# Machine learning-assisted prediction of associated risk factors for depression, anxiety and stress among nursing students

**S. Deivasigamani[1*] Hani J. Hamad[2], Zaher Al-Bashabsheh[2,], Madhanagopal Jagannathan[3] B. Sakthivel[4], K.Gowrishankar[5]**

[2] Department of Clinical Nutrition and Dietetics, Faculty of Allied Medical Sciences, Philadelphia University, Amman 19392, Jordan. hhamad@philadelphia.edu.jo, zalbashabsheh@philadelphia.edu.jo
[3] School of Physiotherapy, Sri Balaji Vidyapeeth, (Deemed to be university), Karaikal campus 609609, India. physiotherapy@k.sbvu.ac.in
[4]Department of ECE, Pandian Sarswathi yadava Engineering College, Sivagangai, TamilNadu, India. 786sakthivel@gmail.com.
[5] Department of Marine Engineering, AMET University, Tamilnadu, India. gowrishankark@ametuniv.ac.in
[*] **Correspondence address**: [1]Faculty of Engineering, Technology & Built Environment, UCSI University, Kuala Lumpur, Malaysia. deivasigamani@ucsiuniversity.edu.my

**Abstract**
During their time as students, nursing students are exposed to numerous stressors that result in physical and mental health issues and poor academic performance. Using the machine learning method, this article investigates the depression, anxiety, and stress among Malaysian university nursing students. The subjects were assured of the secrecy and anonymity of the data collected. Multiple logistic regression and scale 21 was used to identify significant relationships between variables. The sample comprised 83.90 percent female and 16.09 percent male students. The proposed system achieves 80.3% Sensitivity, 80.5% Specificity, 89.8% Accuracy, 89.8% Precision, and 73.23 JSC (Jaccard Similarity coefficient). Therefore, the proposed system's final average CR (Classification Rate) is approximately 89.6%. In this article, the k-fold cross-validation method is utilized to cross-validate the experimental results of the proposed method. According to various university-level surveys, depression, anxiety, and stress affect 47.8 percent, 66.34 percent, and 36.54 percent of students. According to the findings of this study, respondents have a high prevalence of Sp, Se, Acc, Pre, and JSC, was achieved using machine learning method.
**Keywords:** Prevalence; depression; anxiety; stress; university students; machine learning methods

## Introduction:

Depression is a common mental disorder characterized by persistent dissatisfaction and loss of interest in enjoyable activities. It affects around 350 million people globally. Every year, over 700,000 people commit suicide. Being female and aged between 15 and 29 years has been recognized as a risk factor for both depression and suicide - (Van de Velde et al., 2010; WHO, 2015b) - is a group of nursing students. Due to their academic courses, nursing students experience more pressure than students in other fields, such as science and the arts (Shikai et al., 2007). Among students, the literature has consistently reported a number of variables to depression, including the student's age, satisfaction with the program, and academic year, as well as financial strains, academic performance, exam stress, physical activity, working hours, workload, stable relationships, social roles, and cognitive processes (, Sarath and Jeewanthika (2016); South Africa study). In addition to these factors, nursing students also undergo clinical training that requires constant close human contact and emotional engagement during their studies, which are major causes of anxiety and stress, potentially predisposing them to depression (Gibbons, 2010; Jimenez et al., 2010; Melo et al., 2010; Pulido-Martos et al., 2012; Moreira and Furegato, 2013; Rudman and Gustavsson, 2012). These studies highlight the importance of understanding and addressing the unique factors faced by nursing students to prevent and manage depression. However, these factors are identified using conventional statistical techniques (CSTs) like regression analysis, based on resource conservation theory, self-determination theory, and job demands-resources theory.

