WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Pandemic Pulse: Orchestrating Student Well-being with SVM-Enhanced Mental Health Forecasting

Rakesh Shankar Ghosh[1], Shovon Mandal[2], Md. Jakaria[3], Mahmuda Akter Keya[4]

[1,2]Postgraduate Scholar, Department of Computer Science and Engineering, Khulna University of Engineering & Technology, Khulna-9203, Bangladesh
[3]Undergraduate Scholar, Department of Computer Science and Engineering, Imperial college of Engineering, Jessore Road, Boikali, Khulna, Bangladesh
[4]Undergraduate Scholar, Department of Computer Science and Engineering, BRAC University, 66 Mohakhal, Dhaka-1212, Bangladesh
DOI: https://doi.org/10.55248/gengpi.4.1123.113122

A B S T R A C T

Amidst the pervasive impact of the COVID-19 pandemic, student mental health stands as a paramount public health concern. This vulnerability is exacerbated by the hectic lifestyles of students, rendering them particularly susceptible to the implications of social distancing measures. The primary objective of this research is to assess the self-reported levels of stress, anxiety, and depression among Serbian students. This evaluation encompasses a comprehensive examination of demographic traits, living and learning environments, students' activities during the epidemic, potential exposure to the coronavirus, and overall mental and physical health. The core aim of this article is to prognosticate the mental health of students under the unprecedented circumstances posed by the pandemic. Leveraging the power of PandemicPulse, this SVM-enhanced forecasting model, this research meticulously analyzes a dataset obtained from a dedicated student mental health survey website. The application of appropriate classifiers allows us to draw meaningful insights and predictions regarding mental health outcomes. In this discourse, this study unveils a multifaceted approach, employing multiple classifier strategies to ensure the highest accuracy in forecasting student mental health. This research endeavors to harmonize understanding, utilizing PandemicPulse as a guiding rhythm to navigate the complex landscape of student well-being during these challenging times.

Keywords: Pandemic, student mental health, academic stress, machine learning techniques, SVM.

1. INTRODUCTION

Amid the rapid spread of the Covid-19 outbreak, nations globally implemented preventive measures, including complete lockdowns, leading to a significant surge in anxiety levels [1]. Originating in Wuhan, China, the virus swiftly transcended borders, with the World Health Organization officially declaring it a pandemic on January 30, 2020. Governments responded with stringent measures, disrupting daily life and affecting culture, economy, and ecology. This study delves into the profound impact of the pandemic on mental health, particularly among students facing disruptions in academic and extracurricular activities due to ongoing lockdowns. The extended home confinement, school closures, and social exclusion have raised concerns about children's mental well-being globally [3]. The author aims to anticipate these effects using student survey data, employing machine learning (ML), artificial neural networks (ANN), and explainable AI (XAI) techniques. In navigating the complexities of the pandemic's impact, the study employs SVM, a supervised machine learning algorithm known for its versatility in handling both linear and non-linear issues. SVM's premise involves creating a hyperplane to accurately classify data points, addressing challenges in real-world scenarios through parameters such as kernel, margin, regularization, and gamma. This choice enhances the reliability and accuracy of the model, making it a robust tool for forecasting and interpreting students' mental health during these challenging times.

2. LITERATURE REVIEW

Researchers and scientists from all over the world were working around the clock to determine the effects of this pandemic and potential treatments. There are many different kinds of research studies that have been done on associated subjects; some are based on statistics and some on machine learning. For a hypothetical example, Tor ales et al. noted how this pandemic is causing substantial health difficulties globally, such as worry, uneasiness, depressive symptoms, exhaustion, denial, wrath, and dread on a daily basis, leaving us vulnerable to coping with the unknown truth for leading this miserable life in lockdown. [4]. In the past few years, there was an increasing of interest in machine learning algorithms for predicting anxiety disorders. These prediction models are interesting to decision-makers because they can foresee the outcomes of various acts. These tools are valuable for predicting the consequences of public mental health crises and understanding the dynamics of the components involved. Panatelas and colleagues [5]. A recent systematic evaluation of the effects of the COVID-19 lockdown on child and adolescent mental health outcomes found a considerable rise in the symptoms of despair and anxiety as well as loneliness, anger, and impatience. It also emphasized how extremely vulnerable persons with pre-existing mental health issues and/or

