Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Article

An explainable machine learning approach for COVID-19 impact on mood state of children and adolescents during the 1st lockdown in Greece

Charis Ntakolia[1,*], Dimitrios Priftis[1], Mariana Charakopoulou-Travlou[1], Ioanna Rannou[1], Konstantina Magklara[2], Ioanna Giannopoulou[3], Konstantinos Kotsis[4], Aspasia Serdari[5], Emmanouil Tsalamanios[6], Aliki Grigoriadou[7], Konstantina Ladopoulou[8], Iouliani Koullourou[9], Neda Sadeghi[10], Georgia O'Callaghan[10], Argyris Stringaris[10], Eleni Lazaratou[2]

[1] University Mental Health Research Institute, Athens, Greece; charis.nt@gmail.com; icedale@gmail.com; mariana.har.travlos@gmail.com; ioannarannou@gmail.com
[2] First Psychiatric Department, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece, nadia.magklara@gmail.com; elazar@med.uoa.gr
[3] Second Psychiatric Department, 'Attikon' University Hospital, National and Kapodistrian University of Athens, Athens, Greece, igianno@med.uoa.gr; igioannag@gmail.com
[4] Department of Psychiatry, Faculty of Medicine, School of Health Sciences, University of Ioannina, Ioannina, Greece, konkotsis@gmail.com
[5] Department of Child and Adolescent Psychiatry, Medical School, Democritus University of Thrace, University Hospital of Alexandroupolis, Alexandroupolis, Greece, aserdari@yahoo.com; aserntar@med.duth.gr
[6] Department of Child and Adolescent Psychiatry, Division of Psychiatry, 'Asklepieion Voulas' General Hospital, Attica, Greece, emtsalamanios@hotmail.com
[7] Hellenic Centre for Mental Health and Research, Athens, Greece, alikigrigoriadou@gmail.com
[8] Athens Child and Adolescent Mental Health Centre, General Children's Hospital 'Pan. & Aglaia Kyriakou', Athens, Greece, kladopou@gmail.com
[9] Mental Health Center, General Hospital 'G. Hatzikosta', Ioannina, Greece, jkoullourou@gmail.com
[10] Section of Clinical and Computational Psychiatry, National Institute of Mental Health, National Institutes of Health, Bethesda, MD, USA, neda.sadeghi@nih.gov; georgiaocallaghan@gmail.com; argyris.stringaris@nih.gov
* Correspondence: charis.nt@gmail.com

Abstract: The global spread of COVID-19 led the World Health Organization to declare a pandemic on 11 March 2020. To decelerate this spread, countries have taken strict measures that affected the lifestyle and economy. Various studies have been focused on the identification of COVID-19 impact to mental health of children and adolescents via traditional statistical approaches. However, a machine learning methodology must be developed to explain the main factors that contribute to the change of mood state of children and adolescents during the first lockdown. Therefore, to this study an explainable machine learning pipeline is presented focusing on children and adolescents in Greece, where a strict lockdown was imposed. The target group consists of children and adolescents, recruited from children and adolescent mental health services, who present mental health problems diagnosed before the pandemic. The proposed methodology imposes: (i) data collection via questionnaires; (ii) a clustering process to identify the groups of subjects with amelioration, deterioration and stability to their mood state; (iii) a feature selection process to identify the most informative features that contribute to mood state prediction; (iv) a decision-making process based on an experimental evaluation among classifiers; (v) calibration of the best performing model and (v) a post-hoc interpretation of the features' impact on the best performing model. The results showed that a blend of heterogeneous features from almost all feature categories is necessary to increase our understanding regarding the effect of COVID-19 pandemic on the mood state of children and adolescents.

Keywords: COVID-19 pandemic; children and adolescents; machine learning; post-hoc explainability; model calibration

© 2021 by the author(s). Distributed under a Creative Commons CC BY license.

1. Introduction

In December 2019 the World Health Organization (WHO) identified the novel coronavirus (COVID-19) as the cause of pneumonia in Wuhan, China, and on the 11th March the WHO declared COVID-19 as a pandemic [1,2]. Between December 31st 2019 and May 4th 2020, over 184 countries adopted strict measures to limit the spread of COVID-19, such as lockdown restrictions and quarantine time, which lead to socioeconomic, environment and mental health challenges. Within those restrictions, specific measures ranged from working from home, online education (e-learning), social restrictions to border closure (

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 1) [3]. Even though the lockdown policies contributed to the control and decrease of the spread of COVID-19, they also resulted to the deterioration of the mental health of the population worldwide [3–5].