Machine learning (ML) has gained significant attention for identifying risk factors for diseases by studying complex and non-linear interactions between datasets. This holistic approach allows ML algorithms to uncover hidden patterns and relationships that CST may overlook. By analyzing large amounts of data from various sources, ML can identify subtle risk factors that contribute to disease development, helping healthcare professionals take preventive measures and design targeted interventions. The utilization of ML in disease risk assessment holds great potential for improving public health outcomes and reducing the burden of diseases on individuals and healthcare systems. However, CSTs are limited in their ability to reflect real-world complexities and anticipate future data due to their assumption of linearity between variables. ML techniques can aid in illness detection by forecasting disease risk. In this context, few research was conducted to predict associated factors for depression among young children, older persons and university students, not nursing students using ML algorithms. Therefore, the purpose of this study was to find the important associated factors influencing the risk of depression among undergraduate nursing students using random forest (RF).

## Methodology

**Study participants:** This cross-sectional study was conducted among 205 female nursing students (Malay, Chinese, and Indians), aged between 17 and 21 years from AIMST University, Kedah, Malaysia. All the students of nursing program were invited to participate in this study. This study was carried out between March 2019 and January 2020. The participants were asked to fill-up the questionaries mentioned in the table 1. These questionaries were adopted from the literature. The study was approved by the AIMST University Human Ethics Committee and carried out in accordance with AIMST University's ethical standards in Kedah, Malaysia, as well as the tenets of the Helsinki Declaration. All study participants provided written informed consent. The attributes of the students and their score levels are depicted in Table 2.

**Table 1 Strength of the participating students for mental analysis system**

| Mental status | Number of students participated | Total |
|---|---|---|
| **Anxiety-Yes case** | 129 | 205 |
| **Anxiety-No case** | 76 | |
| **Depression-Yes case** | 104 | 205 |
| **Depression-No case** | 101 | |
| **Stress- Yes case** | 75 | 205 |
| **Stress- No case** | 130 | |

**Table 2 Attributes and its scoring levels**

| Attributes | Scoring 1 | Scoring 2 | Other scores |
|---|---|---|---|
| **Age** | 1= (age from 15-20) | 2=(age from 21-25) | |
| **Race** | 1= malay | 2= chinese | 3=Indian; 4=others |
| **Gender** | 1= male | 2= female | |
| **Academic** | 1= year 1 semester 1 | 2= year 1 semester 2 | 3= year 2 semester 1<br>4= year 2 semester 2<br>5= year 3 semester 1<br>6= year 3 semester 2 |
| **Sponsorship** | 1= yes | 2= No | |
| **Illness** | 1= yes | 2= No | |
| **Before Gym** | 1= yes | 2= No | |
| **After Gym** | 1= yes | 2= No | |
| **Before Aerobic Exercise** | 1= yes | 2= No | |
| **After Aerobic Exercise** | 1= yes | 2= No | |
| **Entertaiment** | 1= yes | 2= No | |
| **Sleep** | 1= yes | 2= No | |
| **Before Psy disturbance** | 1= yes | 2= No | |
| **After Psy disturbance** | 1= yes | 2= No | |

| | | | |
|---|---|---|---|
| **Before Smoking** | 1= yes | 2= No | |
| **After Smoking** | 1= yes | 2= No | |
| **Before Drinker** | 1= yes | 2= No | |
| **After Drinker** | 1= yes | 2= No | |
| **Disturbance** | 1= yes | 2= No | |
| **Violence** | 1= yes | 2= No | |
| **Physical Health** | 1= good | 2=bad | |
| **Mental Health** | 1= good | 2= bad | |
| **Income** | 1= below RM 2000 | 2= RM 2001 AND RM 5000 | 3= >RM5000 |
| **Death of Parents** | 1= yes | 2= No | |
| **Death of Relatives** | 1= yes | 2= No | |
| **Result** | 1= <2.00 | 2= >2.01 TO 3.00 | 3= >3.01; 4= Not applicable |
| **Knowledge** | 1= good | 2= poor | 3= Not applicable |

**Random Forest (RF) classification approach**

In this article, the mental status of nursing college students is analyzed into either Anxiety, Depression, or Stress using Random Forest (RF) classification approach. This proposed method consists of two stages training and testing. During the training stage of the RF classification, the attributes of Anxiety, Depression, and Stress affected students are collected and trained by the RF classifier in training mode. During the testing stage of the RF classification, the attributes from the student (to be classified or tested) are fed into RF classifier in the classification phase using the trained patterns, which produces the classification results. The workflow of the proposed methodology is illustrated in Fig.1.

