Modelling an Innovative Machine Learning Model for Student Stress Forecasting

Dr. Anusuya Biswas[1*]

[1*]*Alliance University, India.*

Received: 06 March 2024; Revised: 08 April 2024; Accepted: 07 May 2024; Published: 28 June 2024

Abstract

Student stress has become an important distress in an educational environment, affecting academic presentation and overall well-being. Traditional methods of stress prediction in students lack accuracy and fail to adapt to the dynamic nature of stress factors. This study aims to develop and validate an advanced machine-learning model for predicting student stress. Data was collected from a various group of students through studies and behavioural valuations, capturing numerous stress-related indicators. Data pre-processing for scaling the features of the model was achieved through min-max normalization. Study utilized a Fine-tuned Squirrel Search-driven Light Gradient Boosting Machine (FSS-LGBM) model, which integrates advanced optimization techniques with Light GBM for enhanced predictive performance. The FSS-LGBM model validated superior performance associated with convolutional machine learning models, achieving higher accuracy (98%), recall (99%), precision (97%), and F1-score (97%) in student stress prediction. The advanced machine learning model developed in this study, incorporating FSS-LGBM, offers a promising approach for predicting student stress. By employing advanced data pre-processing and optimization techniques, the model offers a proactive tool for early stress detection, potentially leading to enhanced mental well-being intrusions and support for students.

Keywords: Student's Stress; Students' Well-being; Fine-tuned Squirrel Search-driven Light Gradient Boosting Machine (FSS-LGBM); Machine Learning Model; Predictive Accuracy.

I. INTRODUCTION

Stress has become an issue confronting learner in today's schools, colleges, and universities since the expectations, pressures, and responsibilities increase (PV et al., 2022). Some consequences of stress that can manifest among students are low life quality, psychiatric issues, and poor academic results (Monisha et al., 2020). Several conventional methods of activity monitoring and stress control, such as periodic investigations or psychological tests fail to offer real-time measurement or prompt problem-solving. Identification of student stress levels is important because it contributes to early prevention but also the provision of adequate support measures (Sinha & Sriram, 2022). Educators and mental health professionals recognize mental stress predictors, allowing such issues to be handled in advance rather than after they occur. Stress forecasting can help identify students who are at risk early on and provide appropriate treatments to improve learning and mental health (Colizzi et al., 2020). In the context of student stress forecasting, machine learning models can consider a large number of parameters, such as academic achievement, missed lessons, behavior, or physiological data, and forecast stress indexes with higher reliability (Li & Sano, 2020). Utilizing these models, educators and mental health

https://doi.org/10.66027/GPIM/V2I2/GPM24203

professionals can identify particular stress factors and implement appropriate preventative and treatment measures for every student. The study aims to develop an innovative machine-learning approach for forecasting students' stress.

II. LITERATURE REVIEW

Verma et al., (2020) detected the factors, particularly in engineering schools, that affect mental health disorders, including anxiety, stress, and depression in college students. Two models for machine learning. Support vector machines (SVM) and logistic regression (LR) were used to assess the students' stress levels. The precision of the SVM was 86.84%, whereas the logistic regression's accuracy was 67%. Using machine learning methods, predicted, anxiety, stress, and depression (Malik & Khan, 2023). Data from 400 students were collected using the Depression, Anxiety, and Stress Scale – 21 items (DASS21). The approaches employed for classification were SVM, Naive Bayes, decision trees, logistic regression, and K-Nearest Neighbour (KNN). After implementing each approach, that logistic regression and KNN were observed most effectively. To evaluate stress and treat it early with machine learning techniques, about 190 children between the ages of 14 and 18 were polled using a range of 26 questions in a real-time database developed (Bisht et al., 2022). Diverse machine learning methodologies were employed, such as KNN, Random Forest (RF), Decision Tree (DT), and LR, in which KNN attains the greatest accuracy of 88%. By employing training and test situations, (Firoz et al., 2023) intended to identify the psychological distress in a group of 2116 pupils. The dataset was then subjected to six classification methods, which were used: KNN classifier, Random Forest classifier (RFC), Multinomial Naive Bayes, Support Vector classifier (SVC), and K-Nearest Neighbours Regression (KNNR). The Decision Tree and Random Forest classifiers outperformed each other, achieving the maximum possible test score of 0.99%. By using physiological data gathered via a wearable device, (Dham et al., 2021) intended to automate the process of detecting mental stress. Several Artificial Intelligence models, including the Radial Basis Function (RBF) Network, Stacking Classifier, Support vector machine and artificial neural network support vector machine (ANNSVM), and Artificial Neural Network (ANN), were employed to evaluate in terms of accurately predicting the stress state. The experimental findings show that the identification of mental stress might be significantly influenced by physiological signals. Verma & Verma, (2020) investigated mental stress among Indian college students, focusing on technical programs. Data from 2500 graduating students from northern Indian technical education schools and universities was collected using standardized questionnaires. The aim was to determine early stress levels and provide strategies for managing anxiety, stress, and depression.