A plethora of studies have been conducted to examine the impact of COVID-19 and its restriction policies to the studied population [6–8]. Specifically, multivariable logistic regression analyses were adopted in various studies to: (i) identify the correlations of mental health with other factors [9], such as socio-demographic features [4,10–12] and/or school aspects [12] or health behaviors [13] mostly on university students [14–16]; (ii) assess the prevalence and the risk factors associated with self-reported psychological distress [17]; and (iii) evaluate the effects of COVID-19 measures upon the mental health of children and adolescents, with or without pre-existing diagnoses [18]. Binomial or binary logistic regression analysis was used to: (i) identify sleeping problems of adolescents and young adults (12–29 years) during pandemic [19]; (ii) assess depression and anxiety amongst university students [20]; and (iii) examine the prevalence of anxiety among children and the possible association to COVID-19 [21]. Other studies focused on youths used univariate logistic regression to identify mental health issues [22]. Hierarchical logistic regression analyses were used to examine variables associated with mental health problems during the COVID-19 outbreak to university students [23]. Adjusted logistic regression analyses were used to examine the association between stress due to COVID-19 and worries to children and adolescents [24]. However, limited studies have been employed machine learning prediction models, such as XGBoost model to predict the anxiety and insomnia in undergraduate students during COVID-19 pandemic [25] or Random forest and regression trees to identify predictors of psychological distress during COVID-19 in participants aged 18-85 [26].

Most of the above presented studies are focused on Chinese regions [12,14,23], college students [14,17,23] and used traditional statistical approaches, such as logistic regression and chi-square tests [20–22,24] to identify correlations among risk factors and mental health problems, while only few of them employ machine learning methodologies [26]. Furthermore, to the best our knowledge, there has not been any study focused on children and adolescents with diagnosed mental disorders. Therefore, this study aims to fill this gap by proposing the development of an explainable machine learning pipeline to create a deeper understanding of the consequences and impact of the 1st lockdown in Greece on the mental health of children and adolescents. The study includes 71 heterogenous factors. The proposed methodology consists of: (i) clustering the examined population based on their mood state alteration during lockdown; (ii) identifying the main features that contribute to the mood alteration of the examined population; (iii) developing calibrated machine learning models to predict the alteration of mood state; (iv) post-hoc explainability analysis to rank features in terms of their impact on the final machine learning outputs.

The current study focuses on children and adolescents that have been attending, during the year prior to the pandemic, Children and Adolescents Mental Health Services (CAMHS) in Greece. The study is aligned with the IJMEDI checklist for assessment of medical AI [27].

Table 1. Lockdown policies implemented around the world [3].

	Measures	Explanation
International Measures	Curfew	The effective date when a country announced a restriction on the movement of individuals within a given time of the day
	State of emergency	The effective date when a country announced a state of emergency
	Within country regional lockdown	The effective date when a region within a country announced that it will be entering a total lockdown
	Partial selective lockdown	The earliest effective date for the partial restriction on the movement of people such as through school closures or through limiting the number of people allowed to gather in a group and/or closure of religious institutions
External measures	Selective border closures stage 1	The earliest effective date when a country closed its borders with a region or country significantly affected by COVID-19 (Wuhan, China, Iran, and Italy – individually or as a group)
	Selective border closures stage 2	The earliest effective date after Selective border closure stage 1 when a country closed its border to people from one or multiple other countries in the world significantly affected by COVID-19
	International lockdown	The effective date when a country banned all flights, rail and automotive movements internationally

2. Materials and Methods

To predict the impact of COVID-19 due to the 1st lockdown imposed in Greece during the period March 23rd 2020 to May 4th 2020, we focused on the sensitive group of children and adolescents. The data from the Hellenic COvid-19 imPact survEy (HOPE) were used, a longitudinal study surveying parents of children that have been attending, during the year prior to the pandemic (March 1st, 2019 to March 1st, 2020), CAMHS in Greece (seven in Athens Greater Metropolitan Area, two in Ioannina, one in Alexandroupolis, one in Thessaloniki and one on Crete). A machine learning pipeline is proposed that includes: (i) data collection via questionnaires and medical reports; (ii) data pre-processing; (iii) a competitive evaluation of state-of-the art clustering methods and evaluation metrics; (iv) a feature selection based on a state-of-the-art and robust method, named ReliefF, that has been proven effective to medical data; (v) a competitive evaluation of various ML models following calibration; and (vi) a post-hoc explainability of the best performed model with SHAP to identify the features' impact on the model.

