



Mental Health Identification of Children and Young Adults in a Pandemic Using Machine Learning Classifiers

Xuan Luo* and Youlian Huang

School of Pre-school Education, Yichun Early Childhood Teachers College, Yichun, China

OPEN ACCESS

Edited by:

Ali Ahmadian,
Mediterranea University of Reggio
Calabria, Italy

Reviewed by:

Venkatachalam K,
University of Hradec Králové, Czechia
Ayoub Salimipour,
Quchan University of Advanced
Technology, Iran
Soheil Malekshah,
University of Wisconsin–Milwaukee,
United States

*Correspondence:

Xuan Luo
lxph411255313@sina.com

Specialty section:

This article was submitted to
Educational Psychology,
a section of the journal
Frontiers in Psychology

Received: 19 May 2022

Accepted: 08 June 2022

Published: 29 July 2022

Citation:

Luo X and Huang Y (2022) Mental
Health Identification of Children and
Young Adults in a Pandemic Using
Machine Learning Classifiers.
Front. Psychol. 13:947856.
doi: 10.3389/fpsyg.2022.947856

COVID-19 has altered our lifestyle, communication, employment, and also our emotions. The pandemic and its devastating implications have had a significant impact on higher education, as well as other sectors. Numerous researchers have utilized typical statistical methods to determine the effect of COVID-19 on the psychological wellbeing of young people. Moreover, the primary aspects that have changed in the psychological condition of children and young adults during COVID lockdown is analyzed. These changes are analyzed using machine learning and AI techniques which should be established for the alterations. This research work mainly concentrates on children's and young people's mental health in the first lockdown. There are six processes involved in this work. Initially, it collects the data using questionnaires, and then, the collected data are pre-processed by data cleaning, categorical encoding, and data normalization method. Next, the clustering process is used for grouping the data based on their mood state, and then, the feature selection process is done by chi-square, L1-Norm, and ReliefF. Then, the machine learning classifiers are used for predicting the mood state, and automatic calibration is used for selecting the best model. Finally, it predicts the mood state of the children and young adults. The findings revealed that for a better understanding of the effects of the COVID-19 pandemic on children's and youths' mental states, a combination of heterogeneous data from practically all feature groups is required.

Keywords: machine learning, COVID-19, artificial intelligence, mental health, feature selection, mood state, clustering

INTRODUCTION

The World Health Organization (WHO) confirmed the new coronavirus (COVID-19) as its source of pneumonia in Wuhan, China, in December 2019, and proclaimed COVID-19 a pandemic on 11 March 2020 (Ntakolia et al., 2022). About 184 nations enacted strict precautions to control the growth of COVID-19 between 31 December 2019 and 4 May 2020, including lockdown limitations and quarantined periods, resulting in economical, ecological, and mental health problems. The lockdown tactics helped governments to reduce the spread of COVID-19, but the rise of mental health problems is highly concerning. Furthermore, the lockdown is still in effect, and the total academic sector has been affected, with schools, colleges, and institutions facing partial or complete closures, so both academic and non-academic activities are severely impacted (Demir et al., 2022). However, the lockdown is enforced and maintained to prevent the virus from spreading, and a slew of negative

consequences have emerged. To begin with, all academic and non-academic events have been suspended. Pupils are becoming bored due to the closure of academic institutions, and the drive for gaining information to be successful in a positive manner is dwindling.

The effect of COVID-19 as well as its limitation rules on the researched group has been investigated in a number of studies (McKune et al., 2021; Qin et al., 2021; Błaszczuk et al., 2022). Multivariable logistic regression models were used in many research to: (i) examine the possible risk factors defined as self-emotional problems (Ren H. et al., 2021) and (ii) compare the impact of COVID-19 assessments (Gomathi et al., 2021).

Even during a disease outbreak, binomial or binary logistic multiple regressions have been used to: (i) recognize difficulty sleeping between many adolescents and young adults (12–29 years) (Islam et al., 2020; Zhou et al., 2020), (ii) evaluate anxiety and depression between many college students (Gomathi et al., 2020; Yeasmin et al., 2020), and (iii) investigate the risk of developing depression between many children as well as the potential connection to COVID-19 (Garcia de Avila et al., 2020). Univariate logistic regression analysis was employed in other research focused on youth to detect psychological disorders (Shanmugam et al., 2017; Liang et al., 2020). Then, during the COVID-19 epidemic among University students, hierarchical logistic regression models were employed to investigate characteristics linked to mental health concerns (Ma et al., 2020). The relationship among COVID-19 anxiety and problems in children and teenagers was investigated using modified logistic regression analysis (Ma et al., 2020).

The results of previous studies introduced above centered on Chinese regions and University students and also used conventional regression algorithms to detect correlation coefficients between risk variables and mental health issues, such as logistic regression and chi-square examinations (Ge et al., 2020; Rens E. et al., 2021; Sciberras et al., 2022), although only very few were using machine learning techniques. Moreover, no study focusing on children and teens with identified mental problems has been conducted to their understanding. As a result, this research proposed the establishment of such an understandable machine learning pathway to understand the implications and effects of Greece's first lockdown on the mental health of children and teenagers.

The main contribution of the proposed work is given below:

- In this work, a *post hoc* explainability model is used for detection and it identifies the mental health of youths and adults during the lockdown in Greece,
- Susceptible, Infected, Recovered, and Deceased (SIRD) automatic calibration model effectively detects the mood states and reduces the computation cost.
- **For the feature selection process, it uses chi-square, L1-Norm, and ReliefF method, which efficiently selects the features for detecting mental health.**

The rest of our research article is written as follows: Section Related Work discusses the related work on various COVID-19 lockdown scenarios and mental health. Section Proposed Method shows the algorithm process and general working

methodology of the proposed work. Section Result Analysis evaluates the implementation and results of the proposed method. Section Conclusion concludes the work and discusses the result evaluation.

RELATED WORK

Primarily, during the COVID-19 pandemic, a multivariable logistic regression method was used to test connections between socio-demographic characteristics and mental health issues in Chinese teenagers. There were 8,079 Chinese students aged 12 to 18 in the community. The Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder (GAD-7) questionnaire were used to obtain information to assess symptoms of depression and anxiety. Female students as well as those with top grades were shown to have a higher probability of reporting mental health symptoms (Chakraborty and Maity, 2020).

Similarly, throughout COVID-19, secondary research was completed on the psychological health of Chinese children between the ages of 7–15 years, with a number of 668 families from various provinces of China. The statistical analyses were performed using multiple logistic regression assessment to determine the elements that make a contribution to Chinese children's support and mental health, and it was discovered that public schools and the provincial capital of beginnings were the important considerations associated with treating PTSD, with the majority of individuals showing a better viewpoint regarding online learning (WHO, 2020). The author used information from the General Health Questionnaire (GHQ-12), the PTSD Checklist—Civilian Version (PCL-C), or the Adverse Behavioral Strategies Scale to investigate the impact of COVID-19 on juvenile mental illness in China.

The effectiveness of simulation using epidemiological models (SIR, SEIRV, etc.) was also examined, as was the discovery of design variables (Bonardi et al., 2020; Vindegaard and Benros, 2020; Abas et al., 2021; Ma et al., 2021), which are critical for successful modeling fits into actual data. Furthermore, the contemporary estimating methodology allows the study of design variables as they emerge and continuous updating of their estimates, thanks to the tremendous development in processing capacity in recent years. For instance, (Vizheh et al., 2020) examined several time-series approaches to forecast the number of actual COVID-19 incidences and mortality in Chile. In (Penner et al., 2021) employed a Bayesian method again for an agent-based system. Finally, Prati and Mancini (2021) examined deep learning approaches (based on LSTM neural networks).