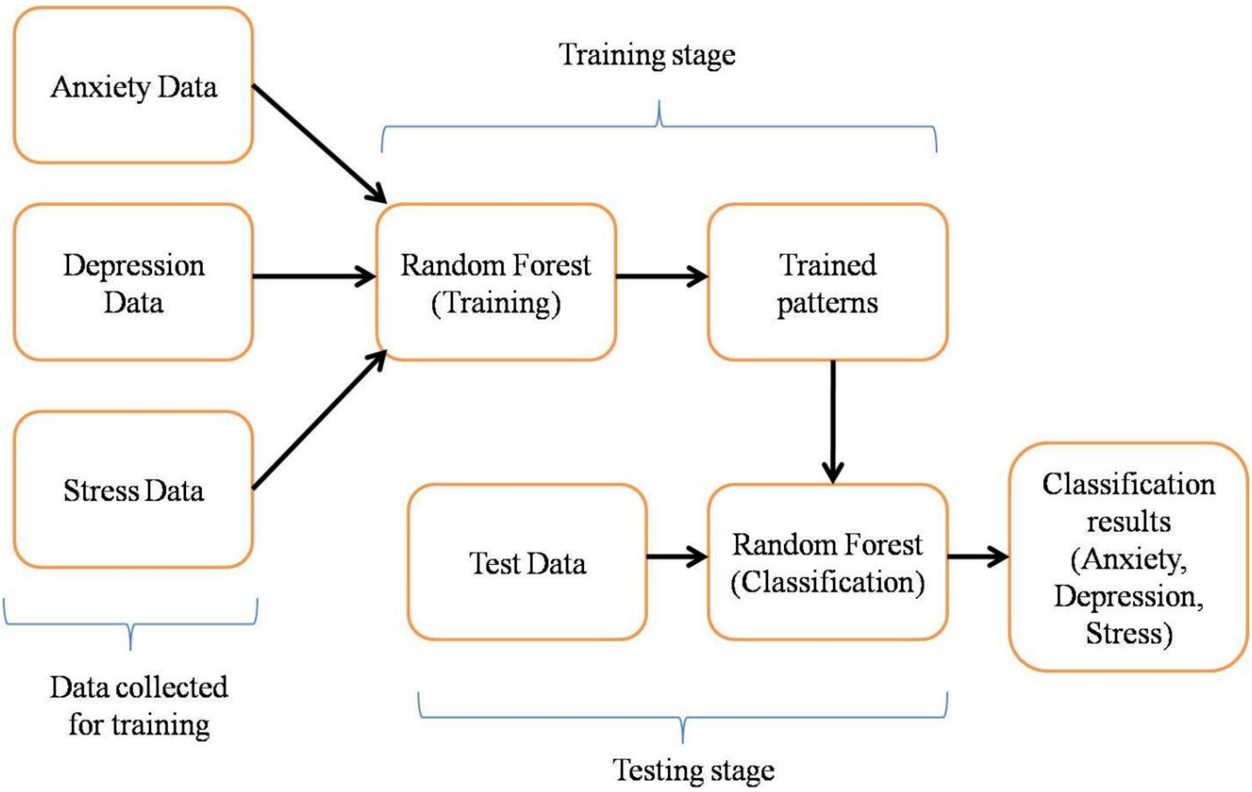


**Figure 1. Proposed mental analysis system for college students**

RF classifier is one machine learning methodology without using hyperparameters to classify tasks. RF classification algorithm can be used in classification and regression tasks due to its diversity mechanism. This algorithm is explained in the following steps:

**Step 1:** The training data is randomly split into some sub-samples.
**Step 2:** Each sub-samples are applied to the decision tree.
**Step 3:** Generate a vote for each decision tree output.
**Step 4:** Combine all the votes to produce the trained patterns.
**Step 5:** Apply the test data into the decision tree to compute the vote using the firefly algorithm.
**Step 6:** This vote for test data is compared with the trained patterns produced during training the classifier.
**Step 7:** The classification result is correlated with the final vote.

**RF classification algorithm**

The proposed systematic workflow of the RF classification algorithm is shown in Figure 2, which describes the seven steps. The entire training data is divided into N number of sub-set models, and each sub-set model is individually applied to the decision tree. A *decision tree* is a probability tree that chooses the required systematic output from the set of input models. This decision tree is categorized into a Regression tree and a Continuous variable tree. A regression decision tree is used if the training samples are non-linear. If the training samples are linear, then a Continuous variable decision tree is used. In this article, the data in each sub-set model belongs to non-linear, and hence regression decision tree is used.

A set of nodes structures the regression decision tree as Chance Node (CN), Decision Node (DN), and End Node (EN). A circle represents the CN in the tree, the DN is represented by a square, and a circle represents EN. The function of DN is to divide the sub-set based on the threshold value, which is the average of the functional values of the data. The firefly algorithm is used to compute the voting of the decision tree. All computed voting is consolidated and compared with the test data's voting to produce the classification results.

