Comparative analysis of stress factors of humanities and technical specialities students

Oleh Pitsun[1)], Yaroslav Dykyi [1)], Maria Lysyk [1)], Anastasiia Sobchak [1)]

[1] West Ukrainian National University, 11 Lvivska st., Ternopil, 46001, Ukraine

Abstract

This article covers and substantiates the importance of using cluster analysis methods to identify groups of students with similar reactions to stress and to identify in detail the main factors that affect their psycho-emotional resilience. The authors of the study consider different approaches to cluster analysis, clustering algorithms and further interpretation of the results. This study shows that the use of cluster analysis makes it possible to effectively identify risk groups among students of WUNU and to identify specific stressors and their impact on the psychological comfort and academic performance of students, obtained results will serve as a basis for improving the process of students' adaptation to studying at a higher education institution.

Keywords

stress factors, clustering, statistical data.

1. Introduction

The problem of psychological research of stress factors as a determining factor in the adaptation of students to studying in higher education is of great relevance in modern conditions. The transience of public life and educational life involves the use of artificial intelligence technologies, namely clustering, to identify predictable and non-obvious root causes of stress among students of the West Ukrainian National University for the purpose of qualitative classification, distribution by specialty, age, gender, lifestyle, the balance of work and leisure time and a number of other equally important aspects. Previous scientific studies of the levels of educational losses and adaptation of students identify stress as the main catalyst for academic failure and low motivation among students. At the same time, the current dynamics of the stressors' growing in their environment causes a tendency to avoid social interaction and mental disorders, and in some cases is a prerequisite for the increasing cases of cardiovascular and pulmonary diseases.

The scientific problem of the study consists of two components:

a) the lack of interdisciplinary research that would combine scientific developments in the field of theoretical and practical psychology with artificial intelligence technologies to study stressors of university students;

b) the limited means and tools of the classical approach to studying the adaptive potential of students to study in higher education institutions actually make it impossible to identify non-obvious causal relationships that can be identified by the artificial intelligence clustering algorithm.

Out of the total number of students of the West Ukrainian National University, 200 respondents aged 17-25 were randomly selected from among the students of FCIT and SHF (technical and humanitarian specialities), of whom X were men, Y were women, and Z were people who classified themselves as non-binary or did not wish to indicate their gender. The questionnaire was developed on the basis of the following tests NEO-FFI-3 and Student Adjustment to College Questionnaire (SACQ) Baker & Siryk, 1984/1989, Dahmus & Bernardin, 1992, The questionnaire was formed on the basis of the approved tests NEO-FFI-3 and Student

The First International Workshop of Young Scientists on Artificial Intelligence for Sustainable Development; May 10-11, 2024, Ternopil, Ukraine
EMAIL: o.pitsun@wunu.edu.ua (O.Pitsun); dykyyyaroslav@gmail.com (Y.Dykyi); marialysyk01@gmail.com (M. Lysysk); a.sob4aak@gmail.com (A. Sobchak);
ORCID: 0000-0003-0280-8786 (O.Pitsun); 0009-0007-8510-0594 (Y.Dykyi); 0009-0000-2114-8272 (M. Lysysk) ; 0009-0001-8662-7001 (A. Sobchak) ;

Adjustment to College Questionnaire (SACQ) Baker & Siryk (1984/1989), Dahmus & Bernardin (1992), namely questions that allow to study the level of personal responsibility of the respondents and motivation to study, taking into account their attitude to the educational process, lifestyle, habits and a number of other subjective factors.

This work consists of the following sections: introduction, literature review, problem statement, dataset, materials and research methods, computer experiments and conclusions.

Object of study - statistical data of the results of the student survey to determine stress factors.

The subject of the study is data clustering algorithms.

The purpose of this article is to analyse the results of a survey of students of technical and humanitarian specialities based on clustering to identify stressors.

2. Literature review

In [1], the authors assess the reporting quality of cluster analysis in health psychology publications. The study concludes that clear guidelines are essential for improving the rigor and transparency of cluster analysis reporting in health psychology.

Students of university highlights the clustering of health-related lifestyles in [2]. In particular, the authors revealed distinct clusters among university students, categorizing their health-related lifestyles.

In work [3], involving 107 severe irritable bowel syndrome (IBS) patients resistant to conventional treatment, the authors employed a K means cluster analysis, which made it possible to reveal three subgroups with distinct characteristics.