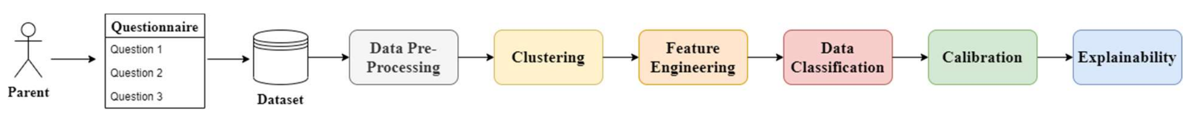


Figure 1. Machine learning pipeline adopted in this study.

3.1 Data collection

To collect the data and form the dataset, children that attended the service of CAMHS participated. Specifically, 744 children whose parents (738 parents) answered on their behalf the online questionnaire participated in this study. This process took place between May 8th and June 1st, 2020. The questionnaire included questions relevant to demographic information, parent's evaluation of the child's condition 3 months (3m) before the lockdown and 2 weeks (2w) after the 1st lockdown in Greece. Table 2 shows the sociodemographic characteristics of the dataset while

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 3 presents the description of the variables used in the study as they were extracted from the questionnaires. Table 4 shows the approval from the participated hospitals' research ethics committee. The study was performed in line with the principles of the Declaration of Helsinki.

Table 2. Sociodemographic characteristics of the dataset

Sociodemographic Characteristics	Population (%)
Age, Mean $\pm$ Standard Deviation	10.7 $\pm$ 4.1
Sex	
Male	466 (63.1%)
Female	273 (36.9%)
Participant parent	
Mother	588 (79.7%)
Father	142 (19.2%)
Other (grandparents, uncle/aunt, foster parents, other)	8 (1.1%)
Child's ethnicity	
Greek	725 (98.2%)
Other	13 (1.8%)
Health insurance type	
National/ Military	650 (87.7%)
Private	63 (8.7%)
Other	9 (1.3%)
None	16 (2.3%)
Residential area	
City	382 (51.8%)
Suburbs of a city	200 (27.1%)
Town/ village	131 (17.7%)
Rural area	10 (1.4%)
Island	15 (2.0%)
Reporting parent's educational level	
Compulsory 9-years education	26 (3.5%)
Senior high school	146 (19.8%)
Institute of Vocational Training	118 (16.0%)
Technical College or University degree	280 (37.9%)
Postgraduate degree (M.Sc./ PhD)	168 (22.8%)
Second parent's educational level	
Compulsory 9-years education	80 (10.8%)
Senior high school	221 (29.9%)
Institute of Vocational Training	105 (14.3%)
Technical College or University degree	211 (28.6%)
Postgraduate degree (M.Sc., PhD)	121 (16.4%)
Essential worker (yes): healthcare, delivery worker, store worker, security, building maintenance	321 (43.5%)
Worker in a facility treating COVID-19 (yes)	105 (14.2%)
Job loss during the pandemic (yes)	38 (5.1%)
Limited ability to earn money (yes)	81 (10.9%)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 3. Dataset description.

Category	Features	Description
Demographics	age_group	Age group of child
	gender child	Gender of child
	parent_area_live	Area of residence
	gender_parent	Gender of the parent or guardian
	parenteducation	Education level of parent or guardian
	school_child	School enrolment and attendance
	2w_essential_worker_r	Whether any adults living with the child are essential workers (health care, delivery services, pharmacies, law enforcement and security, store worker, cleaning services, other)
Social life	3m_outdoors 2w_outdoors	Days per week the child spent outside the house (parks, outdoor spaces) - for 3 months and 2 weeks respectively
	2w_time_outside	Amount of time per week the child spent/dedicated out of the house (e.g. shopping, parks, etc)
	2w_event_cancellat	How difficult the cancellation of important events in the child's life (graduation, vacation, Easter recess) has been for him/her
	2w_recommendations	Difficulty to follow recommendations regarding social distancing
	2w_contact_changed	Change in the contact of the child with people outside home relative to before the coronavirus/COVID-19 crisis
	2w_relationships_friends	Change in the quality of relationships of the child with his/her friends
	3m_soc_media 2w_soc_media	Time spent using social media (e.g Facetime, Facebook, Instagram, Snapchat, Twitter, Tiktok) -for 3 months and 2 weeks respectively
Personal life	2w_positive	Positive changes in the child's life due to the coronavirus/COViD-19 crisis
Family life	Family_impact_any	If any event that affected the family was occurred due to COVID-19
	2w_financial_recod	Financial problems faced by the family due to the Coronavirus/COVID-19 crisis
	2w_relationships_family	Changes to the quality of the relationships between the child and members of his/her family
	2w_family_events_lost_job 2w_family_events_loss_earnings	Whether any of the following have happened to the child's family members because of Coronavirus/COVID-19, such as loss of job and loss of earnings
Daily activities	3m_exercise 2w_exercise	Days per week the child engaged into exercise (e.g., increased heart rate,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