developmental abnormalities are to experiencing more stress [6]. The educational system has progressed online in this epidemic, causing kids to deal with self-study, poor communication with professors, and a reduction in enthusiasm to learn. Thus, it increased their tension, which was daily leading to major mental health concerns. There were excellent research efforts in this field that are ML based, therefore machine learning researchers were not falling behind. For instance, a team of researchers from MIT and Harvard University used the ML approach to find changes in their discussion and an increase in fear as well as strong suicidal tendencies during the pandemic [7]. They have analyzed machine learning prediction methods for anxiety disorders in depth. They came to the conclusion that the effectiveness of these investigations depends on the type of prediction approaches and data collections, such as clinical data or self- or screening tools. Out of the 16-research looked at, they discovered that Hybrid approaches and Support Vector Machine (SVM) were most frequently employed to predict post-traumatic stress disorder (PTSD) and seasonal affective disorder (SAD), respectively. The highest prediction scores are also attained by ensemble methods and artificial neural networks (ANNs) [8]. Similarly, a secondary study on the psychological health of Chinese children aged 7 to 15 was conducted in 668 families from different provinces of China during the duration of COVID-19. Public schools and the provincial capital were found to be the crucial factors associated with treating PTSD, with most people demonstrating a more favorable attitude toward online learning. Multiple logistic regression analysis was utilized in the statistical studies to investigate the elements influencing Chinese children's support and mental health. (WHO, 2020). Due to the huge increase in processing power in recent years, modern estimating methodologies also enabled the analysis of design factors as they emerge and continual updating of their estimations [9]. As predicted mood states, this study considered positive changes in the child's life caused by the COVID-19 pandemic, relationships between family members, time spent watching television, mental health assessment and diagnosis, and resulting stress. on restrictive measures. Although previous work has contributed to understanding the impact of the COVID-19 pandemic on children's emotional responses, research in this area has primarily used cross-sectional designs, general and at-risk samples and traditional inferential statistical analysis. Assumptions do not limit the SVM approach compared to traditional machine learning models and help find patterns in complex data sets [10] [11] samples. One in four children and adolescents were estimated to have these conditions exhibiting symptoms of clinically high despair, and 1 in 5 young people exhibiting symptoms of clinically elevated anxiety. However, because the data are cross-sectional, it is impossible to determine how the clinically high symptoms change over time, such as whether they persist, worsen, or another recent study they conducted looked into how COVID-19 affected people's mental health, using a variety of machine learning prediction models. From every country in the world, as the crisis has led to a number of problems such as stress, anxiety, and unhappiness. Anxiety and depression among children and adolescents doubled during the first year of the COVID-19 pandemic compared to pre-pandemic estimates, according to a recent meta-analysis of 29 studies using general population go away. Clinical samples from cross-sectional research looking at the effects of COVID-19 lockout on mental health outcomes revealed that some symptoms rose while others reduced or stayed largely stable [12]. Similarly, Herbert et al. performed an online survey with students from Egypt and Germany to conduct psychological circumstances, individual perspectives, and behavior using machine learning approaches that aided in predicting the students' personalities [13]. According to a recently released study in a diverse clinical population, 77% of children's emotional and behavioral responses during a six-week lockout were rated by parents, and machine learning techniques were used to predict mood states managed to transform their mood throughout the lockdown from a negative one to one that was stable. Machine learning technology can be used to create an efficient and effective model to predict anxiety and depression from different predictor variables. Their regular daily routine of going to work, spending time with friends or coworkers, and meeting people in public places—things that bring them joy and keep them entertained—has been put on hold. Many people had been frightened by this unexpected arrangement, which also caused corona phobia. Machine learning for psychological problem prediction in Indians during COVID-19 nationwide lockdown can raise anguish, symptoms, and relapse into mental health. COVID-19 can increase one's "normal" levels of stress and worry. Additionally, stigmatization of those who are afflicted, authorities, and medical personnel results from epidemics. Various news outlets have carried reports of landlords threatening and shunning medical personnel doctor, nurse, and paramedics caring for COVID-19 patients as well as flight crew. Additionally, a person's fear of contracting a disease may cause them to develop an obsession with hand hygiene and an accompanying need to wash or clean, which will exacerbate their obsessive-compulsive disorder. The psychological impacts can cause a variety of reactions, including fear, anxiety, boredom, melancholy, rage, and revulsion. They can also cause feelings of stigma and wrath. While maintaining mental health is important, it is also crucial to seek professional assistance if symptoms of this Corona-related "Coro-anxiety" interfere with daily activities or interpersonal or career connections [14]. Here this study article explains mental health analysis in words that varied target audiences may comprehend. They found a direction for measuring a person's mental health problem, and prediction models were built utilizing this framework of the machine learning algorithm. Prior to developing models, grouping procedures were used to determine the number of structures. They were used to appraise the created majority class before being prepared for the training of the classifier. KNN, SVM, and Random Forest are exceptionally almost equally well in the experimentations. Group classifiers were also shown to significantly enhance mental health prediction skills, with a 90% high accuracy [15]. They employed discourse analysis throughout their work to better understand representation patterns in human-centered machine learning (HCML). They located 55 collaborative studies on case prediction of disorders of mental health using massive amounts of data. Their analysis revealed that conflicting interaction discourses may emerge throughout the dataset, giving individuals agency. Their findings show how the five meanings produce a contradictory object and subject perceptions in kids, sometimes inadvertently affecting them [16]. Other models were also produced on a regular basis that predicts and categorized future occurrences using machine learning methodologies such as support vector machines, random forests, and artificial neural networks. [17]. Studies on the effects of the pandemic and associated issues on student mental health as a result, the goal of this project is to employ a carefully designed machine learning method, using SVM, to predict anxiety using real-world data from COVID-19 and to investigate how the lockdown affects students' mental health, social lives, and academic performance. Moreover, feature selection was carried out systematically, providing the top 4 features with the maximum accuracy out of the 19 available characteristics.

