testing with 95% confidence says that the difference is not significant.

Association between gender and stress relief activities: We have seen how different factors affect stress on a student. We also want to find the association between gender and stress relief activities. To find this, we used the chi-square test for this purpose. Let's say the following are the hypotheses:
H_0: There is no significant association between Gender and Stress Relief Activities.
H_1: There is a significant association between Gender and Stress Relief Activities.
We used the p-value test to verify the claim. The p-value after the calculation is 0.02297 which is lesser than alpha = 0.05. So we will reject the null hypothesis. After this hypothesis testing, we are 95% confident that there exists a significant association between gender and stress relief activities. Let's have a look at the following contingency table for gender and stress relief activities.

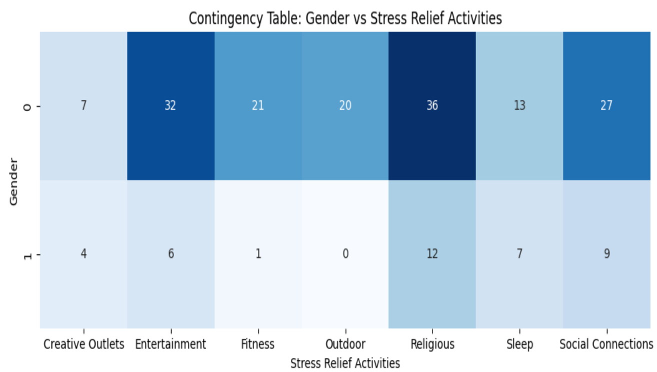

Fig. 11: Contingency Table: Gender vs Stress Relief Activities

The table also shows that there exists a significant dependency between gender and stress relief activities. Both males and females do mostly religious activities for stress relief. Males more focus on entertainment and fitness activities while females focus on social connections and sleep.

V. Machine Learning

First, it is important to know why machine learning is helpful in the student's mental health. We saw several online healthcare sites that predict a person's health score and most of them are based on PSS means Perceived Stress Scale but these predicting models are for a universal audience. After several interactions and suggestions from the students, we feel that it would be great to make a predictive model that is trained specifically for the computer sector. We plan to design a model that greatly handles mental health specifically stress levels and depression among the individuals based on specific parameters.

Before applying machine learning it is essential to know about the data. After applying the exploratory data analysis which you find out in the demographics section there is one part missing which is the correlation coefficients of all the features so that it helps us to identify high as well as low correlated features. The features such as anxiety, isolation, university workload, academic pressure, and financial pressure of the students are highly correlated. Following is the heatmap of all the correlation coefficients sorted from most correlated to least correlated features.

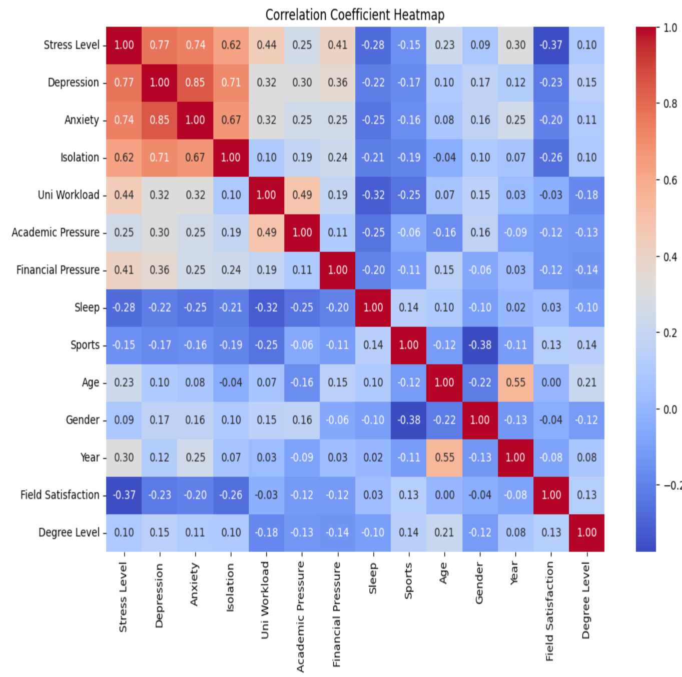

Fig. 12: Correlation Coefficients

After experimenting with well-known machine learning models such as logistic regression, decision trees, and support vector machines, we found that random forest is the most efficient for our dataset. We learned from several sources that random forest is an ensemble strategy that avoids high variances and predicts optimal results through hyperparametric tuning. One of the hidden benefits of random forest is that it reduces the effort of feature selection and does not require as much feature preprocessing as other models, without losing the accuracy.