In [4], a cluster analysis aimed to differentiate individuals based on coping profiles and examine their associations with perceived stress and health-related behaviors. The authors emphasized the importance of considering coping profiles for tailored interventions.

In the study [5], the authors investigated and confirmed the opinion that persons over 50 years of age are characterized by a stronger correlation between 4 behavioral risk factors. Therefore, an awareness of these clustered unhealthy behaviors enables more effective prevention and intervention measures targeting multiple behaviors simultaneously.

In [6], the authors examined the prevalence of lifestyle behaviors, identified sociodemographic and lifestyle profiles through cluster analysis. The analysis has shown the importance of understanding lifestyle patterns among a diverse population for effective health promotion strategies.

The study [7], identified the health profile of Hong Kong Chinese adults through a descriptive correlational survey involving 702 participants. Based on the conducted analysis, the authors suggested that nurses can utilize these identified health profiles to develop targeted interventions for improving the health of the population under study.

The challenge of monitoring and understanding soil state and change at a national scale by presenting a dataset of topsoil samples is addressed in work [8]. The analysis reveals consistent differences in soil properties across habitat types and identifies four soil functional classes.

In paper [9], the authors investigated gender inequalities in health among adolescents across 45 countries, analyzing data from the 2018 Health Behaviour in School-aged Children survey. The study reveals systematic gender inequalities, identifying seven distinct groups with similar patterns.

In [10], explores the application of data mining to uncover hidden patterns relevant to medical diagnoses and recovery processes for patients admitted due to substance abuse. Utilizing observational data and employing techniques such as K-means clustering and Hierarchical Clustering with Minitab, the authors identified correlations between factors, providing insights into health service databases.

Papers [11-12] provide the principles for developing software for the tasks of processing a large amount of data using artificial intelligence. The principles of data processing with the achievement of artificial intelligence with the use of wasteful means are considered in the work

[13-14]. Neural networks are used in many directions, in particular, one of them is considered in the work [15].

.

3. Problem statement

To conduct the research, the following tasks need to be implemented:
- create the structure of the questionnaire file. The survey includes elements of both psychological and physiological state;
- make a statistical analysis of the research data. Basic elements of data visualization in various formats are used;
- conduct a cluster analysis of the research results to form patterns. Clustering algorithms were used to search for interdependencies.

4. Dataset

There is a small number of available datasets on the Internet for researching the psycho-emotional state of students. Most of the available datasets have a limited set of items (characteristics) and were conducted for foreign students. The purpose of this paper is to develop our dataset based on the indicators of Ukrainian students, in particular, students of the West Ukrainian National University. The developed dataset "WUNUTH-24" was formed on basis of a technical and humanitarian specialities students' survey to cover students with different specifics of educational and cognitive activities. The dataset consists of about 200 records and is being updated. The gender distribution of the sample is shown in Figure 11.

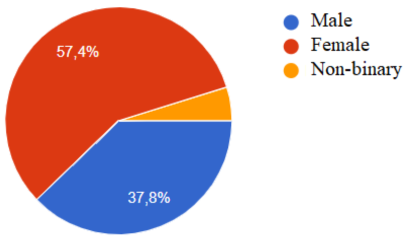


Figure 1: Gender distribution of the sample

The number of students of technical and humanitarian faculties is almost equally distributed.

5. Materials and research methods

Thus, it is an unsupervised clustering we do not have a tagged function to evaluate or score our model. The purpose of this section is to study the patterns in the generated clusters and determine the nature of the cluster structures.

The problem with supervised learning is that we provide our model with the initial labels to learn from, i.e. we have both X and Y for the entire training set and the model can see them while learning.

Unsupervised learning is when our model only sees X and not Y during training. The model doesn't know what the actual ground truth labels are for the training set and has to make inferences based on X alone.

Data clustering involves dividing a set of objects into subsets called clusters. The generalised algorithm consists of the following steps.

1) Let's represent the characteristics in the form of a matrix $X_{n\times m}$, where m – number of features, and n –values of features.

2) We use the principal components method, which results in the following main components $y_j = w_{1j}x_1 + w_{2j}x_2 + ... + w_{pj}x_p$, where w_{ij} – characteristics of the load factor. After data reduction, we get many most informative functions p : $X = \{X_1, X_2, ..., X_p\}$, where $p < m$.