III. METHODOLOGY

The methodology section shows a collection of datasets, pre-processing using min-max normalization, and a Fine-tuned Squirrel Search-driven Light Gradient Boosting Machine (FSS-LGBM) model for student stress prediction. Figure 1 shows the research framework.

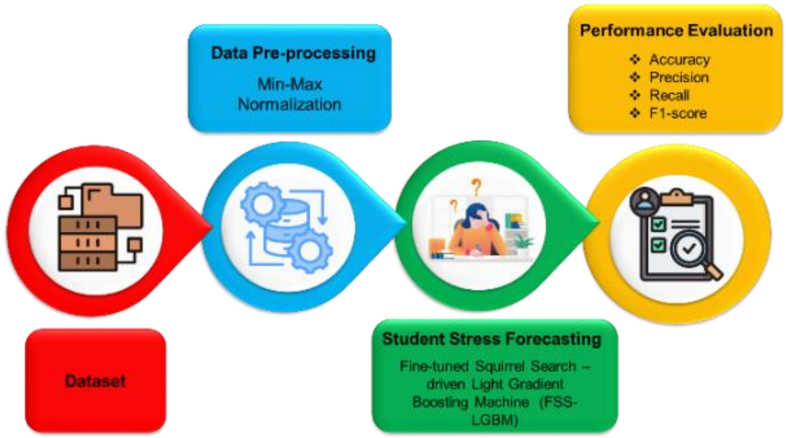

Figure 1: An Overview of the Research Procedure

3.1 Data Collection

For this research, a comprehensive dataset from kaggle was employed, providing extensive insight into student stress factors [https://www.kaggle.com/datasets/rxnach/student-stress-factors-a-comprehensive-analysis]. The dataset includes around 20 factors that contribute to student stress, which are classified into five major factors: These can be: psychological, physiological, social environment, and academic factors. The dataset is separated into two categories: The dataset consists therefore of 80% for training and 20% for testing.

3.2 Data Pre-processing USING Min-Max Normalization

In student stress forecasting, min-max normalization is used to scale the feature value to a consistent range, which is commonly [0,1], often based on the min-max equation (1).

$$u' = \frac{u - min_X}{min_X - max_X}(new\_max_X - new\_min_X) + new\_min_X \qquad (1)$$

Where u'represents the new normalized value of the feature and $u$indicates the feature's initial value, In the initial dataset, max_X, and min_X represent the maximum and minimum values of feature X, and new_max_X and new_min_Xrepresent the new ranges' maximum and minimum bounds, typically [0,1] for standard normalization. By accelerating convergence throughout the training phase, this step improved the model's performance.

Features such as psychological, physiological, social, environmental, and academic factors are scaled to a consistent range using min-max normalization for students' stress forecasting. This assures the various data types are treated regularly, consequently improving the model's accuracy and efficiency. Standardizing data is the essential first step toward achieving the best possible model performance because it significantly increases the efficiency of training and forecasting dependability.

3.3 Fine-tuned Squirrel Search-driven Light GBM (FSS-LGBM)

The fine-tuned squirrel search-driven light GBM is certainly novel in forecasting students' stress levels. It integrates the accurate prediction feature of LightGBM with the optimal parameters gained by the Fine-tuned squirrel search optimization algorithm. To increase the accuracy and the efficiency of stress prediction, FSS-LGBM makes use of Fine-tuned squirrel search to further tune the hyperparameter of LightGBM. This integration enabled the model, to address challenges with handling large and intricate data related to student stress and would be useful in managing student stress.