The relationship between socio-demographic characteristics and COVID-19-related characteristics, as well as their impact on sadness, tension, and pressure among teenagers in Spain, was investigated (Masten, 2021). A final tally of 523 adolescents (13–17 years old) finished the Depressive episodes, Nervousness, and Anxiety Scale (DASS-21) and the Oviedo Infrequency Scale (INFO-OV), with findings denoting those girls who worked knowingly and willingly and decided to stay home more commonly, and were much more inclined to display the signs of depression, anxiousness, and strain. Furthermore, when

performing COVID-19 study, the researchers discovered a link between mental anguish and chronic stress. Finally, those in a romantic connection, as well as those who had previously been afflicted with COVID-19, were much more likely to also have better psychological health. A summary of research on the first COVID-19 pandemic is stated in **Table 1**.

The network reliability is discussed by Ansari and Malekshah (2019). Security of the system is the key assessment (Malekshah and Javad, 2021) which is implemented for environmental-related data distribution. Deep reinforcement learning (Malekshah et al., 2022) is used for distributing the power in the dynamic network with high reliability.

The limitations of the existing work are studied above. The main issue is that some techniques fail to estimate real situations before taking statistics. The situation in the sense is that if some people are infected with COVID, then the families have high mental pressure. It is a different scenario. Also, the accuracy of the prediction is not as expected.

PROPOSED METHOD

In this work, we concentrated on the marginalized populations of children and teenagers to forecast the effect of COVID-19 and the first lockdown implemented in Greece from 23 March to 4 May 2020. The Hellenic COVID-19 impact survey (HOPE) information is used, in a large study of parents with children who were planning to attend CAMHS in Greece each year even before the global epidemic (1 March 2019 to 1 March 2020). The proposed work consists of six stages. Initially, the data are collected through a questionnaire. Then, the collected information is pre-processed. Then, the clustering process is done, and feature selection is based on chi-square, L1-Norm, and ReliefF. Next, it uses the automatic calibration method Susceptible, Infected, Recovered, and Deceased (SIRD) along with the machine learning models that were used for evaluation. Finally, a *post-hoc* explainability helps to detect the best feature using the SHAP model. **Figure 1** shows the overall architecture of the proposed model.

Collection of Data

Children who used CAMHS services were recruited to effectively gather data and create the dataset. This research included 744 children who had their parents (738 parents) complete an online survey on their own. Through 8 May and 1 June 2020, this process was performed. The questionnaire asked about demographic data as well as the parent's assessment of their patient's care 3 months (3 m) well before lockdown and 2 weeks (2 w) following the very first lockdown in Greece.

Pre-processing

There was no requirement for data imputation because the final dataset had no missing values for categorical or numerical variables. Moreover, the dataset was standardized, which is a frequent prerequisite for many ML classifiers.

Data Cleaning

Removing the dataset is the first step in processing it, and it entails carefully removing redundant items and attributes. To start, several categorical variables in the dataset must be deleted for the concerns of confidentiality.

Categorical Encoding

Age, parent, symptoms, and all data points in the dataset must be encoded. Encoding is essential if ML techniques demand numerical data and therefore can handle categorical variables. This study employed one-hot encoding, which creates "dummy" values for each non-numeric attribute's potential classes.

Data Normalization

Most of the items in the data are integers, as noted previously. A handful of these features were also obtained with the help of additional measuring instruments. The performance of the algorithms may be harmed if these attributes are handled without normalization. Normalization is needed to dynamically expand all numerical values into a range between 0 and 1.

$$NOR = \frac{Val - MIN}{MAX} \quad (1)$$

Here, MAX and MIN are the maximum and minimum column values, respectively. Equation 2 shows the normalization formula.

Clustering Process

Several clustering methods were compared. Mini Batch K-Means (Peng et al., 2018), spectral clustering (Von Luxburg, 2007), ward (Ward, 1963; Murtagh and Legendre, 2014), average linkage (Yim and Ramdeen, 2015), balanced iterative reducing and clustering utilizing hierarchies (Birch) (Anchang et al., 2016), and Jenks natural breaks optimization approach (Jenks) (North, 2009; Zhang et al., 2021) were used in the clustering process. The occurrences of the parameter mood change, which reflects the changes in mood state, were clustered.

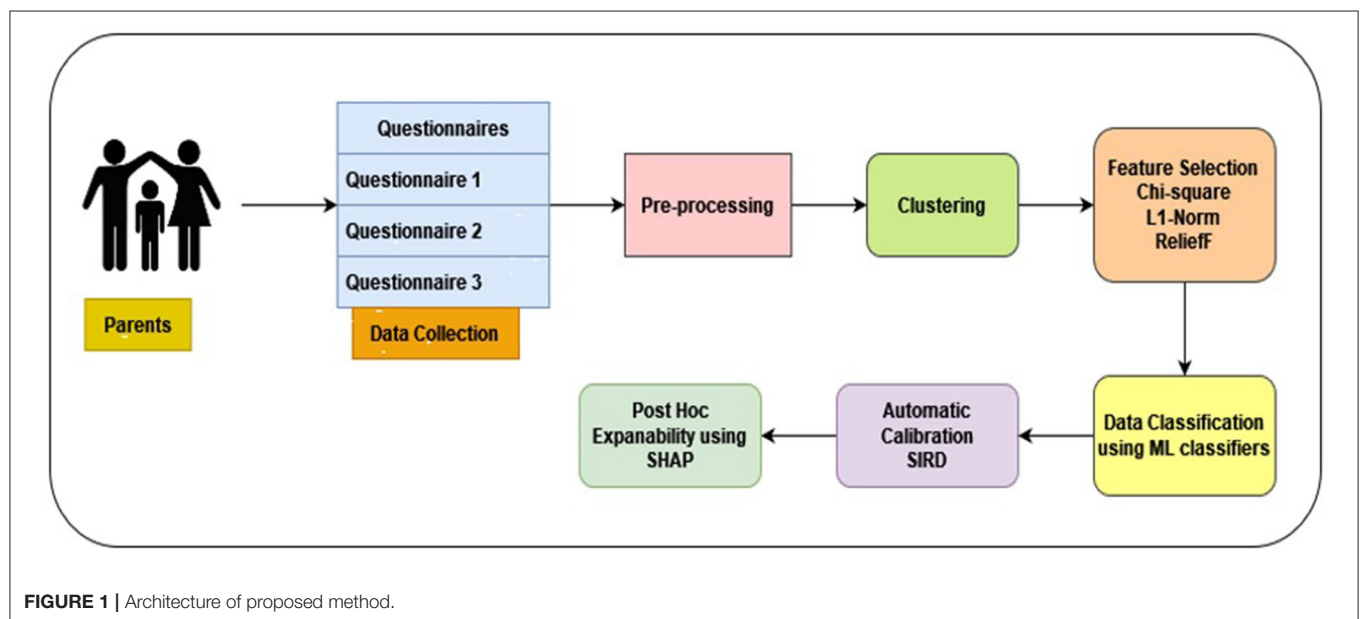
The parameters such as worry, sorrow, anxiety, uneasiness, anhedonia, isolation, irritation, focus, weariness, and rumination are used to construct the emotional condition rating before and during the lockdown. The difference in mood state throughout the last 2 weeks and 3 months until the first lockdown in Greece is the variation in emotional experience. As a result, a negative value for the forecasted variable mood change suggests that the user's mood state score has improved overall, whereas a positive value suggests that the user's attitude condition score has worsened on the whole. Results near to 0 indicate that the subject's mood state score did not change during the lockdown. **Figure 2** shows the clustering methods used.