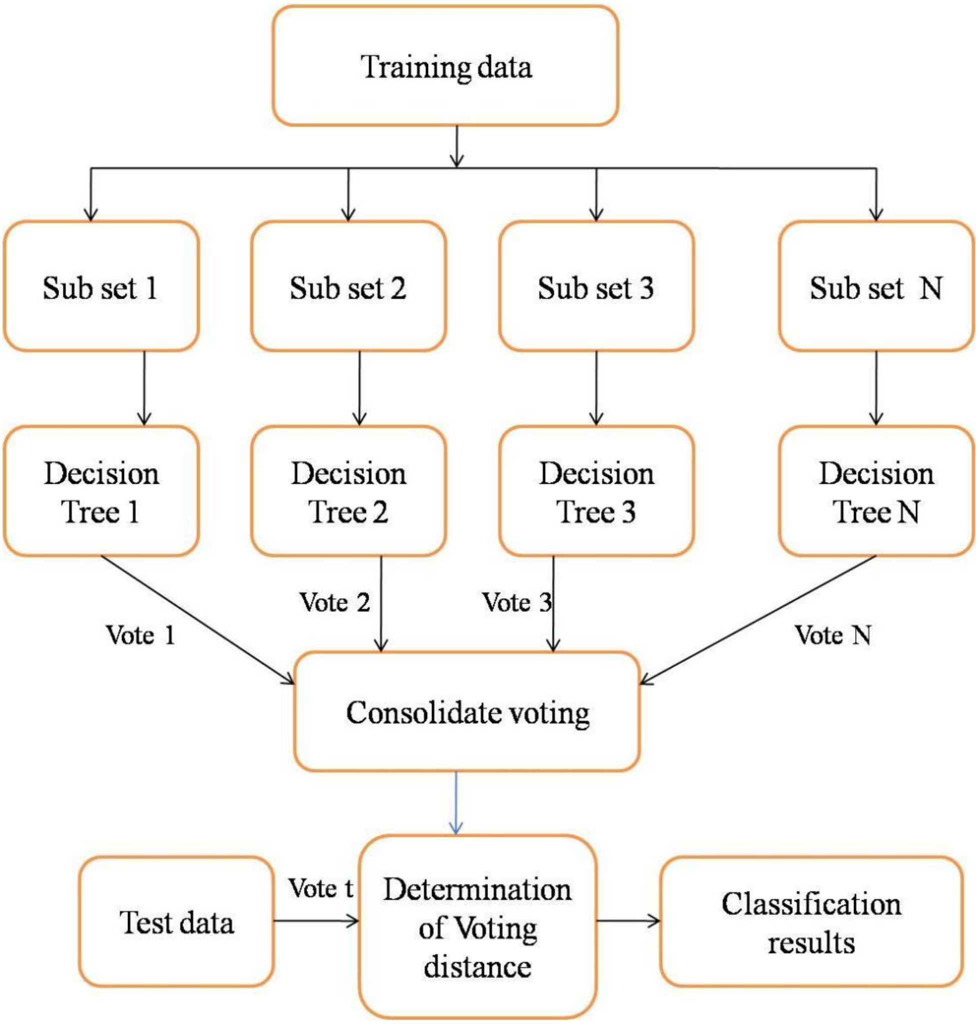


**Figure 2 Proposed RF classification Algorithm**

**Results and Discussions**

This article analyses the student's cognitive behaviour using the RF classification approach. The proposed system is experimentally analysed in Intel i3 multi-core 3.5 GHz processor with

8 GB internal RAM. 615 data are collected from university students in the category of Anxiety (205 data), Depression (205 data), and Stress (205 data). These collected data are split into a training set (63%) and a testing set (37%).

- For the Anxiety case, 127 data are used for training, and 78 are used for testing. Among the 78-testing data, 69 data belong to the 'Anxiety-Yes case,' and 9 data belong to the 'Anxiety-No case.'
- For the Depression case, 147 data are used for training, and 58 are used for testing. Among the 58-testing data, 29 data belong to the 'Depression-Yes case,' and 29 data belong to the 'Depression-No case.'
- For the Stress case, 114 data are used for training, and 91 data are used for testing. Among the 91-testing data, 10 data belong to the 'Stress-Yes case,' and 81 data belong to the 'Stress-No case.'

The following equations (S. Deivasigamani. et. al.(2022) are used to analyse the performance of the proposed mental analysis system to identify the mental status of the student as either Anxiety or depression, or stress.

$$Sensitivity\ (Se) = \frac{TP}{TP + FN} \times 100\%$$

$$Specificity\ (Sp) = \frac{TN}{TN + FP} \times 100\%$$

$$Accuracy\ (Acc) = \frac{TP+TN}{TP+TN+FP+FN} \times 100\%$$

$$Precision\ (Pre) = \frac{TP}{TP+FP} \times 100\%$$

$$Jaccard\ Similarity\ Coefficient\ (JSC) = \frac{TP}{TP+FP+FN} \times 100\%$$

Whereas TP is the True Positive which represents the total number of correctly identified Yes cases in each status category, TN is the True Negative, which represents the total number of correctly identified No casein each status category. FP is the False Positive, representing the total number of wrongly identified Yes cases in each status category. FN is the False Negative, representing the total number of wrongly identified No casein each status category.

The proposed system achieves 98.4% of Se, 53.3% of Sp, 89.7% of Acc, 89.8% of Pre, and 88.5% of JSC for Anxiety cases. The proposed system achieves 89.6% of Se, 89.6% of Sp, 89.6% of Acc, 89.6% of Pre, and 81.2% of JSC for depression cases. The proposed system achieves 52.9% of Se, 98.6% of Sp, 90.1% of Acc, 90% of Pre, and 50% of JSC for stress cases. The experimental results for the mental analysis system are depicted in Table 3.