3. METHODOLOGY

In this study, introduced Support Vector Machines (SVM) as primary machine learning algorithm to analyze and predict the impact of the COVID-19 pandemic on student mental health. The chosen model, named PandemicPulse, utilizes SVM's capabilities in handling both linear and non-linear issues, providing a robust approach for forecasting mental health outcomes. The dataset, obtained from the 'Kaggle' website (COVID-19 and its impact on education, social life, and mental health of students: A Survey), comprises 1182 student opinions on 19 different features representing various survey elements. The dataset is divided into training (80%) and testing (20%) sets for a comprehensive analysis. Feature selection was systematically conducted, and the top 4 features, out of the 19 available, were identified for optimal accuracy in predicting anxiety. PandemicPulse, powered by SVM, aims to provide insights into the complex landscape of students' mental well-being during the ongoing pandemic.

4. MODELING AND ANALYSIS

This methodology was selected due to the absence of a comprehensive literature review on the use of machine learning algorithms for predicting students' mental health during the COVID-19 pandemic. The proposed project is structured into six sections. Initially, the original data, consisting of 1182 rows, a header, and 19 features, was loaded as a CSV file. Subsequent preprocessing involved handling missing data using NaN (Not a Number) deletion as part of data preparation. Following this, the data underwent encoding from categorical to numerical, with a significant portion involving text-to-number conversion through natural language processing. Non-numerical data, which cannot be processed by the algorithms, often contains one or more missing columns or values. Missing values can arise due to various reasons, such as inaccurate measurements or unavailability, making it essential to address them for machine learning algorithms, which necessitate numerical input values. This process is commonly referred to as the imputation of missing data. To address missing data, here employed the K-Nearest Neighbor (KNN) model, specifically known as "Nearest Neighbor Imputation" or "KNN imputation." This method involves selecting a distance measure and determining the number of contributing neighbors for each prediction, denoted by the KNN algorithm's k variable. The KNN classifier utilizes a majority voting principle to identify the class of a data file. For predicting missing values, here opted for five KNN values: K=1, K=2, K=3, K=4, and K=5. After labeling the data, a standard scalar was used to scale independent variables to unit variance and eliminate mean values. This formatted the data such that Logistic Regression and KNN (K-Nearest Neighbor models could more easily understand it. To complete the data scalability, real-valued input and output variables can be normalized or standardized. Here used standardization and normalization to enhance the effectiveness of a Multilayer Perceptron model in a predictive modeling issue. The aspects that relied on "yes" or "no" replies were determined to be changed to numerical values, with 1 signifying "yes" and 0 indicating "no".

After doing Split test procedure here divided the dataset model into two part 20% of the data are for testing, and 80% are for training. As a necessary consequence, the train dataset and the test dataset are the two key concepts in machine learning, where the training dataset is used to fit the model and the test dataset is used to validate the model. Training data is a segment of the existing data used to train the machine learning model, whereas testing data is used to validate the model's correctness. The training dataset is often bigger than the testing dataset. The majority of ML models are believed to perform best when trained using supervised learning techniques like classifier algorithms. In addition, K-Nearest Neighbor (KNN) model, a tried-and-true