Feature Selection

The ReliefF, chi-square, and L1-Norm algorithms were used in the feature selection procedure because of their usefulness in diagnosing diseases and categorization difficulties. ReliefF is an approximation of the traditional relief that, thanks to its noise resistance increase, can cope with multiclass difficulties and is thus regarded as ideal again for the present healthcare

TABLE 1 | A summary of research on the first COVID-19 pandemic, which included teens and young people.

| Author | Region | Number of Populations | Focus | Methodologies used |
|-------------------------------|------------|---|--|---------------------------------------|
| Ren Z. et al. (2021) | China | After the reopening of the institution, 478 college students | COVID-19's psychological influence was investigated. | Multivariate logistic regression |
| Cost et al. (2021) | Canada | One thousand thirteen infants and young people between the ages of 6 and 18, even without prior illnesses | To assess the impact of COVID-19 on mental wellbeing. | Multinomial logistic regression |
| Garcia de Avila et al. (2020) | Brazil | 6–12 years old: 157 females and 132 boys | To see how common anxiousness was there during COVID-19. | Logistic regression |
| Yeasmin et al. (2020) | Bangladesh | Three hundred eighty-four families with at minimum one child under the age of eight | Detecting psychological problems during COVID-19 | Binary logistic regression |
| Wathelet et al. (2020) | France | 69,054 college students in France | COVID-19-related mental health disorders will be investigated. | Multivariate logistic regression |
| Sciberras et al. (2022) | Australia | Families of 213 ADHD-diagnosed children and adolescents ages 5 to 17 years old | To determine the COVID-19 limitations' effect | Adjusted logistic regression analyses |



multiclassification challenge. **Figure 3** shows the feature selection methods used.

Chi-Square

A significant feature s_{fi} is picked in the chi-square methods based on its association with a Ch_j class, and the differentiating capability of features s_{fi} followed Ch_j class is determined using the formula:

$$X^2(s_{fi}, Ch_j) = \frac{M \times (a_{ij}d_{ij} - b_{ij}c_{ij})^2}{(a_{ij} + b_{ij}) \times (a_{ij} + c_{ij}) \times (b_{ij} + c_{ij}) \times (c_{ij} + d_{ij})} \quad (2)$$

where M represents the number of observations in the Ch_j class that contain feature s_{fi} is a_{ij} , and the number of samples in the Ch_j class that do not have feature s_{fi} is b_{ij} (Anchang et al., 2016). The sample in the Ch_j class that includes features s_{fi} is a_{ij} , whereas

the number of samples in the Ch_j class that do not have feature s_{fi} is b_{ij} (Guru et al., 2018). The number of observations with characteristic s_{fi} which are not in the Ch_j class is given by c_{ij} . Finally, d_{ij} is the sample size that does not have either as the feature t_i or even the Ch_j class.

Relief Algorithm

When targeted categories are multiclass categorical variables, the ReliefF method calculates the predictive weight. Predictors who show various ratings to neighbors in the same class are punished, but predictors who give the same scores to neighbors in the very same class are encouraged (Tuncer et al., 2020). The predictor variables (WE_j) in the ReliefF method are originally set to 0. The ReliefF algorithm selects a randomized forecast (x_s) on a regular basis, then determines the k -nearest forecasts to x_s within every class, and modifies every nearest neighbor (x_t) (Demir et al., 2020;

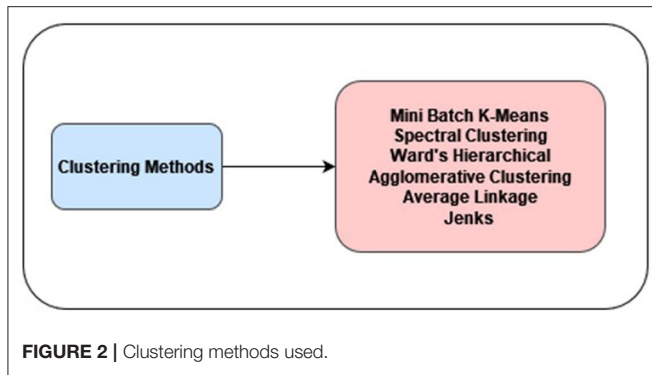


FIGURE 2 | Clustering methods used.

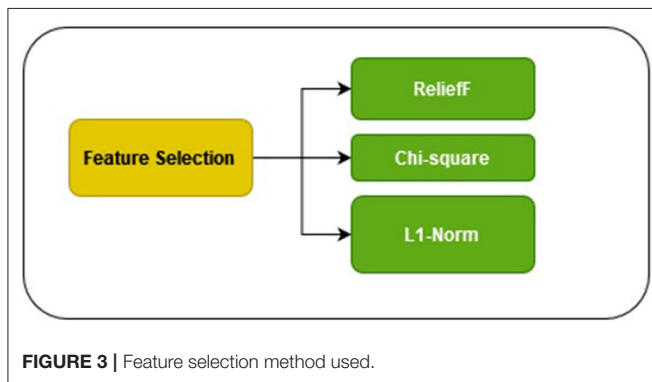


FIGURE 3 | Feature selection method used.

TABLE 2 | Clustering technique's parameter settings.

| Methods | Parameter settings |
|--|--|
| Spectral clustering | Three classes, arpack eigen solver, and affinities for nearest neighbors |
| Average linkage | Average linkage, cityblock affinity, and symmetric connection are three kinds. |
| Mini Batch K-means | Three classes |
| Ward's hierarchical agglomerative clustering | Symmetric connection, three classes, ward linkage |
| Birch | Three classes |
| Jenks | There are three classes, including the lowest value. |

Turkoglu, 2021). If somehow the x_s and x_t classes would be the same, all of the forecasters' values (P_i) are as continues to follow:

$$WE_i^j = WE_i^{j-1} - \frac{\Delta_j(x_s, x_t)}{n} d_{st} \quad (3)$$

If such x_s and x_t categories vary, all of the predictors' values (P_i) were just as described as follows:

$$WE_i^j = WE_i^{j-1} = \frac{Py_s}{1 - Py_t} \cdot \frac{\Delta_j(x_s, x_t)}{n} d_{st} \quad (4)$$

where py_s denotes the prior possible of a category toward which x_s corresponds, py_t indicates the prior probability of the category

toward which x_t originally belonged, n represents the total of repetitions modified by updating, and $\Delta_i(x_s, x_t)$ is the differential with in scores of a predictors P_j among occurrences x_s and x_t . The P_i , $\Delta_i(x_s, x_t)$, could be described as follows for preciseness:

$$i(x_s, x_t) = \begin{cases} 0, & x_{s(i)} = x_{t(i)} \\ 1, & x_{s(i)} \neq x_{t(i)} \end{cases} \quad (5)$$

$$dis_{st} = \frac{dis_{st}^{\sim}}{\sum_{t=1}^l dis_{st}^{\sim}} \quad (6)$$

$$dis_{st}^{\sim} = e^{-(rank(s,t)/sigma)^2} \quad (7)$$

Here, $rank(s,t)$ would be the position of the t^{th} information among the s^{th} observation's closest neighbors, arranged by proximity. The value l represents the nearest neighbors.