		breathing) for at least 30 minutes - for 3 months and 2 weeks respectively
	2w_video_games 3m_video_games	Time spent playing video games - for 3 months and 2 weeks respectively
	3m_tv 2w_tv	Time spent watching TV or digital means (e.g. Netflix, Youtube, or web surfing) - for 3 months and 2 weeks respectively
	2w_reading	How frequently the child asked questions, read, or talked about Coronavirus/COVID-19
Health concerns	2w_worry_self_infected	Child's worry about getting infected
	2w_worry_family_inf	Child's worry about members of the family or friends getting infected
	2w_worry_phys_healt	Worry that physical health will be affected by the Coronavirus/COVID-19
	2w_worry_ment_health	Worry that the child's mental/emotional health will be affected by the Coronavirus/COVID-19
Behavioral effects	2w_stress_restrict	Stress caused by the curfew
	2w_stress_family	Stress caused to the child by changes in family contacts
	2w_worry_food_reco	Worry about food in the family running out due to loss of income
	2w_stress_social	Stress caused to the child by changes to his/her social contacts
	2w_living_stability	Child's concern about the stability of the family's living situation
	2w_hopeful_end	How hopeful the child is that the Coronavirus/COVID-19 crisis will end
Sleeping habits	3m_sleep_hours 2w_sleep_hours_rec	Average sleep duration on weekdays - for 3 months and 2 weeks respectively
	3m_sleep_time 2w_sleep_time_reco	Sleep schedule on weekdays - for 3 months and 2 weeks respectively
	3m_sleep_hours_weeke 2w_sleep_hours_wee	Average sleep duration on weekends - for 3 months and 2 weeks respectively
	3m_sleep_time_weeken 2w_sleep_time_week	Sleep schedule on weekends - for 3 months and 2 weeks respectively
Medical diagnosis / rehabilitation	2w_child_health_evaluation	Parental evaluation of the child's overall physical health before the Coronavirus/COVID-19 crisis
	2w_mental_health_eval	Parental evaluation of the child's overall mental/emotional health before the Coronavirus/COVID-19 crisis
	diagnosis_1_group	Diagnosis defined by the medical expert
	Diagnosis_FINAL_groups	Final diagnostic category defined by the medical expert
	2w_symptoms_tot	Symptoms the child had
	2w_all_exposure_tot	Child exposed to someone likely to have Coronavirus/COVID-19

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

	2w_support_activit	Supports which were in place for the child and have been disrupted
	2w_family_diagnosis	Whether any members of the child's family have been diagnosed with COVID-19
	2w_family_events_ho	Whether any of the following have happened to the child's family members because of Coronavirus/COVID-19, such as hospitalization, self-quarantine, death, physical illness, and total number of the above family events
	2w_family_events_qu	
	2w_family_events_di	
	2w_family_events_il	
	2w_family_events_to	
Mood state	2w_general_worry	How worried the child generally was during the past two weeks
	2w_sadness	How happy versus sad the child was during the past two weeks
	2w_anxiety	How relaxed versus anxious the child was during the past two weeks
	2w_restlessness	How fidgety or restless the child was during the past two weeks
	2w_anhedonia	Ability of the child to enjoy his/her usual activities
	2w_loneliness	How lonely the child was during the past two weeks
	2w_irritability	How irritable or easily angered the child was during the past two weeks
	2w_concentration	How well the child was able to concentrate or focus
	2w_tiredness	How fatigued or tired the child was during the past two weeks
	2w_rumination	How often the child was expressing negative thoughts during the past two weeks

Table 4. Approval from the research ethics committee of the hospitals particiapted in the stsudy.