3) To determine the number of clusters, we use the Elbow method, which is based on the calculation of the minimum summed square error (MSSE)

$$K = \min_{c_1,...,c_k} \sum_{i=1}^{n} \min \left\{ \|x_i - c_1\|^2, ..., \|x_i - c_k\|^2 \right\}.$$

4) Using the K-means method, we find the centers of the clusters

$$C_k = \frac{\sum_{x_i \in C_k} x_i}{|c_k|}.$$

5) We estimate the clustering accuracy on the basis of the Silhouette Coefficient parameter,

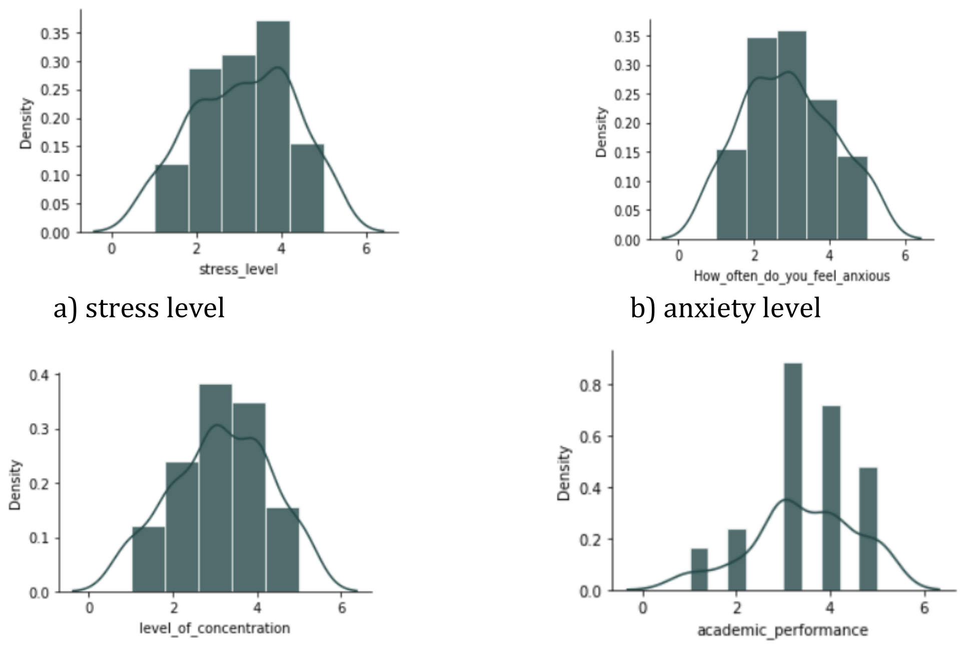
$$s = \frac{b-a}{\max(a,b)},$$

where a is the average distance between element a and all points in the same class

b - the average distance between element a and all other points in the neighbouring cluster.

6. Computer experiments

The visualisation of statistical indicators for the distribution by the criteria of "stress level", "anxiety", "concentration", "academic performance" for FCIT students is shown in Figure 2.



a) stress level

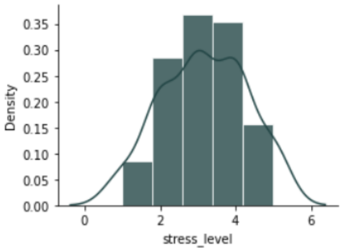
b) anxiety level

c) concentration level

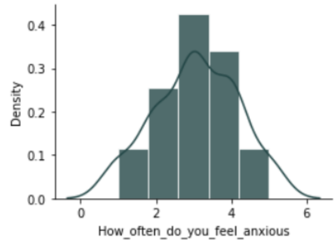
d) academic performance level

Figure 2: Visualisation of statistical indicators for FCIT students

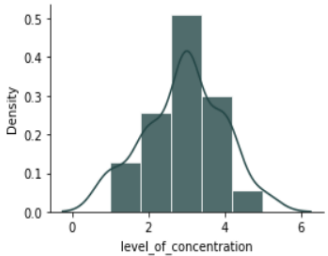
The visualisation of statistical indicators for the distribution by the criteria of "stress level", "anxiety", "concentration", "academic performance" for students of the SHF is shown in Figure 3.