L1-Norm SVM

The cost variable for both the features range of choice L1-Norm SVM (Haq et al., 2019) was used to compute the amount of variables. The continuity formula represents the dataset using m elements:

$$SE = ::(x_i, y_i) | x_i \in Rf^n, y_i \in ::-1, 1; ; ; i=1^k \quad (8)$$

If x_i would be the i^{th} samples with n characteristics and a class label, y_i is the y_i^{th} sample (y_i).

Together in the two-class classification task (Equation 10), the SVM finds the separation of the hyperplane which maximizes the boundary length.

$$y_i = (Wx_i - b) \geq 1, i = 1, \dots, k \quad (9)$$

$$\min \frac{1}{2} ||W||^2 \quad (10)$$

As just an outcome of sparse answers, Bradley and Mangasarian utilized Equation (10) by recognizing Equation (11) as a restriction for feature selection-based L1-Norm SVM.

$$\min ||W|| + B \sum_{i=1}^k \max(0.1 - y_i(\alpha^T x_i + b))^2 \quad (11)$$

here, α the weight matrix produced by the optimizing multipliers seems to be the Lagrange (Guo et al., 2017).

Furthermore, the size of the feature set is measured by the amount of the C variable in Equation (11).

Data Classification

For data classification, there are seven popular classifiers used and evaluated to overcome the specified multiclass classification issue: random forest (RF), multilayer perceptron (MLP), eXtreme Gradient Boosting (XG Boost), logistic regression (LR), support vector machine (SVM), k-nearest neighbor (KNN), and decision trees (DTs). The accepted methods are often utilized for clinical classification tasks, and they include tree-based, linear, and neural network prediction models.

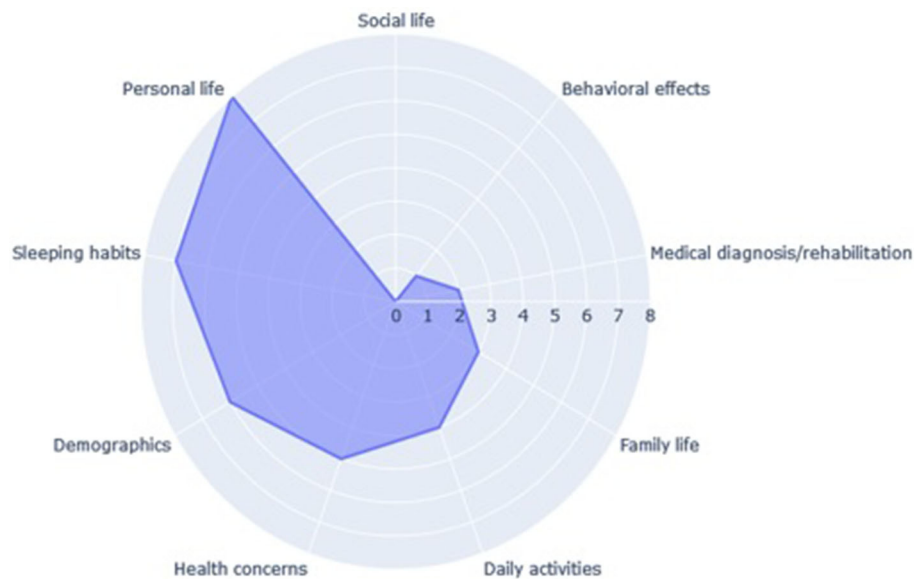


FIGURE 4 | For the first 40 features in which the greatest result was attained, a spider plot showing the amount of characteristics that correspond to every characteristic group was created.

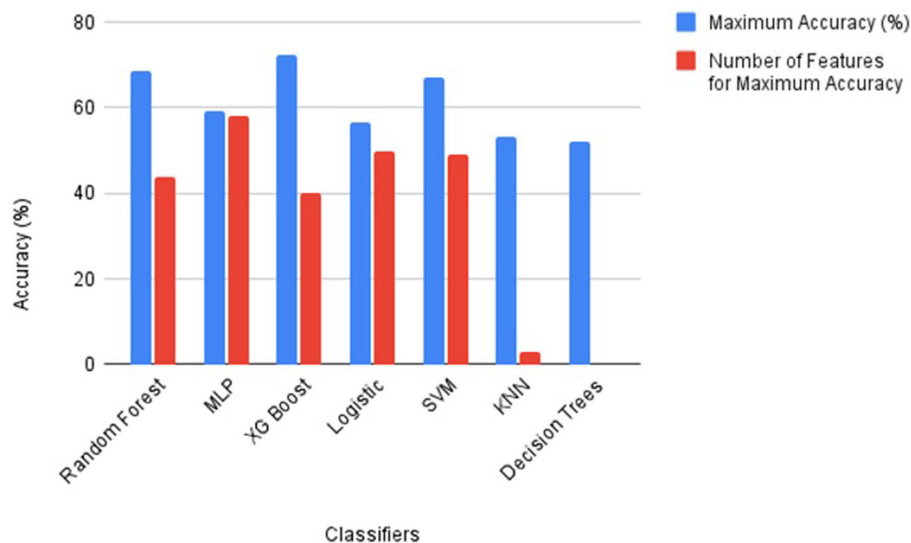


FIGURE 5 | Maximum accuracy for classification methods.

Random Forest

A decision tree forecasts future occurrences using several classifiers instead of a single classifier to arrive at a precise and reliable forecast. RF creates an enormous amount of decision trees. Every decision tree produces a class forecast, and the model's prediction seems to be the category with more scores.

Extreme Gradient Boosting (XG Boost)

The eXtreme Gradient Boosting is a versatile and trimming gradient-boosting software program. Gradient boosting is a

technique that uses a fresh method to forecast the residuals of past versions which are then combined to form the prediction accuracy. While introducing additional models, it utilizes a gradient descent approach to minimize loss.

Multilayer Perceptron

Multilayer perceptron is the most frequent neural network in the area of artificial neural networks (ANNs). MLP uses a supervised learning technique to create a non-linear prediction system. It has several layers, including an input layer, an output layer, and

TABLE 3 | Maximum accuracy for classification methods.

| Models | Maximum accuracy (%) | Number of features for maximum accuracy |
|---------------------|----------------------|---|
| Random forest | 68.6 | 44 |
| MLP | 59.38 | 58 |
| XG Boost | 72.43 | 40 |
| Logistic regression | 56.74 | 50 |
| SVM | 66.95 | 49 |
| KNN | 53.28 | 3 |
| Decision trees | 52.23 | 5 |

TABLE 4 | Results of using SIRD, isotonic regression to calibrate the XG Boost classifier.

| Models used | Log-loss | Accuracy |
|-------------------|----------|----------|
| XG Boost | 1.196 | 68.98 |
| XG Boost+Isotonic | 0.513 | 73.12 |
| XG Boost+SIRD | 0.352 | 78.98 |

the hidden layers. As a result, MLP is a multilayer feedforward neural network in which data are transported unidirectionally from source to the destination layer *via* the hidden layers.

Logistic Regression

The link between data and binary dependent variables is described by a mathematical model. The logistic function $f(x) = 1/(1+e^{-x})$ is used in the model, with $x \in (-\infty, \infty)$ and $0 \leq f(x) \leq 1$. As a result, compared to the value of x , the framework is meant to explain the data in an S-shaped graph with a probability between 0 and 1.

SVM

Support vector machine is a supervised learning model that uses the VC theory statistical learning approach. SVM aims to build a binary classifier, the hyperplane, among two categories that allow for label prediction through one or even more extracted features while maximizing the separation between the nearest distance of every class, termed weight vector, and the hyper-plane.

KNN

K-nearest neighbor is a non-parametric classification approach that attempts to categorize an unknown substance based on its neighbors' known categorization.