method with several applications, was also used to train data. Then, here assessed the accuracy of each of the models. Implementing proper machine learning (ML), K- Nearest Neighbor (KNN) model, and a survey, the dataset can be used to summarize the methodology. Here used SVM to increase the study's interpretability. Support Vector Machine (SVM). Here used SVM Machine Learning Algorithm to make the model more explanatory. Support Vector Machine is a supervised machine learning algorithm used in classification and regression problems. Here used SVM, linear model to solve classification and regression issues. A powerful supervised linear machine learning approach is Support Vector Machine (SVM). Here choose SVM because of its exceptional capacity to handle both linear and nonlinear data and support a wide range of kernel functions. Another reason here utilize SVMs is that they may identify significant relationships between your data without requiring you to make extensive changes to your own data. Because of their strong tendency for working with tiny, complicated datasets, they usually generate more accurate readings than other algorithms. However, SVM major benefit is its capacity to avoid the dimensionality problem and perform well with little data by applying a generalization control technique. Both classification and regression issues can be solved with SVM. at first, here train a support vector machine and then cross-validate the classifier, just like with any supervised learning model. After That Classify (predict) new data using the trained machine. Additionally, here utilize different SVM kernel functions and fine-tune their parameters to get enough predicted accuracy. Applications for binary classification were primarily used. Where it analyses the training examples and picks a hyperplane to split the two groups. To generate an ideal hyperplane, the distance between the support vectors and the hyperplane must be maximized. The margin is also called street width. The distance from a point. Through the kernel trick, which involves maximizing the distance between support vectors, here intuitively aim to maximize the margin or street width. When systematically selecting input values from a permissible set and evaluating the function, an optimization problem typically involves either maximizing or minimizing a real function. In that case, here seek to achieve the same objective, ensuring that the support vectors remain distant from the street, not residing on it or in its middle. Consequently, here can categorize this as a common constrained optimization problem or scenario. The resolution to such problems often involves the utilization of Lagrangian multipliers or the Lagrangian formula. Before employing Lagrange multipliers to solve the linearly separable plane, it's essential to establish the margin or decision boundary. The decision boundary must accurately classify all points. Solving the ensuing constrained optimization problem helps determine this decision boundary. The optimization problem is constrained, requiring the use of Lagrange multipliers. The Lagrangian is subject to certain conditions, including $a \geqslant 0$ and $||w||^2 = w^T w$. The gradient with respect to w and b is calculated, and the Lagrange multipliers are employed to solve the optimization problem. The mathematical equation for the hyperplane involves the weight vector (w), a value from the set of labeled training pairs (x), and the bias (b). To identify the best hyperplane and ensure generalization control, minimizing the weight vector is necessary. The mathematical formula utilizes Lagrangian duality theory, providing flexibility to the function and determining the best hyperplane. During data preprocessing, certain values in

the dataset were identified, and not-answered (N/A) values were replaced with NAN. Attributes dependent on "yes" or "no" responses were converted to numerical values, with 1 representing "yes" and 0 indicating "no."