**Table 3 Experimental results for mental analysis system**

| Mental status | Computational parameters | Se (%) | Sp (%) | Acc (%) | Pre (%) | JSC (%) |
|---|---|---|---|---|---|---|
| **Anxiety** | TP=61 | 98.4 | 53.3 | 89.7 | 89.8 | 88.5 |
| | TN=8 | | | | | |
| | FP=7 | | | | | |
| | FN=1 | | | | | |
| **Depression** | TP=61 | 89.6 | 89.6 | 89.6 | 89.6 | 81.2 |
| | TN=8 | | | | | |
| | FP=7 | | | | | |
| | FN=1 | | | | | |
| **Stress** | TP=61 | 52.9 | 98.6 | 90.1 | 90 | 50 |
| | TN=8 | | | | | |
| | FP=7 | | | | | |
| | FN=1 | | | | | |
| Average | | 80 .3 | 80 .5 | 89. 8 | 89. 8 | 73. 23 |

**Classification Rate (CR)**

CR is defined as the ratio between the number of samples correctly detected and the total number of samples in each case. It is measured in % and varies between 0 and 100. The proposed methodology stated in this article correctly classifies 62 Anxiety-Yes data over 69 and achieves 89.8% of CR and also correctly classifies 8 Anxiety-No data over nine and achieves 88.8% of CR. The average CR for the Anxiety case is, therefore, 89.3%. The proposed methodology stated in this article correctly classifies 26Depression-Yes data over 29 and achieves 89.6% of CR and also correctly classifies 26Depression-No data over 29 and achieves 89.6% of CR. The average CR for Depression cases is, therefore, 89.6%. The proposed methodology stated in this article correctly classifies 9Stress-Yes data over ten and achieves 90% of CR and also correctly classifies 73Stress-No data over 81 and achieves 90.1% of CR. The average CR for the Stress case is, therefore, 90.05%. Hence, the final average CR of the proposed system is about 89.6%.

**Cross validation**

It is a methodology to cross-check the experimental results of the proposed student's mental analysis system. In this article, the k-fold cross-validation method is used to cross-validate the proposed method's experimental results. This article uses a 4-fold cross-validation method which randomly chooses 200 data from the test dataset, and these data are split into four equal sub-datasets (fold). Hence, each fold contains 50 test data, and the proposed method is applied to each fold dataset. The accuracy of each fold is computed individually, and the average accuracy is determined using the individual fold accuracy.

The folded one (k=1) correctly identifies 18 data over 20 and achieves 90% of CR. The fold two (k=2) correctly identifies 19 data over 20 and achieves 95% of CR. The fold three (k=3) correctly identifies 16 data over 20 and achieves 80% of CR. The fold four (k=4) correctly identifies 18 data over 20 and achieves 90% of CR. Therefore, the average CR using the 4-fold validation method is 88.7%. The cross-validation results are nearby equal to the experimental CR results, as illustrated in Table 4. Hence, the proposed methodology stated in this article is most suitable for checking or identifying the student's mental status.

**Table 4 4-fold cross validation results**

| **Fold number** | **Number of data tested** | **Correctly classified data** | **CR in %** |
|---|---|---|---|
| Fold 1 | 20 | 18 | 90 |
| Fold 2 | 20 | 19 | 95 |
| Fold 3 | 20 | 16 | 80 |
| Fold 4 | 20 | 18 | 90 |
| | 80 | 71 | 88.7 |


**Figure 3 4-fold cross validation**

## CONCLUSION

Even though various studies have been done around the world regarding the prevalence of DSA among university students, there are still some or few studies that revealed a high prevalence of DSA among University students. This study finding shows that respondents have a high prevalence of DSA and a significant positive correlation between depression, anxiety, and stress. This article examines the best cost-effective approaches for assessing university students for DSA eligibility using an automated computer system. Researchers can detect DSA anomalies during extensive population screening using the method proposed in this article.