3.3.1 Light Gradient Boosting Machine (LGBM)

The LightGBM approach incorporates the use of two innovative techniques: bundles of special features and gradient-based one-side sampling. LightGBM is particularly beneficial in the case of stress forecasting for students, as it allows for working with numerous predictors reflecting different aspects of stress in students. It employs a fast, accurate, and robust machine learning approach known as LightGBM, which is a decision tree. To train, LightGBM combines a lot of single decision tree estimators: Every single decision tree first separates the model's predicted values from its actual ones, and then merges them. The representation of this algorithm and equation (2) is as follows:

$$F_m(x) = F_1(x) \sum_{n=1}^{M} h_n(x) \tag{2}$$

Here, $F_1(x)$ is the primary forecast model, which is boosted by different other decision tree $h_n(x)$. Compared with other models, LightGBM enhances performance and adaptability where there are large sets of features and data volumes, including student stress prediction. This enhancement is made possible through the use of gradient-based one-side sampling (GOOS), and exclusive feature bundling (EFB), two novel techniques. In this case, GOOS aids in the reduction of data sampling of gradient-based and integrity-based data so that learning does not have to be undertaken on the entire dataset of stress indicators for a student. EFB allows LightGBM to quickly find the branch sites by allowing to select the chosen sparse characteristics instead of evaluating all the possible characteristics. With these techniques, LightGBM is immune to and/or capable of addressing and analyzing complex stress patterns among students and consistently has better, if not the best, training time and memory usage, compared to other approaches.

3.3.2 Fine-tuned Squirrel Search Algorithm (FSS)

The FSS optimizes the machine learning model by efficiently exploring the parameter space, improving stress forecasting accuracy. It adapts to intricate stressors, offering more accurate forecasts and assisting with early intervention in the learning environment. A Fine-tuned Squirrel Search optimization methodology includes four new methods to increase the program's capacity for searching. The sections that follow will provide a detailed explanation of each of the four methods.

- **Adaptive strategy for estimating predator presence probability:** similar to how predators influence the behaviors of flying squirrels, various stressors (psychological, physiological, social, environmental, and academic factors) impact students differently over time and the adaptive strategy dynamically adjusts the probability of these stressors represented by these variable O_{dp}, affecting the model based on the number of iterations, allowing the FSS's to focus on the most significant stress factors as the prediction process evolves. The following equation (3) represent the probability of dynamically based iteration:

$$O_{dp} = \left(O_{dpmax} - O_{dpmin}\right) \times \left(1 - \frac{Iter}{Iter_{max}}\right)^{10} + O_{dpmin} \tag{3}$$

Where O_{dpmax} and O_{dpmin} represents the maximum and minimum probability of the existence of stressors, respectively.

- **Cloud-based random position for flying squirrel:** When$Q_1, Q_2, Q_3 < O_{dp}$, the flying squirrels randomly move on to the next possible food places. Students' stress responses modeled by the variable such as FS_{at}^{new}, it can be unpredictable and vary gently, much like the erratic foraging behavior of squirrels. The FSS uses a cloud-based model to generate new positions for these stress level variables, capturing the diverse and ambiguous nature and allowing the algorithm to identify more nuanced factors in stress prediction. The following equations are used in place of the equations (4) to (6):

$$FS_{at}^{new} = \begin{cases} FS_{at}^{old} + c_h H_d\left(FS_{ht}^{old} - FS_{at}^{old}\right), if\ Q_1 \geq O_{dp} \\ Dw\left(FS_{at}^{old}, En, He\right), \quad otherwise \end{cases} \tag{4}$$

$$FS_{nt}^{new} = \begin{cases} FS_{nt}^{old} + c_h H_d\left(FS_{at}^{old} - FS_{nt}^{old}\right), if\ Q_2 \geq O_{dp} \\ Dw\left(FS_{nt}^{old}, En, He\right), \quad otherwise \end{cases} \tag{5}$$

$$FS_{nt}^{new} = \begin{cases} FS_{nt}^{old} + c_h H_d\left(FS_{ht}^{old} - FS_{nt}^{old}\right), if\ Q_3 \geq O_{dp} \\ Dw\left(FS_{nt}^{old}, En, He\right), \quad otherwise \end{cases} \tag{6}$$