Hospital	Ethic Committee Name	Approval Code	Approval Date
Hellenic Centre for Mental Health and Research,	Board of the HCMHR	1016	27 April 2020
University Hospital of Ioannina	Scientific Committee	10	03 June 2020
University Hospital of Alexandroupolis	Ethics Committee, General University Hospital of Alexandroupolis	ΕΣ6/ΘΕΜΑ10/10-05-20	21 January 2021
Asklepieio Voulas General Hospital	Scientific Counsil	8353/6	20 May 2020
Attikon General University Hospital	Ethics Committte	ΒΨΥΧ, 6265/19-2-2020	17 Mars 2020
General Hospital 'G. Hatzikosta'	Scientific Committee	8/ 28-05-2020	28 May 2020
General Children's Hospital 'Pan. & Aglaia Kyriakou'	Ethics Committte	11816/24-06-2020	24 June 2020

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Aiginition Hospital	Ethics Committee Aiginiteion Hospital	628Σ46Ψ8N2-8ΤΩ 221/ 27-05-2020	27-05-20

3.2 Data Pre-processing

Data imputation was not needed since there was no missing values of categorical or numerical variables in the final dataset. Furthermore, as a common requirement for many ML classifiers, the standardization of the dataset was implemented.

3.3 Clustering methods

For the clustering process, six popular methods were employed, such as Mini Batch K-Means [28,28], Spectral Clustering [29], Ward [30,31], Average Linkage [32,33], Balanced Iterative Reducing and Clustering using Hierarchies (Birch) [34,35], and Jenks natural breaks optimization method (Jenks) [36–38]. Clustering was performed on the values of the variable that represents the change of mood state (Figure 2). Specifically, the mood state score is calculated by the sum of the variables 2w_general_worry, 2w_sadness, 2w_anxiety, 2w_restlessness, 2w_anhedonia, 2w_loneliness, 2w_irritability, 2w_concentration, 2w_tiredness and 2w_rumination (

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 3). The change of the mood state is the difference between their mood state score 2 weeks after the 1st lockdown and 3 months before the 1st lockdown in Greece. Hence, a negative value indicates a positive change of the patient's mood state, while a positive value indicates a negative change of the patient's mood state.

3.4 Feature engineering

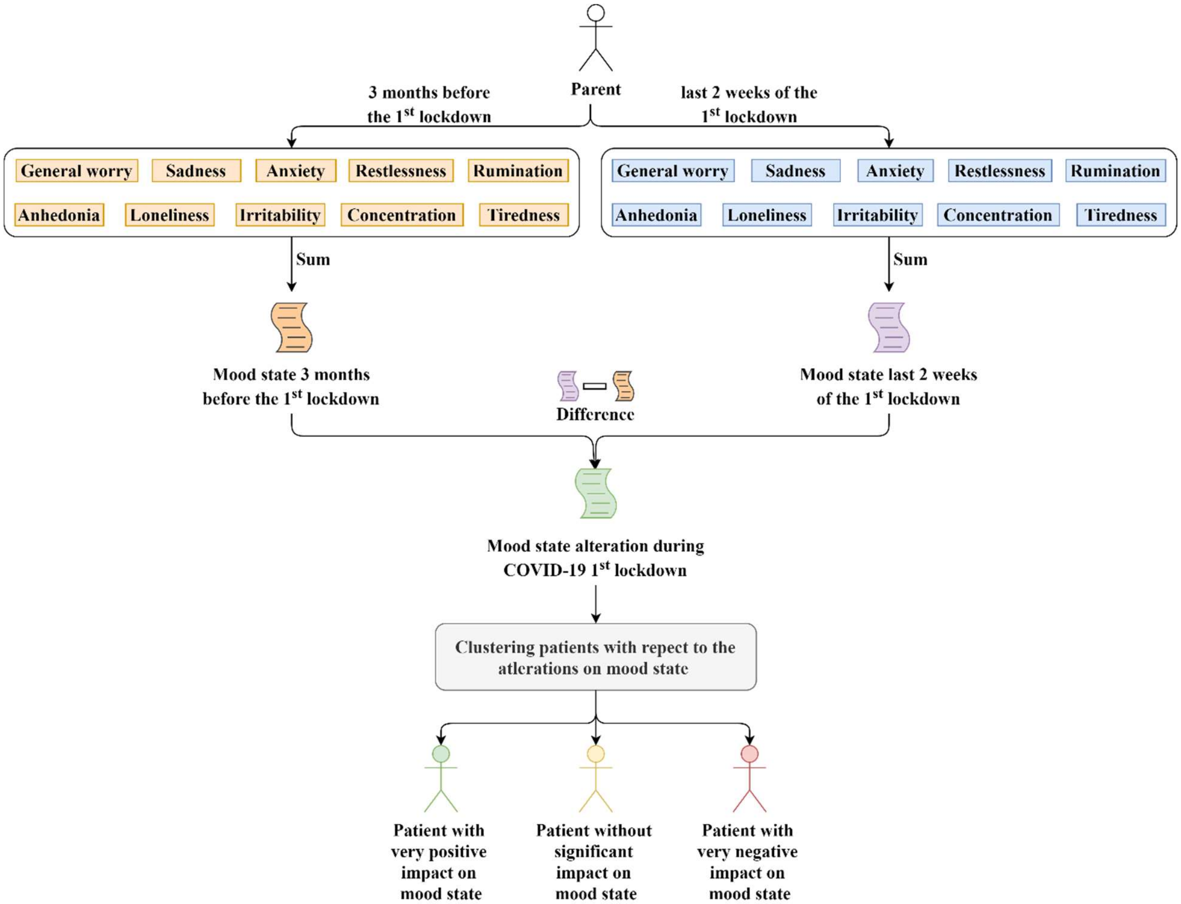
The feature selection process was performed by using the ReliefF algorithm due to its effectiveness to medical diagnosis and classification medical problems [39–43]. ReliefF is an extension of the original Relief which can deal with multiclass problems due to the enhancement with noise resistance [44,45] and therefore it is considered suitable for the current medical multi-class classification problem as it is defined in section 3.3, Figure 2.