Decision Trees

DTs are sequencing designs that integrate a series of simple tests systematically. Every test was used to compare a numerical characteristic to a threshold level or a nominal property to a range of outcomes.

Calibration Using SIRD Model

We looked at two fundamental fractional transmissions of infection theories that can be applied to information combining locally and nationally epidemiological data. The driving concern

was whether certain simple models might assist in predicting future trends in daily epidemiology data. The most basic mathematical method to simulate the evolution of communicable diseases is the SIR paradigm (Błaszczuk et al., 2022), which assumes a community of size N split into S vulnerable, I infected, and R removing people. These three parameters are time-dependent and reflect the number of patients in each group at a specific moment in time.

The pattern suggests that mortality is limited to a subgroup of resistance people, as determined by R development, and also that sickness does not quite introduce additional vulnerable persons after recovering. The SIR system is defined by the following set of equations when studied on the time period $[a, b]$ $[0, T]$ even without vital movements (births and deaths).

$$SH' = -\frac{\beta}{N}SI, \quad IN' = \frac{\beta}{SI} - \gamma I, \quad RE' = \gamma I, \quad DE' = \mu I \quad (12)$$

where $t \in [a, b]$ seems to be the transmitting frequency (gets to control overall amount of spreading), γ is indeed the recovering sample mean, and $R_0 = \frac{\beta}{\gamma}$ is the fundamental reproductive ratio. Several nations track not just the amount of newly confirmed samples and infected people changing constantly, as well as the number of deaths caused by COVID-19.

The tracking is based on the primary goals of nations where Covid case values are determined by standard systems of government. We introduced three low-level cost features presuming that the series $y_i; i=1, \dots, n$ includes numbers corresponding to the case of the original $Y \in \{I, R, D\}$, for every sequential day from 1 to n , and that: $y_i; i=1, \dots, n$ refers to the relating obtained values from the Euler scheme [applicable to (1)] with initial state $\hat{y}_1 - y_1$.

$$MXSE(Y) = \max_{i=1, \dots, n} e_i^2 \quad (13)$$

$$MSE(Y) = \frac{1}{n} \sum_{i=1}^n e_i^2 \quad (14)$$

$$MAE(Y) = \frac{1}{n} \sum_{i=1}^n |e_i| \quad (15)$$

$$MAPE(Y) = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{e_i}{y_i} \right| \quad (16)$$

The second method assumes that now, the major goal function is influenced by all 3 sections, particularly I , R , and D . It is worth noting that its second option is significantly more computationally demanding. Because each compartment of the system may well have data across distinct categories, the relative error of the proposed low-level cost equations with the exception of the MAPE example may be significantly different. As a result, we were making it easy correct normalization to treat each section identically. Then, have a look at the latest set of services.

$$f_Y(y) = \frac{y - \min_{i=1, \dots, n} y_i}{\max_{i=1, \dots, n} y_i - \min_{i=1, \dots, n} y_i} \quad (17)$$

where $y_i; i=1, \dots, n$ signifies a series of observed Y values from consecutive weeks $1, \dots, n$. For divisions I , R , and D , the

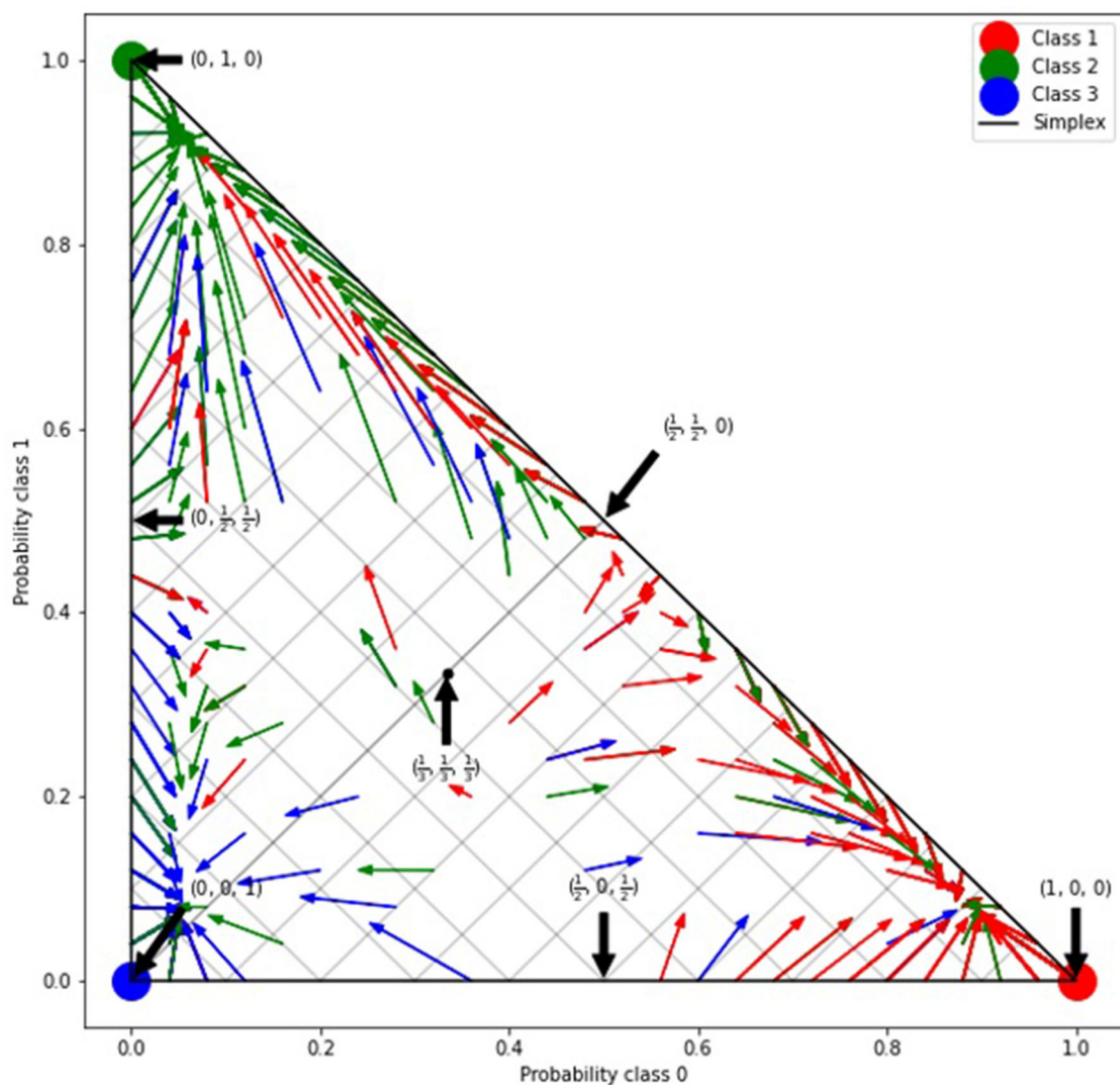


FIGURE 6 | Following calibration using isotonic and SIRD model, change in expected probability on test samples.

proposed methods were utilized to rescale not just given numbers but also values acquired from the Euler scheme.

We require an independent benchmarking data variable. Finally fitness correction is evaluated and performance is compared with other optimal solutions of optimization methods. Consequently, one of them might be our first choice. In these trials, we opted to use an R^2 value to assess the matching quality of low-level cost models that were not described.

$$R^2(YE) = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (18)$$

This was generally determined using only the $Y = D$ variable. Although we employ many optimization methods, the finest potential segment D forecast is our end objective and that is the primary reason for concentrating on this section of the design.