Flying squirrels' foraging behavior is erratic and ambiguous, with individual squirrels exhibiting varying perceptions and orientations. A typical cloud model might be used to incorporate and synthetically explain these features. For each flying squirrel, new locations are generated using a conventional cloud model generator instead of evenly distributed random functions. The environmental parameter is decreased using equation (7) to finish iteration:

$$En = En_{max} \times \left(1 - \frac{Iter}{Iter_{max}}\right)^{10} \tag{7}$$

- **Strategies for successive positioning choices:** This implies continuously updating the new positions of stress factors based on the fitness value. In every iteration, if a new position provides a better prediction of stress levels than the previous one, otherwise it retains the previous position. This process mimics the ongoing assessment and refinement needed in student stress prediction. There is a mathematical description for this tactic. Equation (8) compares and chooses the optimal location for stress factors.

$$FS_j = \begin{cases} FS_j^{new}, if\ e_j^{new} < e_j^{old} \\ FS_j^{old}, \quad oterwise \end{cases} \tag{8}$$

Where $En_{max} = \frac{FS_V - FS_K}{4}$ is the maximum search radius.

- **Enhancing dimensional search technique:** The fundamental FSS is updating each of a single dimension of stress factor for students' stress prediction simultaneously. However, this approach has a flaw in that certain dimensions are dependent on others, thus modifications to one stress factor might have undesirable impacts on others. Every iteration involves the following procedures to enhance the search: identify the optimal best place for the flying squirrel (stress factor of combination for predicting student stress), then create a new solution based on this optimal combination, compare the predictive accuracy of the new solution with the original, and repeat these processes for each stress dimension. An equation (9) is used to accurately refine the parameters of the stress components. By contrasting the initial stress prediction with the newly generated solutions, the FSS ensures that the most accurate and effective stress model is developed.

$$FS_{best,i}^{new} = Dw\left(FS_{best,i}^{old}, En, He\right), i = 1,2, \ldots \ldots, m \tag{9}$$

IV. PERFORMANCE EVALUATION METRICS

The experimental setup for the student stress forecasting model consisted of an Intel Core i7 CPU with 16 GB of RAM and a machine learning framework that used Python 3.8 together with TensorFlow for model development and evaluation. The efficacy of the proposed and existing methods was assessed employing the following metrics: accuracy, recall, precision, and F1 score. The suggested strategy was contrasted with the existing methods of Multilayer Perceptron (MLP), Deep Learning (DL), and Decision Tree (DT) (Zainudin et al., 2021).

The model's accuracy graph shows its ability to forecast student stress levels over time, while its loss graph shows its error rate, indicating improving training efficiency. The graph illustrates the model's ability to forecast student stress, exhibiting variations in accuracy and error reduction throughout various epochs. Figure 2 demonstrates the outcome of the accuracy and loss graph.

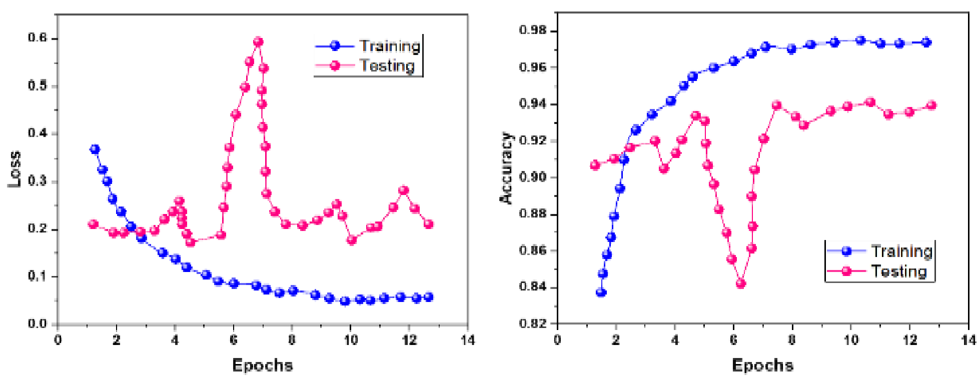

Figure 2: Result of Training and Testing Graph

- **Accuracy:** the accuracy metrics quantify the percentage of all tested students' stress levels, both stressed and non-stressed, that was accurately predicted. The proposed method FSS-LGBM (98%) demonstrates superior performance compared to the existing approaches, such as MLP (86%), DL (91%), and DT (95%). Figure 3 and Table 1 demonstrate the outcome of accuracy.