3.5 Data Classification

To solve the defined multiclass classification problem seven popular classifiers are employed and tested: Random Forest, Multi-Layer Perceptron, XG Boost, Logistic Regression, Support Vector Machine, K-Nearest Neighbor and Decision Trees. The adopted models are frequently used for medical classification problems while covering various types of prediction models such as tree-based, linear or neural networks [46–50].

3.6 Post-Hoc Explainability

In the current study the SHapley Additive exPlanations (SHAP) is employed to rank the features of the dataset with respect to their impact on the final machine learning outputs. SHAP calculates optimal shapley values from coalitional game theory. These values show how fairly the impact on model's prediction is distributed among the features of the dataset. Then, SHAP develops a mini explainer model that corresponds to a single row-prediction pair in order to explain how this prediction was achieved [51].



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Figure 2. Clustering process.

3. Results

3.1. Evaluation Methodology

The proposed methodology was applied in the context of predicting the change of the mood state in children and youths that are diagnosed with a mental disease by using the medical data derived from the dataset (section 3.1). Initially, an evaluation for the best performed clustering method is performed, then based on the results of the feature selection method, various prediction models are evaluated to choose the best performed based on the accuracy metric following a calibration process. For the best performing calibrated model, a post-hoc explainability analysis is performed for a deeper understanding and interpretation of the most contributing features to model's output (Figure 3).

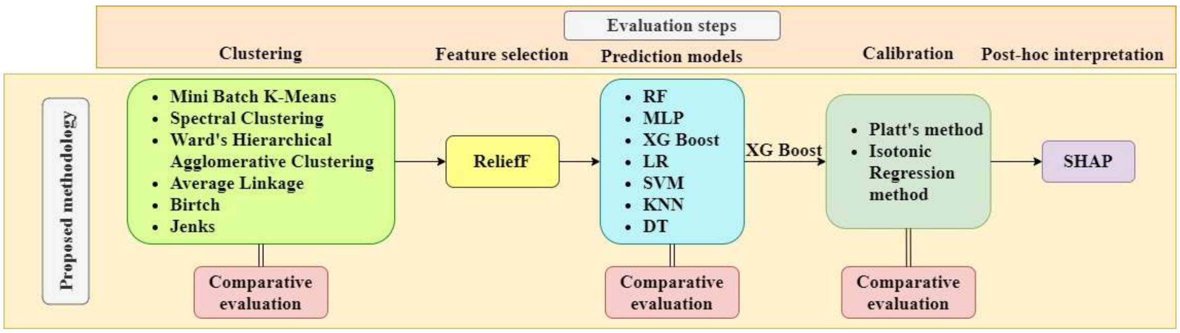


Figure 3. Evaluation methodology.