5. RESULTS AND PERFORMANCE EVALUATION

By employing the SVM classifier, this prediction accuracy reached 88.00%, as indicated in the table detailing the SVM model analysis. The F1 score, recall, and accuracy metrics stand at 0.88, 1.00, and 0.94, respectively.

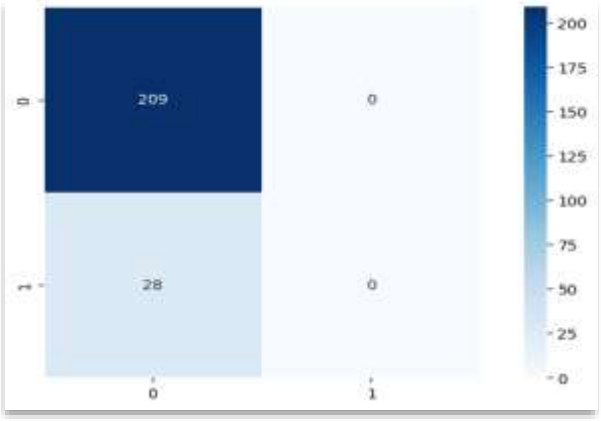

Fig.1: Confusion Matrix of Pandemic Pulse

Classification Report

	precision	recall	f1-score	support
0.0	0.88	1.00	0.94	209
1.0	0.00	0.00	0.00	28
accuracy			0.88	237
macro avg	0.44	0.50	0.47	237
weighted avg	0.78	0.88	0.83	237

Fig.2: Classification Output

From the fig.2. out of 237 values, 209 were able to predict and the remaining 28 were not predict correctly. 209 = TP, 0 = FP; 28 = FN, 0 = TN;

Given that the model accurately predicted all data in class, the precision, recall, and F1 score support are 0.88, 1.00, and 0.94, respectively. To calculate the recall, both the macro average and weighted average are computed Eqn:

(1) Macro average (recall) = (R1 + R2) / 2 = (1.00 + 0.00) / 2 = 0.50

(2) Weighted average (recall) = ((R1 x s1) + (R2 x s2)) / (s1 + s2) = ((1.00 x 209) + (0.00 x 28)) / (209 + 28) = 0.88

These calculations apply to matrices, allowing a comprehensive understanding of the classification report and confusion matrix for any number of classes.