Post-hoc Explainability

The Shapley Additive exPlanations (SHAPs) are used in this research to evaluate the dataset's characteristics in terms of its effects just on the last machine learning results. SHAP uses a coalitional game model to compute optimal Shapley ratios. These numbers indicate how evenly the impact on the model's predictions is divided across the dataset's characteristics. Next, in terms of explaining how well this forecast was made, SHAP creates a mini-explainer architecture that correlates with just a single-row forecast pairing.

RESULT ANALYSIS

Employing healthcare data collected from the dataset, the research framework was used to predict the changes in the emotional experience of children and teens who have

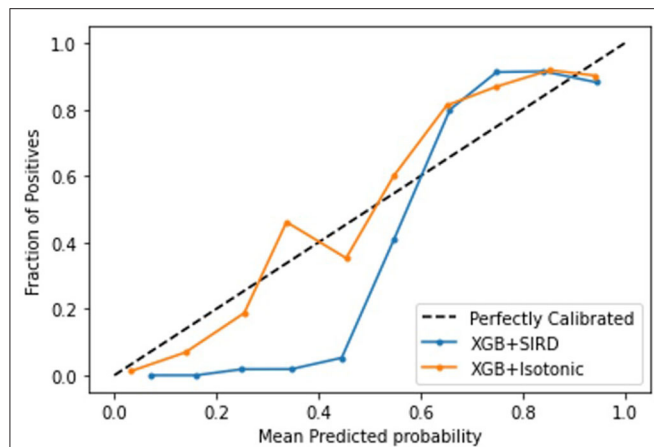


FIGURE 7 | XG Boost classifier calibration graph for class 0.

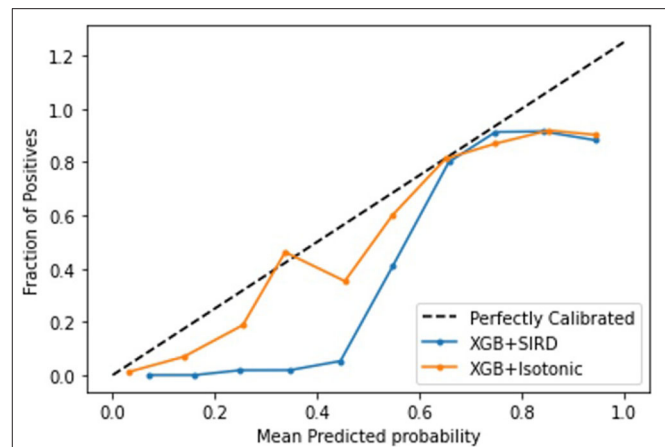


FIGURE 9 | XG Boost classifier calibration graph for class 3.

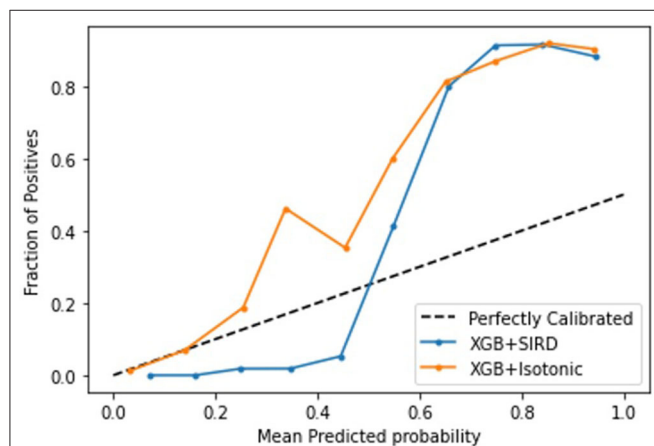


FIGURE 8 | XG Boost classifier calibration graph for class 1.

been diagnosed with chronic disorders. During a calibration procedure, the best-performing clustering approach is reviewed, and multiple prediction methods are assessed based on the outcomes of the feature selection technique to identify the best-performing depending on the accuracy metric. A *post hoc* explainability study is conducted just on best-performing calibration to obtain a greater understanding and interpretations of the most significant contributing aspects to the model's outputs.

Environmental Setup

The Silhouette coefficient, the Calinski–Harabasz Index, and the Davies–Bouldin Index are the three clustering determining factors that are integrated. To calculate a cumulative average rating, the normalized values of the assessment criteria are added together. The clustering algorithms were used with standard system parameters from the sklearn.cluster modules (<https://scikit-learn.org/stable/modules/classes.html#module-sklearn.cluster>, viewed on 1 August 2021), whereas the model variables were seen in Table 2. The feature selection is then conducted just

on 3 clusters obtained by the predominant clustering algorithm using ReliefF, L1-Norm, and chi-square.

The minority classes were classified using a repeating stratification 5-fold cross-validation using grid search using the SMOTE technique—oversampling to the training sample. Components of characteristics with growing complexity were used to assess the predictive model. The reliability of the prediction methods was selected as the method of evaluating their effectiveness.

Feature Selection

Figure 4 shows the spider graph with the number of attributes from every category for the first 40 characteristics where the greatest result was attained, resulting from ReliefF, chi-square, and L1-Norm. The features are listed in Table 2 (Ntakolia et al., 2022).

Classification and Calibration

The efficiency of the comparison prediction methods for the range of features is shown in Figure 5. Table 3 illustrates the optimum value accuracy of every estimation method employed in the various experiments, as well as the number of characteristics where it was attained.

We calibrate the XG Boost model using the SIRD model and compare it to current models such as isotonic regression to improve its efficiency. To evaluate the models, we are using the logistic regression loss (Log-loss) as well as the correctness. Table 4 demonstrates the results of using SIRD, and isotonic regression to calibrate the XG Boost classifier.

Figure 6 measures the variation in projected probability on sample material following isotonic regression and SIRD calibration, correspondingly. True categories 0, 1, and 2 are represented by the red, green, and blue hues of an arrow, accordingly. Class 0, class 1, and class 2 participants, alternately, have a negative, neutral, or positive alteration in their emotional experience. The calibration charts for every class above all others are shown in Figures 7–9.

DISCUSSION

Feature Selection

The findings indicated that social life factors had a considerable impact on predicting outcomes (Table 1, Figure 4). Furthermore, the spider graph (Figure 4) shows that nine characteristics from people's social groups were in the top 40 most important qualities. Moreover, with six elements inside the attribute selection subgroup, everyday routines are the second least important category. Finally, behavioral and demographic effects each provide five characteristics. Health treatment, sleep pattern, medical problems, home life, and private life are the other characteristics (Figure 4). The foregoing findings clearly show that characteristics in all areas are required to properly predict the effect of COVID-19 on the emotional responses of children and teenagers.

Post-hoc Explainability

The mood change, or the changing in emotional experience prior to actually and after the first lockdown in Greece, was chosen as the predicted variable in this research. The findings revealed that the family's perspectives as to whether the COVID-19 emergency caused changes for the better in their child's growth (2 w positive), their family life (2 w relationships family), and their psychological health assessment prior to the COVID-19 emergency (3 m tv) were all strongly linked to a child's emotional experience. Furthermore, during the 3 months prior to and 2 weeks just after lockdown, the amount of time spent watching television or using electronic devices by children increased significantly. As a result, we can see that lockdown had a harmful effect on children who did not originally spend significant time watching television but now do. It is worth noting that the first diagnostic given by a healthcare practitioner played a significant effect in the child's personality change.