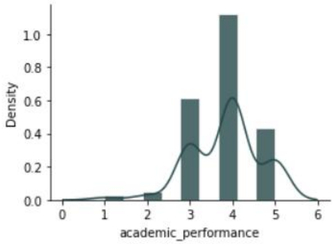
a) stress level



b) anxiety level



c) concentration level



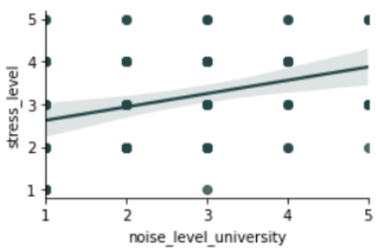
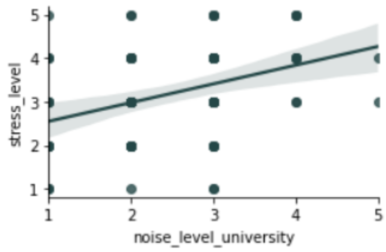
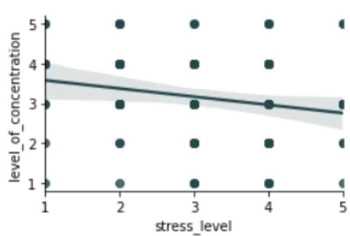
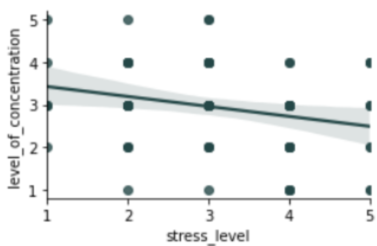
d) academic performance level

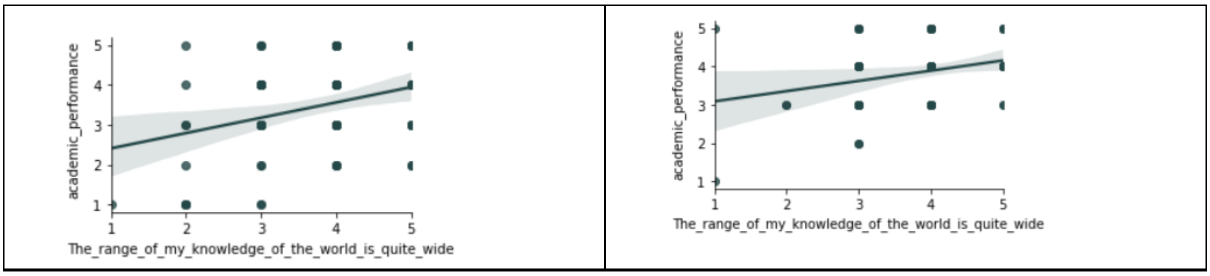
Figure 3: Visualisation of statistical indicators for students of the SHF

Analysing the results from Figures 2 and 3, it can be concluded that the indicators of technical and humanities faculties are similar in general. A more detailed analysis reveals that students of the technical faculty are characterised by a higher level of stress and anxiety. The range of values is from 0 to 5.

A comparative analysis of statistical indicators between students of technical and humanitarian specialities is presented in Table 1.

Table 1
Comparative analysis of statistical indicators between students of technical and humanitarian specialities

FCIT	SHF
	
	

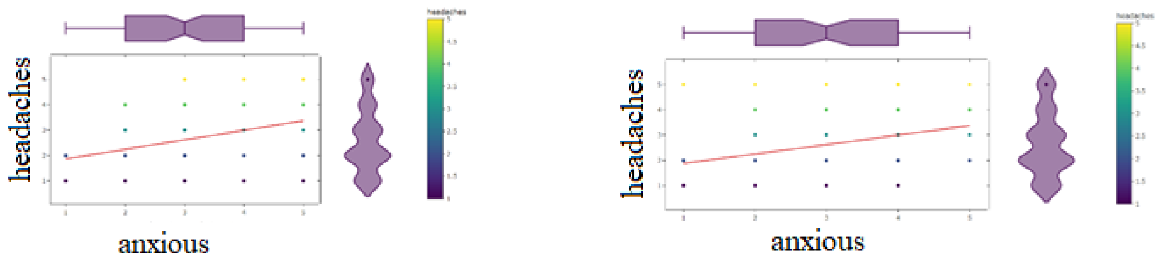


The comparisons were made by the ratio of the indicators "stress level - noise level", and "stress level - level of student concentration".