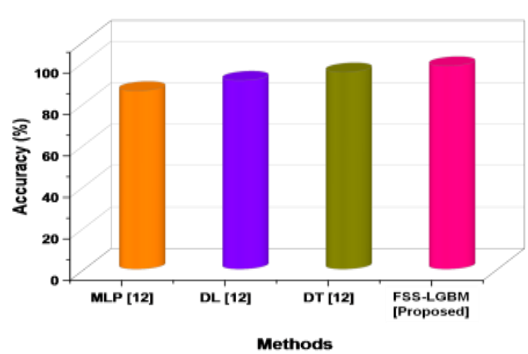

Figure 3: Accuracy Performance Result

- **Recall**: The fraction of accurately predicted stressed-out students, among actual stressed students. The suggested technique **FSS-LGBM (99%)** performs better than existing approaches MLP (87%), DL (91%), and DT (96%). Figure 4 and Table 1 demonstrate the result of recall.

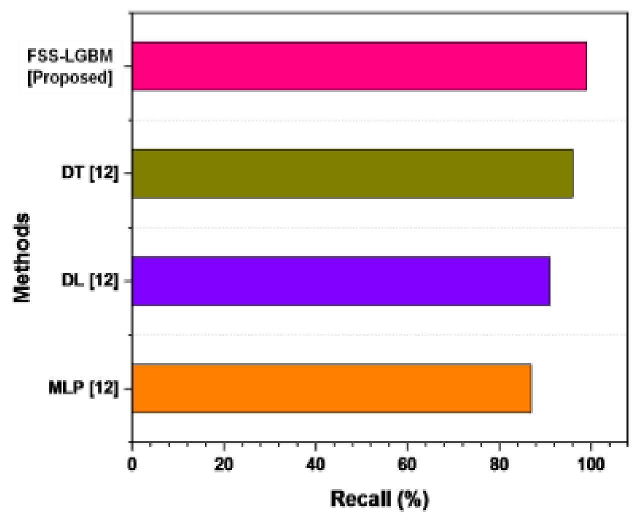

Figure 4: Recall Performance Result

- **Precision: It** indicates the percentage of stressed-out students who were stressed out. In comparison to existing methods MLP (84%), DL (92%), and DT (96%), the suggested methodology **FSS-LGBM (97%)** exhibits better performance. Figure 5 and Table 1 determines the outcome of precision.

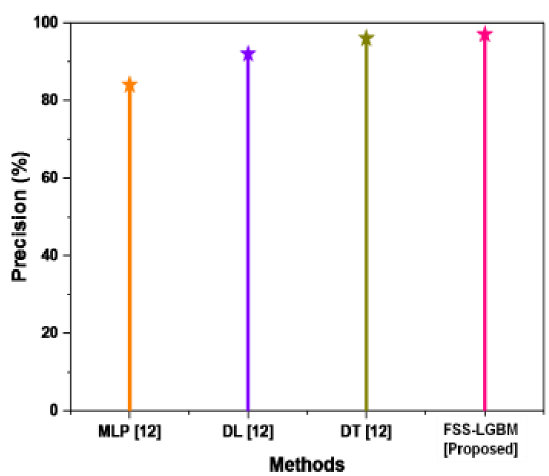

Figure 5: Precision Performance Result

- **F1score**: It is a statistic that assesses the model's ability to detect stressed students by finding the balance between recall and accuracy. When compared to the existing approaches MLP (85%), DL (91%), and DT (96%), the recommended strategy **FSS-LGBM (97%)** performs better. Figure 6 and Table 1 shows the result of the f1score.

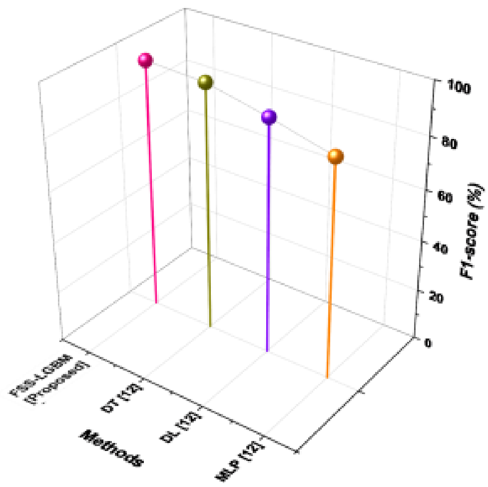

Figure 6: F1score Performance Result

Table 1: Comparisons Metrics of Accuracy, Recall, Precision, and F1score

Models	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**F1score (%)**
Multilayer Perceptron (MLP) (Zainudin et al., 2021)	86	87	84	85
Deep Learning (DL) (Zainudin et al., 2021)	91	91	92	91
Decision Tree (DT) (Zainudin et al., 2021)	95	96	96	96
FSS-LGBM [Proposed]	**98**	**99**	**97**	**97**