While doing the hyperparameteric tuning, the main focus was on the 'n estimator' which sets the number of decision trees in the random forest and this is obtained through the cross-validation method. Our target columns are stress level and depression level and this is the classification problem.

The stress level is from the range 1 to 3 where 1 means low stress, 2 means moderate and 3 is high stress. Similarly, the depression level ranges from 1 to 5 representing lowest to highest depression.

Our model gives approximately 90 percent accuracy after cross-validation. The one of main features of random forest is the feature importance. When a model is trained then it identifies the important features and gives them the weights. It helps to identify and reduce the number of features by retaining the model's accuracy. The following table shows the features importance of model1 and model2 and are sorted on their sum scores.

Feature	Model1	Model2	Total Importance
Stress Score	0.434691	0.203910	0.638601
Anxiety	0.096186	0.193034	0.289220
Isolation	0.088368	0.095503	0.183871
Future Insecurity	0.039346	0.058154	0.097500
Financial Pressure	0.037055	0.059915	0.096970
Year	0.062900	0.031987	0.094887
Campus Networking	0.040695	0.049492	0.090187
Field Satisfaction	0.052441	0.031797	0.084238
Academic Pressure	0.032764	0.039322	0.072086
Age	0.028016	0.037754	0.065770
CGPA	0.024120	0.034357	0.058477
Workload	0.015996	0.042341	0.058338
Sports	0.007862	0.035823	0.043685
Campus Mistreat	0.016603	0.019893	0.036497
Sleep	0.006299	0.025991	0.032289
Gender	0.008055	0.022966	0.031021
Residential Status	0.003389	0.015109	0.018499
Degree Level	0.005214	0.002652	0.007865

TABLE I: Feature Importance Table

VI. Conclusion

We have thoroughly analyzed each student's mental health in the computer science domain. Results show that there is no significant difference between stress scores among males and females, on-campus and off-campus, with and without campus mistreatment, and with sports. On the other hand, we have found that some factors have significant statistical differences with stress scores which are years of university, university majors, levels of sports engagements, and sleep. Also, there is a significant association between Gender and Stress Relief Activities. In the end, after a deep understanding of mental health factors, machine learning models are designed to help identify such students, and these models have an impressive accuracy of 90 percent.

This study represents a significant advancement in addressing mental health concerns and their aftereffects within the student community.

References

[1] Rosie Allen, Chathurika Kannangara, Mahimna Vyas, and Jerome Carson. European university students' mental health during covid-19: Exploring attitudes towards covid-19 and governmental response. *Current Psychology*, 42(23):20165–20178, 2023.

[2] Jeffrey Jensen Arnett. Emerging adulthood: A theory of development from the late teens through the twenties. *American psychologist*, 55(5):469, 2000.

[3] Carlos Blanco, Mayumi Okuda, Crystal Wright, Deborah S Hasin, Bridget F Grant, Shang-Min Liu, and Mark Olfson. Mental health of college students and their non–college-attending peers: results from the national epidemiologic study on alcohol and related conditions. *Archives of general psychiatry*, 65(12):1429–1437, 2008.

[4] Emily G Lattie, Sarah Ketchen Lipson, and Daniel Eisenberg. Technology and college student mental health: challenges and opportunities. *Frontiers in psychiatry*, 10:246, 2019.

[5] Thema Monroe-White and Ebony Mcgee. The critical role of race-related stress and racial activism on stem graduate students' career aspirations: An intersectional perspective. 2024.

[6] Usha Rout. Gender differences in stress, satisfaction and mental wellbeing among general practitioners in england. *Psychology, Health & Medicine*, 4(4):345–354, 1999.

[7] Nahal Salimi, Bryan Gere, William Talley, and Bridget Irioogbe. College students mental health challenges: Concerns and considerations in the covid-19 pandemic. *Journal of College Student Psychotherapy*, 37(1):39–51, 2023.

[8] Nor Safika Mohd Shafiee and Sofianita Mutalib. Prediction of mental health problems among higher education student using machine learning. *International Journal of Education and Management Engineering (IJEME)*, 10(6):1–9, 2020.

[9] Adrian BR Shatte, Delyse M Hutchinson, and Samantha J Teague. Machine learning in mental health: a scoping review of methods and applications. *Psychological medicine*, 49(9):1426–1448, 2019.

[10] Diana M Zuckerman. Stress, self-esteem, and mental health: How does gender make a difference? *Sex roles*, 20(7-8):429–444, 1989.