Based on the indicators, it can be concluded that, regardless of the type of speciality, with an increase in noise level, the level of stress increases, and with an increase in stress level, the level of concentration decreases.

Additionally, students were asked to answer the questions "The spectrum of my knowledge of the world is quite wide" and "My goals and achievements require a lot of hard work." The analysis shows that students with a wide range of knowledge have better academic performance in the learning process.

A visual comparison of the dependence of the "headache-anxiety level" indicators for the faculties of FCIT and SHF is shown in Figure 4.

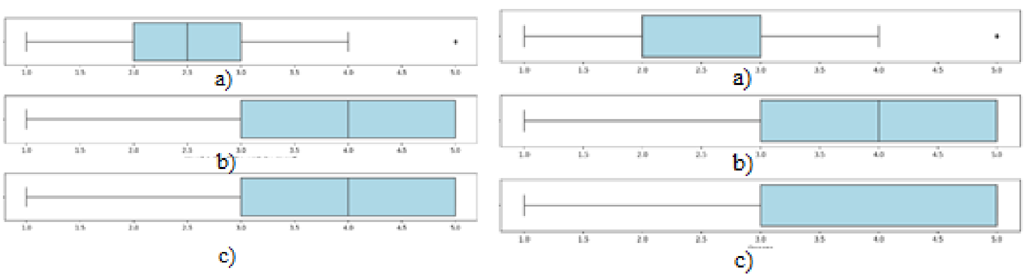


a) FCIT b) SHF

Figure 4: Visual comparison of the relationship between headache and anxiety levels for the faculties of FCIT and SHF

The obtained indicators are almost identical for both faculties

The results of the survey on the criteria "I often want someone else to solve my problems", "I usually do all my work on my own", "I donate to meetings and charities" for both faculties are shown in Figure 6.



FCIT SHF

Figure 5: Comparative analysis of indicators by the criteria "a) I often want someone else to solve my problems", "b) I usually do all my work on my own", "c) I donate to fees and charities"

The indicators are identical, which indicates a high level of self-organisation and responsibility of students.

The silhouette score for a set of sample data points is used to measure how dense and well-separated the clusters are. The silhouette score takes into account the intra-cluster distance between the sample and other data points within the same cluster. A visual representation of the silhouette scores for the 4 clusters is shown in Figure 6.

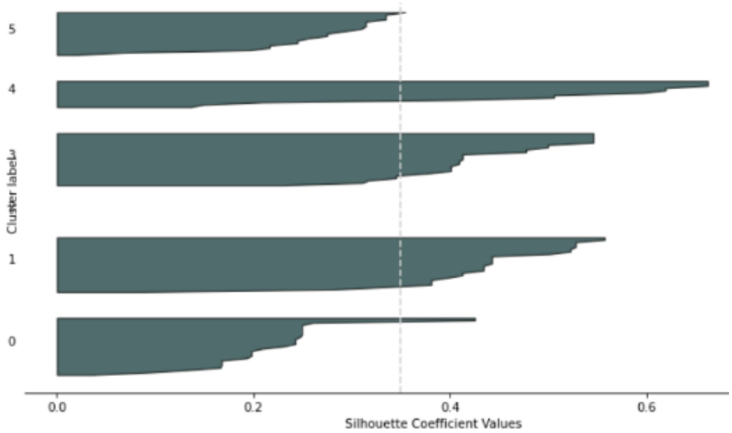


Figure 6: Silhouette coefficient

A three-dimensional visualisation of the 6 clusters is shown in Figure 7.

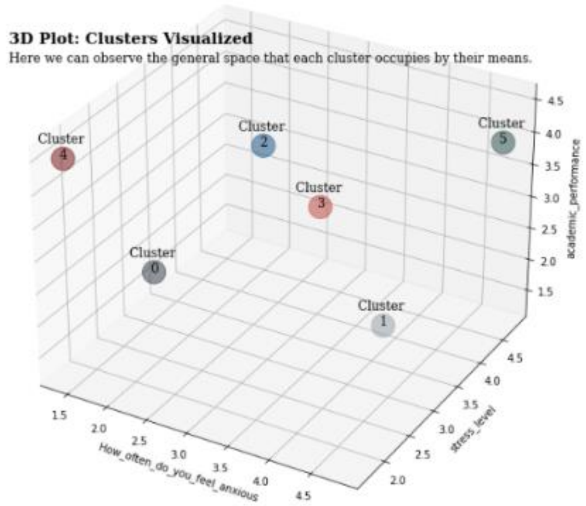


Figure 7: Three-dimensional visualisation of the 6 clusters

Cluster 0 - low anxiety, medium stress, low academic performance

Cluster 1 - higher than average level of anxiety, average level of stress, low academic performance