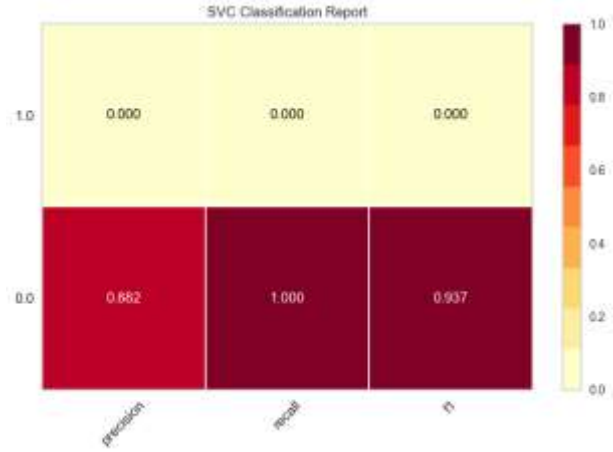

Fig.3: SVC Classification

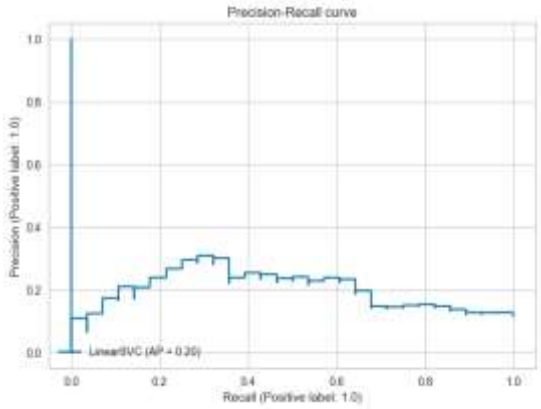

Fig.4: Precision-Recall Curve

The SVC Classification Report indicates precise predictions by the model, with precision at 0.88, recall at 1.00, and a F1 score support of 0.93. To assess accuracy and recall, here employed two key metrics. The confusion matrix, which efficiently gauges precision and recall based on predicted and actual labels from a model. In a binary classification confusion matrix, the four potential outcomes—true positive, false positive, true negative, and false negative—are displayed. Each outcome is denoted at the intersection of the matrix's rows and columns. For instance, if a data point is predicted to be positive but is actually negative, it constitutes a false positive.

6. DISCUSSION

In this study, an explainable machine learning (ML) approach was adopted to identify student mental health. The goal of this study was to learn more about the consequences of the COVID-19 epidemic on student mental health. Data collection includes 1182 student opinions on 19 different features

representing various components of the survey. It was gathered through the 'Kaggle' website (COVID-19 and its impact on education, social life, and mental health of students: A Survey) and the survey analysis focused on the students' impacts on mental health prediction. The findings revealed that, of the initial 19 identified features, the change in perceived stress derived from the restriction measures, the change in perceived positive changes in students' lives during the COVID-19, and the change in time spent on social media were the most significant contributors to mood state change. Four features were selected as significant out of a total of 19, with the SVM classifier reaching 88% accuracy.

7. CONCLUSION

Several years ago, the world was engulfed by the devastating pandemic, COVID-19, which disrupted our way of life in profound ways. It had far-reaching effects on various aspects of our daily existence, encompassing the economy, health, education, and social interactions. This research centers on the student population, specifically exploring how this global crisis has impacted their mental well-being during this tumultuous period. Students, grappling with academic pressures, limited interaction with educators, prolonged confinement at home, interpersonal challenges, and feelings of isolation, are undergoing a challenging phase during this unsettling time. Many of them are grappling with significant psychological consequences, manifesting in issues such as sadness, anxiety, and insomnia. The COVID-19 outbreak has undeniably left an indelible mark on our lives, with many struggling to adapt to the new normal. This dire situation has led to profound feelings of dejection for some. To enhance the comprehensibility of the model, SVM was utilized in this study.

References

[1] Y. Bar, I. Diamant, L. Wolf and H. Greenspan, “Deep learning with non-medical training used for chest pathology identification,” *Proceedings of Society for Optics and Photonics*, pp. 94140V-94140V. doi: 10.1117/12.2083124, 2015.

[2] C. Ntakolia, D. Prifits and I. Rannou, "An Explainable Machine Learning Approach for COVID-19’s Impact on Mood States of Children and Adolescents during the First Lockdown in Greece," 13 January 2022.

[3] A. U. Hussna, I. I. Trisha, I. J. Ritun and M. G. R. Alam, "COVID-19 impact on students' Mental Health: Explainable AI and Classifiers," International Conference on Decision Aid Sciences and Application (DASA), pp. 847-851, 2021.

[4] X. Luo and Y. Huang, "Mental Health Identification of Children and Young Adults in a Pandemic Using Machine Learning Classifiers," *Frontiersin*, pp. 847-851, 2021.

[5] J. Torales, M. O’Higgins and A. Ventriglio, "The outbreak of COVID-19 coronavirus and its impact on global mental health," *International Journal of Social Psychiatry*, vol. 66, no. 4, p. 317– 320, 2020.