## References

Brahima Sory keita, Ir.Pratap Nair, Haarindra Prasad, Sudhakara pandian, S.Deivasigamani, "Classification of Benign and Malignant MRIs using SVM Classifier for Brain Tumor Detection," *International Journal of Engineering Trends and Technology*, vol. 70, no. 3, pp. 234-240, 2022. Crossref, https://doi.org/10.14445/22315381/IJETT-V70I3P226

Chatterjee, S., Saha, I., Mukhopadhyay, S., Misra, R., Chakraborty, A., & Bhattacharya, A. (2014). Depression among nursing students in an Indian government college. British Journal of Nursing, 23(6), 316–320. doi:10.12968/bjon.2014.23.6.316.

Chernomas, W. & Shapiro, C. (2013). Stress, Depression, and Anxiety among Undergraduate Nursing Students. International Journal of Nursing Education Scholarship, 10(1), 255-266. https://doi.org/10.1515/ijnes-2012-0032

Chernomas, Wanda & Shapiro, Carla. (2013). Stress, Depression, and Anxiety among Undergraduate Nursing Students. International journal of nursing education scholarship. 10. 10.1515/ijnes-2012-0032.

Cheung, Teris & Wong, Siu & Wong, Kit & Law, Lap & Ng, Karen & Tong, Man & Wong, Ka & Ng, Man & Yip, Paul. (2016). Depression, Anxiety and Symptoms of Stress among Baccalaureate Nursing Students in Hong Kong: A Cross-Sectional Study. International Journal of Environmental Research and Public Health. 13. 779. 10.3390/ijerph13080779.

Dawood, Eman & Mitsu, Rufa & Ghadeer, Hind & Alrabodh, Fatimah. (2017). Assessment of Depression and Its Contributing Factors among Undergraduate Nursing Students. International Journal of Nursing. 4. 10.15640/ijn.v4n2a6.

Fuad, Maher & Nasir, Balsam & Qader, Mohammed & Bakar, Mohammed & Elnajeh, Maged & Abdullah, Mohd. (2015). Prevalence and Risk Factors of Stress, Anxiety and Depression among Medical Students of a Private Medical University in Malaysia. Education in Medicine. 7. 52. 10.5959/eimj.v7i2.362.

https://www.who.int/health-topics/depression#tab=tab_1

Mofatteh M. (2020). Risk factors associated with stress, anxiety, and depression among university undergraduate students. AIMS public health, 8(1), 36–65. https://doi.org/10.3934/publichealth.2021004.

Puthran, R., Zhang, M. W. B., Tam, W. W., & Ho, R. C. (2016). Prevalence of depression amongst medical students: a meta-analysis. Medical Education, 50(4), 456–468. doi:10.1111/medu.12962.

Ramón-Arbués, E., Gea-Caballero, V., Granada-López, J. M., Juárez-Vela, R., Pellicer-García, B., & Antón-Solanas, I. (2020). The Prevalence of Depression, Anxiety and Stress and Their Associated Factors in College Students. International journal of environmental research and public health, 17(19), 7001. https://doi.org/10.3390/ijerph17197001.

Rathnayake, Sarath & Ekanayaka, Jeewanthika. (2016). Depression, Anxiety and Stress among Undergraduate Nursing Students in a Public University in Sri Lanka. International Journal of Caring Sciences. 09. 1020.

S.Deivasigamani, G.Narmadha, C.Senthilpari, Wong Hin Yong and P.K.Rajesh (2022), "Epileptic EEG signal classifications based on DT-CWT and SVM classifier", Journal of Engg. Research Vol.10 No. (2A) pp. 95-104, DOI: 10.36909/jer.10523

Shabbir MH, Bashir U (2016) Depression Among Medical Students. J Psychol Clin Psychiatry 6(5): 00371. DOI: 10.15406/jpcpy.2016.06.00371.

Tung YJ, Lo KKH, Ho RCM, Tam WSW. Prevalence of depression among nursing students: A systematic review and meta-analysis. Nurse Educ Today. 2018 Apr;63:119-129. doi: 10.1016/j.nedt.2018.01.009. Epub 2018 Feb 9. PMID: 29432998.