Cluster 2 - low anxiety, high stress, average academic performance

Cluster 3 - medium level of anxiety, high level of stress, below average academic performance

Cluster 4 - low anxiety, below average stress, above average academic performance

Cluster 5 - high level of anxiety, high level of stress. High level of academic performance

Consider an example of clustering based on 4 classes.

The correlation matrix for the parameters "stress level", "sleep quality", "concentration level", and "academic performance" is shown in Figure 8.

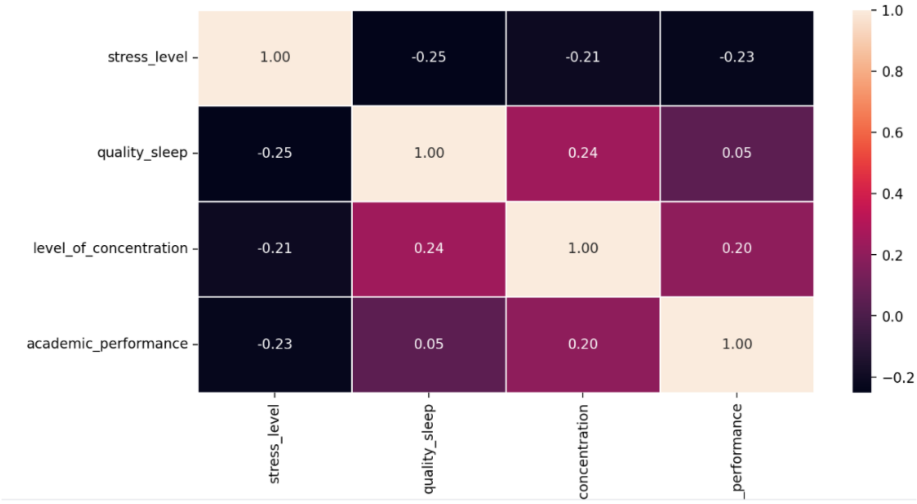


Figure 8: Correlation matrix

The results of the distribution of clustering results between clusters are shown in Table 2.

Table 2
Results of the distribution of clustering results between

stress_level	quality_sleep	level_of_concentration	academic_performance	Cluster
4	2	2	4	3
1	4	4	3	0
3	5	4	5	0
3	3	4	4	1
2	1	2	5	1
...	...	...	...	...
5	2	3	3	3
4	3	3	3	3
3	3	1	1	3
3	1	2	1	3
2	1	4	5	1

The distribution of records between clusters is as follows:
Cluster 3 - 45 items
Cluster 1 - 28 items
Cluster 0 - 20 items
Cluster 2 - 11 items
The Min and Max of stress_level in Cluster = 0 : 1 and 3
The Min and Max of stress_level in Cluster = 1 : 1 and 5
The Min and Max of stress_level in Cluster = 2 : 3 and 5
The Min and Max of stress_level in Cluster = 3 : 2 and 5

Clusters are characterised by the following parameters:
Cluster 0 - high sleep quality, high concentration level, low-stress level, average to above average academic performance
Cluster 1 - average level of sleep and concentration, below average level of stress, above average level of academic performance
Cluster 2 - below average sleep and concentration, average academic performance
Cluster 3 - High levels of stress, average or poor academic performance, below average sleep.

7. Conclusions

Based on the analytical approach, the article analyses modern approaches to data analysis using artificial intelligence, clustering, in particular.

The authors developed their own dataset structure and conducted research among students of technical and humanitarian specialities at the West Ukrainian National University.

The statistical analysis of the data shows that the level of stress depends on the academic success of students, and the level of anxiety and stress among students of both specialities is also determined. The results demonstrate that students of technical specialities have a higher level of stress, however this indicator does not differ significantly from representatives of humanitarian specialities.