CONCLUSION

An understandable machine learning pathway was developed in this work to study and identify its most essential factors related to the emotional state shifts of children and teens throughout Greece's first lockdown. The focus of this research is to discover and explain the elements that influenced the psychological health of the studied group throughout the first COVID-19-related lockdown using the chosen ML workflow. As a result, the issue was defined as a three-class classification issue to detect the

variations in the emotional experience of the persons under investigation. Persons with positive (class 0) and negative (class 2) mood alterations, as well as those with no major change in emotional experience, were divided into classes (class 1). To find the optimum clustering approach and prediction model for this topic, a detailed comparative evaluation was conducted. The Jenks approach was chosen for clustering, followed by feature selection using ReliefF, chi-square, and L1-Norm. After that, the best-performing prediction, the XG Boost with SIRD and isotonic model, was utilized for calibration and a *post-hoc* explainability study to validate the key characteristics that contributed to a model's estimate places. Furthermore, the impact of every characteristic on the various classes was discussed.

We can conclude that the positive developments in a child's life as a result of the first lockdown—family relationships, time was spent watching TV, parenting practices assessment of the child's mental health, and stress created by COVID-19 restrictions—could play a critical role in the change in the child's mood state. These findings are consistent with those of previous research that used pre-pandemic healthcare instances or population-based groups of children who are at risk of transitioning between subclinical to medically severe levels of psychosis. The findings of this study can be effectively employed for the primary requirements of physicians which can be properly prepared for future emergencies or restrictions.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author/s.

AUTHOR CONTRIBUTIONS

Both authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This work was supported by the Social Science Research Planning Project of Jiangxi Province, China, "The Influence of Moral Emotion on Online Helping Behavior of College Students" (19JY46).

REFERENCES

- Abas, M. A., Weobong, B., Burgess, R. A., Kienzler, H., Jack, H. E., Kidia, K., et al. (2021). COVID-19 and global mental health. *Lancet Psychiatry* 8, 458–459. doi: 10.1016/S2215-0366(21)00155-3
- Anchang, J. Y., Ananga, E. O., and Pu, R. (2016). An efficient unsupervised index based approach for mapping urban vegetation from IKONOS imagery. *Int. J. Appl. Earth Obs. Geoinf.* 50, 211–220. doi: 10.1016/j.jag.2016.04.001
- Ansari, J., and Malekshah, S. (2019). A joint energy and reserve scheduling framework based on network reliability using smart grids applications. *Int. Trans. Electr. Energy Syst.* 29, e12096. doi: 10.1002/2050-7038.12096
- Błaszczak, P., Klimczak, K., Mahdi, A., Oprocha, P., Potorski, P., Przybyłowicz, P., et al. (2022). On automatic calibration of the SIRD epidemiological model for COVID-19 data in Poland. *arXiv [Preprint]*. arXiv: 2204.12346. doi: 10.48550/arXiv.2204.12346
- Bonardi, J.-P., Gallea, Q., Kalanoski, D., and Lalive, R. (2020). Fast and local: how did lockdown policies affect the spread and severity of the covid-19. *Covid Econ.* 23, 325–351. Available online at: <https://cepr.org/sites/default/files/news/CovidEconomics23.pdf>
- Chakraborty, I., and Maity, P. (2020). COVID-19 outbreak: Migration, effects on society, global environment and prevention. *Sci. Total Environ.* 728, 138882. doi: 10.1016/j.scitotenv.2020.138882