[6] E. G. Pintelas, T. Kotsilieris, I. Livieris and P. E. Pintelas, "A review of machine learning prediction methods for anxiety disorders," pp. 8-15, 2018.

[7] U. Panchal, G. S. d. Pablo, M. Franco, C. Moreno, M. Parellada, C. Arango and P. Fusar-Poli , "The impact of COVID-19 lockdown on child and adolescent mental health: systematic review," *Eur Child Adolesc Psychiatry*, 2021.

[8] "Using machine learning to track the pandemic’s impact on mental health," 5 November 2020.

[9] E. A. Boeke, A. J. Holmes and E. A. Phelps, "Toward Robust Anxiety Biomarkers: A Machine Learning Approach in a Large-Scale Sample," no. 8, pp.

[10] M. A. Abas, B. Weobong, R. A. Burgess, H. Kienzler, H. E. Jack and e. al., "COVID-19 and global mental health," vol. 8, no. 6, 01 June 2021.

[11] G. Orrù, M. Monaro, C. Conversano, A. Gemignani and G. Sartori, "Machine Learning in Psychometrics and Psychological Research," vol. 10, 2020.

[12] C. Comito and C. Pizzuti, "Artificial intelligence for forecasting and diagnosing COVID-19 pandemic: A focused review," Artif Intell Med, vol. 128, p. 102286, 2022. N. R., "Impact of covid-19 pandemic on mental health using machine learning," british, 1 january 1970.

[13] C. Herbert, A. E. Bolock and S. Abdennadher, "A survey using psychological and linguistic self-report measures, and machine learning to investigate mental health, subjective experience, personality, and behaviour during the COVID-19 pandemic among university students," *BMC Psychology*, 2021.

[14] F. M. Albagmi, A. Alansari, D. S. A. Shawan and H. Y. AlNujaidi, "Prediction of generalized anxiety levels during the Covid-19 pandemic: A machine learning-based modeling approach," 19 January 2022.

[15] "Predicting psychological state among Chinese undergraduate students in the covid-19 epidemic: A longitudinal study using a machine learning," 2020.

[16] I. P. Jha, R. Awasthi, A. Kumar, V. Kumar and T. Sethi, "Learning the mental health impact of covid-19 in the united states with explainable artificial intelligence: Observational study," *JMIR mental health*, vol. 8, no. 4, p. 25097, 2021.

[17] M. tubio-Fungueirino and E. Cernadas, "Viability Study of Machine Learning-Based Prediction of COVID-19 Pandemic Impact in Obsessive-Compulsive Disorder Patients," 10 February 2022.

[18] M. Rezapour and S. K. Elmshaeuser, "Artificial intelligence-based analytics for impacts of COVID-19 and online learning on college students’ mental health," 18 November 2022.

[19] C. Ntakolia, D. Priftis, K. Kotsis, K. Magklara, M. C. Travlou, I. Rannou and K. Ladopoulou, "Explainable Ai-Based Identification of Contributing Factors to the Mood State Change of Children and Adolescents with Pre-Existing Psychiatric Disorders in the Context of COVID-19 Related Lockdowns in Greece," 22 December 2021.

[20] Z. Ma, S. Idris, Y. Zhang, L. Zewen, A. Wali, Y. Ji, Q. Pan and Z. Baloch, "The impact of COVID-19 pandemic outbreak on education and mental health of Chinese children aged 7–15 years: an online survey," BMC Pediatrics, vol. 21, 2021.

[21] S. Changwon and e. al., "Effects of COVID-19 on college students' mental health in the United States: Interview survey study," *Journal of medical internet research*, vol. 22.9, 2020.

[22] M. Aarah-Bapuah, S. S. Oppong, A. O. Yawson, G. Dzansi and S. Adjorlolo, "Covid-19 and mental health of children and adolescents: A systematic review, *Cogent Psychology*," vol. 9, no. 1, p. 2111849, 2022.