It was found that the optimal number of clusters for analysis is 4-6.

As a result of the use of three-dimensional clustering, it can be concluded that with an increase in the level of stress, students' academic success decreases. However, the exception is the case when the level of stress and anxiety is high, in this case academic performance will be at a high level.

Taking into account the connection between stress level, sleep quality, level of concentration and academic success, we came to the conclusion that as the quality of sleep increases and the level of stress decreases, the level of students' academic performance improves.

The use of clustering allows to determine the optimal parameters for the analysis of psychological indicators of students, which characterizes the theoretical significance of the study. The practical significance of the study lies in the possibility of determining the level of stress of students based on the use of modern libraries and the approach used in this study.

Future work will be devoted to the development of a decision support system based on the results of the clustering of student survey indicators.

References

[1] Chaudhari, Aayushi, Chintan Bhatt, Achyut Krishna, and Pier Luigi Mazzeo. 2022. "ViTFER: Facial Emotion Recognition with Vision Transformers" Applied System Innovation 5, no. 4: 80. https://doi.org/10.3390/asi5040080

[2] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, “The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression,” in 2010 ieee computer society conference on computer vision and pattern recognition-workshops, pp. 94–101, IEEE, 2010

[3] Magdin, Martin, Ľubomír Benko, and Štefan Koprda. 2019. "A Case Study of Facial Emotion Classification Using Affdex" Sensors 19, no. 9: 2140. https://doi.org/10.3390/s19092140

[4] Minaee, Shervin, Mehdi Minaei, and Amirali Abdolrashidi. 2021. "Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network" *Sensors* 21, no. 9: 3046. https://doi.org/10.3390/s21093046

[5] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops, San Francisco, CA, USA, 2010, pp. 94-101, doi: 10.1109/CVPRW.2010.5543262.

[6] DANDIL, Emre, and Rıdvan Özdemir. "Real-time Facial Emotion Classification Using Deep Learning Article Sidebar." International Journal of Data Science and Applications 2, no. 1 (2019): 13-17.

[7] M. Moolchandani, S. Dwivedi, S. Nigam and K. Gupta, "A survey on: Facial Emotion Recognition and Classification," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 1677-1686, doi: 10.1109/ICCMC51019.2021.9418349.

[8] Hu, Tianming, Liyanage C. De Silva, and Kuntal Sengupta. "A hybrid approach of NN and HMM for facial emotion classification." Pattern Recognition Letters 23, no. 11 (2002): 1303-1310.

[9] Khaireddin, Yousif, and Zhuofa Chen. "Facial emotion recognition: State of the art performance on FER2013." arXiv preprint arXiv:2105.03588 (2021).
[10] Duncan, Dan, Gautam Shine, and Chris English. "Facial emotion recognition in real time." Computer Science (2016): 1-7.
[11] Berezsky, O., Pitsun, O., Melnyk, G., Koval, V., Batko, Y. (2023). Multi-threaded Parallelization of Automatic Immunohistochemical Image Segmentation. In: Hu, Z., Wang, Y., He, M. (eds) Advances in Intelligent Systems, Computer Science and Digital Economics IV. CSDEIS 2022. Lecture Notes on Data Engineering and Communications Technologies, vol 158. Springer, Cham. https://doi.org/10.1007/978-3-031-24475-9_23
[12] Berezsky, O., Liashchynskyi, P., Pitsun, O., Liashchynskyi, P., Berezkyy, M. Comparison of Deep Neural Network Learning Algorithms for Biomedical Image Processing. CEUR Workshop Proceeding, 2022, 3302, pp. 135–145. https://ceur-ws.org/Vol-3302/paper7.pdf
[13] R. M. Wallace et al., "Applications of neural-based spot market prediction for cloud computing," 2013 IEEE 7th International Conference on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), Berlin, Germany, 2013, pp. 710-716, doi: 10.1109/IDAACS.2013.6663017
[14] M. Dyvak, A. Pukas and O. Kozak, "Tolerance estimation of parameters set of models created on experimental data," 2008 International Conference on "Modern Problems of Radio Engineering, Telecommunications and Computer Science" (TCSET), Lviv, UKraine, 2008, pp. 24-26.
[15] A. Sachenko et al., "Sensor errors prediction using neural networks", Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium, vol. 4, pp. 441-446, 2000.