- Cost, K. T., Crosbie, J., Anagnostou, E., Birken, C. S., Charach, A., Monga, S., et al. (2021). Mostly worse, occasionally better: impact of COVID-19 pandemic on the mental health of Canadian children and adolescents. *Eur. Child Adolesc. Psychiatry* 31, 671–684. doi: 10.1007/s00787-021-01744-3
- Demir, F., Siddique, K., Alswaiti, M., Demir, K., and Sengur, A. (2022). A simple and effective approach based on a multi-level feature selection for automated Parkinson's disease detection. *J. Pers. Med.* 12, 55. doi: 10.3390/jpm12010055
- Demir, F., Turkoglu, M., Aslan, M., and Sengur, A. (2020). A new pyramidal concatenated CNN approach for environmental sound classification. *Appl. Acoust.* 170, 107520. doi: 10.1016/j.apacoust.2020.107520
- Garcia de Avila, M. A., Hamamoto Filho, P. T., Jacob, F. L. D. S., Alcantara, L. R. S., Berghammer, M., Jenholt Nolbris, M., et al. (2020). Children's anxiety and factors related to the COVID-19 pandemic: an exploratory study using the children's anxiety questionnaire and the numerical rating scale. *Int. J. Environ. Res. Public Health* 17, 5757. doi: 10.3390/ijerph17165757
- Ge, F., Di Zhang, L. W., and Mu, H. (2020). Predicting psychological state among Chinese undergraduate students in the COVID-19 epidemic: a longitudinal study using a machine learning. *Neuropsychiatr. Dis. Treat.* 16, 2111. doi: 10.2147/NDT.S262004
- Gomathi, R., Maheswaran, S., Sathesh, S., and Indhumathi, N. (2021). Covid-19 impact on various sectors in india-a detailed analysis. *NVEO* 8, 227–235. Available online at: <https://www.nveo.org/index.php/journal/article/view/348>
- Gomathi, R. D., Radhika, S., Maheswaran, S., Sathesh, S., and Savitha Sri, N. (2020). Impact of Covid-19 in engineering online mode educational system. *Adv. Comput. Commun. Autom. Biomed. Technol.* Available online at: https://www.researchgate.net/publication/348353530_Impact_of_Covid-19_in_Engineering_Online_Mode_Educational_System
- Guo, S., Guo, D., Chen, L., and Jiang, Q. (2017). A L1-regularized feature selection method for local dimension reduction on microarray data. *Comput. Biol. Chem.* 67, 92–101. doi: 10.1016/j.compbiolchem.2016.12.010
- Guru, D., Suhil, M., Raju, L. N., and Kumar, N. V. (2018). An alternative framework for univariate filter based feature selection for text categorization. *Pattern Recognit. Lett.* 103, 23–31. doi: 10.1016/j.patrec.2017.12.025
- Haq, A. U., Li, J. P., Memon, M. H., Malik, A., Ahmad, T., Ali, A., et al. (2019). Feature selection based on L1-norm support vector machine and effective recognition system for Parkinson's disease using voice recordings. *IEEE Access* 7, 37718–37734. doi: 10.1109/ACCESS.2019.2906350
- Islam, M. A., Barna, S. D., Raihan, H., Khan, M. N. A., and Hossain, M. T. (2020). Depression and anxiety among University students during the COVID-19 pandemic in Bangladesh: a web-based cross-sectional survey. *PLoS ONE* 15, e0238162. doi: 10.1371/journal.pone.0238162
- Liang, L., Ren, H., Cao, R., Hu, Y., Qin, Z., Li, C., et al. (2020). The effect of COVID-19 on youth mental health. *Psychiatr. Q.* 91, 841–852. doi: 10.1007/s11126-020-09744-3
- Ma, Z., Idris, S., Zhang, Y., Zewen, L., Wali, A., Ji, Y., et al. (2021). The impact of COVID-19 pandemic outbreak on education and mental health of Chinese children aged 7–15 years: an online survey. *BMC Pediatr.* 21, 1–8. doi: 10.1186/s12887-021-02550-1
- Ma, Z., Zhao, J., Li, Y., Chen, D., Wang, T., Zhang, Z., et al. (2020). Mental health problems and correlates among 746 217 college students during the coronavirus disease 2019 outbreak in China. *Epidemiol. Psychiatr. Sci.* 29, e181. doi: 10.1017/S2045796020000931
- Malekshah, S., and Javad, A. (2021). A novel decentralized method based on the system engineering concept for reliability-security constraint unit commitment in restructured power environment. *Int. J. Energy Res.* 45, 703–726. doi: 10.1002/er.5802
- Malekshah, S., Rasouli, A., Malekshah, Y., Ramezani, A., and Malekshah, A. (2022). Reliability-driven distribution power network dynamic reconfiguration in presence of distributed generation by the deep reinforcement learning method. *Alex. Eng. J.* 61, 6541–6556. doi: 10.1016/j.aej.2021.12.012
- Masten, A. S. (2021). Resilience of children in disasters: a multisystem perspective. *Int. J. Psychol.* 56, 1–11. doi: 10.1002/ijop.12737
- McKune, S. L., Acosta, D., Diaz, N., Brittain, K., Beaulieu, D. J., Maurelli, A. T., et al. (2021). Psychosocial health of school-aged children during the initial COVID-19 safer-at-home school mandates in Florida: a cross-sectional study. *BMC Public Health* 21, 1–11. doi: 10.1186/s12889-021-10540-2
- Murtagh, F., and Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: which algorithms implement Ward's criterion? *J. Classif.* 31, 274–295. doi: 10.1007/s00357-014-9161-z
- North, M. A. (2009). "A method for implementing a statistically significant number of data classes in the jenks algorithm," in *2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery* (Tianjin: IEEE Press). doi: 10.1109/FSKD.2009.319
- Ntakolia, C., Priftis, D., Charakopoulou-Travlou, M., Rannou, I., Magklara, K., Giannopoulou, I., et al. (2022). An explainable machine learning approach for COVID-19's impact on mood states of children and adolescents during the first lockdown in Greece. *Healthcare* 10, 149. doi: 10.3390/healthcare10010149
- Peng, K., Leung, V. C., and Huang, Q. (2018). Clustering approach based on mini batch kmeans for intrusion detection system over big data. *IEEE Access* 6, 11897–11906. doi: 10.1109/ACCESS.2018.2810267
- Penner, F., Ortiz, J. H., and Sharp, C. (2021). Change in youth mental health during the COVID-19 pandemic in a majority Hispanic/Latinx US sample. *J. Am. Acad. Child Adolesc. Psychiatry* 60, 513–523. doi: 10.1016/j.jaac.2020.12.027
- Prati, G., and Mancini, A. D. (2021). The psychological impact of COVID-19 pandemic lockdowns: a review and meta-analysis of longitudinal studies and natural experiments. *Psychol. Med.* 51, 201–211. doi: 10.1017/S0033291721000015
- Qin, Z., Shi, L., Xue, Y., Lin, H., Zhang, J., Liang, P., et al. (2021). Prevalence and risk factors associated with self-reported psychological distress among children and adolescents during the COVID-19 pandemic in China. *JAMA Net. Open* 4, e2035487. doi: 10.1001/jamanetworkopen.2020.35487
- Ren, H., Luo, X., Wang, Y., Guo, X., Hou, H., Zhang, Y., et al. (2021). Psychological responses among nurses caring for patients with COVID-19: a comparative study in China. *Transl. Psychiatry* 11, 1–9. doi: 10.1038/s41398-020-00993-1
- Ren, Z., Xin, Y., Ge, J., Zhao, Z., Liu, D., Ho, R., et al. (2021). Psychological impact of COVID-19 on college students after school reopening: a cross-sectional study based on machine learning. *Front. Psychol.* 12, 1346. doi: 10.3389/fpsyg.2021.641806
- Rens, E., Smith, P., Nicaise, P., Lorant, V., and Van den Broeck, K. (2021). Mental distress and its contributing factors among young people during the first wave of COVID-19: a Belgian survey study. *Front. Psychiatry* 12, 575553. doi: 10.3389/fpsyg.2021.575553
- Sciberras, E., Patel, P., Stokes, M. A., Coghill, D., Middeldorp, C. M., Bellgrove, M. A., et al. (2022). Physical health, media use, and mental health in children and adolescents with ADHD during the COVID-19 pandemic in Australia. *J. Atten. Disord.* 26, 549–562. doi: 10.1177/1087054720978549
- Shanmugam, M., Nehru, S., and Shanmugam, S. (2017). A wearable embedded device for chronic low back patients to track lumbar spine position. *Biomed. Res.* S118–S123. doi: 10.4066/biomedicalresearch.29-17-1304
- Tuncer, T., Dogan, S., and Ozyurt, F. (2020). An automated residual exemplar local binary pattern and iterative relief based COVID-19 detection method using chest X-ray image. *Chemometr. Intell. Lab. Syst.* 203, 104054. doi: 10.1016/j.chemolab.2020.104054
- Turkoglu, M. (2021). COVIDetectioNet: COVID-19 diagnosis system based on X-ray images using features selected from pre-learned deep features ensemble. *Appl. Intell.* 51, 1213–1226. doi: 10.1007/s10489-020-01888-w
- Vindegaard, N., and Benros, M. E. (2020). COVID-19 pandemic and mental health consequences: systematic review of the current evidence. *Brain Behav. Immun.* 89, 531–542. doi: 10.1016/j.bbi.2020.05.048
- Vizheh, M., Qorbani, M., Arzaghi, S. M., Muhidin, S., Javanmard, Z., and Esmaeili, M. (2020). The mental health of healthcare workers in the COVID-19 pandemic: a systematic review. *J. Diabetes Metab. Disord.* 19, 1967–1978. doi: 10.1007/s40200-020-00643-9
- Von Luxburg, U. (2007). A tutorial on spectral clustering. *Stat. Comput.* 17, 395–416. doi: 10.1007/s11222-007-9033-z
- Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845
- Wathelet, M., Duhem, S., Vaiva, G., Baubet, T., Habran, E., Veerapa, E., et al. (2020). Factors associated with mental health disorders among University students in France confined during the COVID-19 pandemic. *JAMA Net. Open* 3, e2025591. doi: 10.1001/jamanetworkopen.2020.25591
- WHO, G. (2020). *Statement on the Second Meeting of the International Health Regulations (2005) Emergency Committee Regarding the Outbreak of Novel Coronavirus (2019-nCoV)*. Geneva: World Health Organization.

- Yeasmin, S., Banik, R., Hossain, S., Hossain, M. N., Mahumud, R., Salma, N., et al. (2020). Impact of COVID-19 pandemic on the mental health of children in Bangladesh: a cross-sectional study. *Child. Youth Serv. Rev.* 117, 105277. doi: 10.1016/j.childyouth.2020.105277
- Yim, O., and Ramdeen, K. T. (2015). Hierarchical cluster analysis: comparison of three linkage measures and application to psychological data. *Quant. Meth. Psych.* 11, 8–21. doi: 10.20982/tqmp.11.1.p008
- Zhang, L., Zhang, X., Yuan, S., and Wang, K. (2021). Economic, social, and ecological impact evaluation of traffic network in Beijing–Tianjin–Hebei urban agglomeration based on the entropy weight TOPSIS Method. *Sustainability* 13, 1862. doi: 10.3390/su13041862
- Zhou, S.-J., Wang, L.-L., Yang, R., Yang, X.-J., Zhang, L.-G., Guo, Z.-C., et al. (2020). Sleep problems among Chinese adolescents and young adults during the coronavirus-2019 pandemic. *Sleep Med.* 74, 39–47. doi: 10.1016/j.sleep.2020.06.001

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Luo and Huang. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.