V. CONCLUSION

The innovative machine learning model established in this study, employing a Fine-tuned Squirrel Search-Driven Light Gradient Boosting Machine (FSS-LGBM), represents a substantial development in forecasting student stress. By leveraging Min-Max Normalization for data pre-processing and advanced optimization techniques, the model outperformed conventional approaches, representing superior accuracy, recall, precision, and f1-scores of (98%, 99%, 97%, and 97%). This approach offers a proactive tool for timely stress detection, which can improve mental well-being interventions and support systems for students, addressing the dynamic nature of stress factors in the educational environment. Model performance might vary depending on data quality and the education environment. The flexibility and real-time integration of the model is used to enable more extensive and customized applications.

REFERENCES

[1] PV, V. R., Rao, A., Neha, S., & Thingalaya, S. S. (2022, October). Classification of Stress in Students using Machine Learning Algorithms. In *2022 International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER)* (pp. 229-233). IEEE. https://doi.org/10.1109/DISCOVER55800.2022.9974872

[2] Monisha, S., Meera, R., & Swaminath, V. (2020, January). Predictive analysis of student stress level using Naïve Bayesian classification algorithm. In *2020 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-7). IEEE. https://doi.org/10.1109/ICCCI48352.2020.9104113

[3] Sinha, S., & Sriram, R. (2022, May). An educational based intelligent student stress prediction using ml. In *2022 3rd International Conference for Emerging Technology (INCET)* (pp. 1-7). IEEE. https://doi.org/10.1109/INCET54531.2022.9824636

[4] Colizzi, M., Lasalvia, A., & Ruggeri, M. (2020). Prevention and early intervention in youth mental health: is it time for a multidisciplinary and trans-diagnostic model for care?. *International journal of mental health systems*, *14*, 1-14. https://doi.org/10.1186/s13033-020-00356-9

[5] Li, B., & Sano, A. (2020). Extraction and interpretation of deep autoencoder-based temporal features from wearables for forecasting personalized mood, health, and stress. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *4*(2), 1-26. https://doi.org/10.1145/3397318

[6] Verma, G., Adhikari, S., Khanduri, V., Tandon, S., Rawat, S., & Singh, P. (2020, August). Machine learning model for prediction of stress levels in students of technical education. In *AIJR Proceedings* (pp. 45-52). AIJR Publisher. https://doi.org/10.21467/proceedings.100.5

[7] Malik, S. S., & Khan, A. (2023, April). Anxiety, Depression and Stress prediction among College Students using Machine Learning Algorithms. In *2023 Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)* (pp. 1-5). IEEE. https://doi.org/10.1109/ICEEICT56924.2023.10157693

[8] Bisht, A., Vashisth, S., Gupta, M., & Jain, E. (2022, April). Stress prediction in Indian school students using machine learning. In *2022 3rd International Conference on Intelligent Engineering and Management (ICIEM)* (pp. 770-774). IEEE. https://doi.org/10.1109/ICIEM54221.2022.9853200

[9] Firoz, M., Islam, M. M., Shidujaman, M., Islam, A., & Habib, M. T. (2023, September). University student's mental stress detection using machine learning. In *Seventh International Conference on Mechatronics and Intelligent Robotics (ICMIR 2023)* (Vol. 12779, pp. 757-767). SPIE. https://doi.org/10.1117/12.2690039

[10] Dham, V., Rai, K., & Soni, U. (2021, August). Mental stress detection using artificial intelligence models. In *Journal of Physics: Conference Series* (Vol. 1950, No. 1, p. 012047). IOP Publishing. https://doi.org/10.1088/1742-6596/1950/1/012047

[11] Verma, G., & Verma, H. (2020). Model for predicting academic stress among students of technical education in India. *International Journal of Psychosocial Rehabilitation*, *24*(04), 2702-2714. https://doi.org/10.37200/IJPR/V24I4/PR201378

[12] Zainudin, Z., Hasan, S., Shamsuddin, S. M., & Argawal, S. (2021, August). Stress detection using machine learning and deep learning. In *Journal of Physics: Conference Series* (Vol. 1997, No. 1, p. 012019). IOP Publishing. https://doi.org/10.1088/1742-6596/1997/1/012019