The three clustering evaluation criteria that are combined are the Silhouette Coefficient, the Calinski Harabasz Index and the Davies Boulding Index. Specifically, the normalized scores of the evaluation criteria are summed for calculating a cumulative evaluation score (Figure 4). The default parameter settings from sklearn.cluster module[1] were used for the clustering methods while the additional settings are shown in Table 5. Then, the feature selection is performed with ReliefF on the three clusters derived by the prevailing clustering method (Figure 4). For the classification, a repeated stratified 5-fold cross validation with grid search was adopted with SMOTE method [52,53] – oversampling to training dataset for the minority classes. The prediction models were evaluated in subsets of features with increasing dimensionality. The accuracy was chosen as the evaluation criterion for the performance of the prediction models.

[1] https://scikit-learn.org/stable/modules/classes.html#module-sklearn.cluster

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 6 presents the hyperparameters of the classification models for tuning.

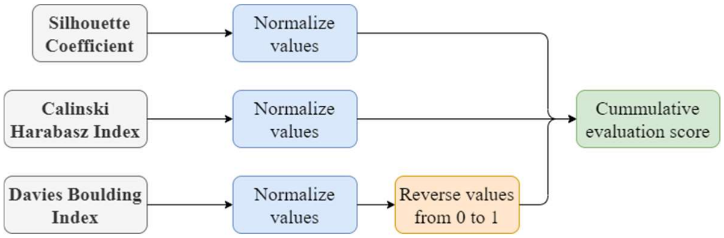


Figure 4. Evaluation process of clustering methods.

Table 5. Parameter settings for clustering methods.

Clustering method	Parameter settings
Mini Batch K-Means	3 classes
Spectral Clustering	3 classes, arpack eigen solver, nearest_neighbors affinity
Ward's Hierarchical Agglomerative Clustering	3 classes, ward linkage, symmetric connectivity
Average Linkage	3 classes, average linkage, cityblock affinity, symmetric connectivity
Birch	3 classes
Jenks	3 classes, include lowest value

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 6. Hyper parameter settings for tuning.

Classification model	Hyper Parameters Tuning
Random Forest	n_estimators = [int(x) for x in np.linspace(start=10, stop=500, num=10)]; max_features = ['auto', 'sqrt']; max_depth = [int(x) for x in np.linspace(3, 10, num=1)]; min_samples_split = [3, 4, 5, 6, 7, 10]; min_samples_leaf = [1, 2, 4]; bootstrap = [True, False].
Multi-Layer Perceptron	hidden_layer_sizes = [(2, 5, 10), (5, 10, 20), (10, 20, 50)]; activation = ['tanh', 'relu']; solver = ['sgd', 'adam']; alpha = [0.0001,0.05]; learning_rate = ['constant', 'adaptive']
XG Boost	max_depth = [2, 3, 4, 5, 6, 7, 8]; min_child_weight = [1, 2, 3, 4, 5, 6]; gamma = [0, 0.4, 0.5, 0.6]
Logistic Regression	C = [0.001, 0.01, 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]; warm_star = [True, False]; multi_class = ['ovr', 'multinomial']; solver = ['newton-cg', 'lbfgs', 'sag', 'saga']
Support Vector Machine	C= [0.001, 0.01, 0.1, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10]; kernel = ['linear', 'sigmoid', 'rbf', 'poly']
K-Nearest Neighbor	n_neighbors = [5, 7, 9, 12, 14, 15, 16, 17]; leaf_size= [1, 2, 3, 5]; weights = ['uniform', 'distance']; algorithm = ['auto', 'ball_tree', 'kd_tree', 'brute']
Decision Trees	max_features = ['auto', 'sqrt', 'log2']; min_samples_split = [2, 3, 4, 5, 6, 7, 8, 10, 12, 15]; min_samples_leaf = [1, 2, 3, 4, 5, 6, 7, 8, 10]

3.2 Results

3.2.1 Clustering

Table 7 shows the results from the clustering methods that were employed to group the population among the individuals with positive change to their mood state (Cluster 0), without significant change (Cluster 1) and with negative change (Cluster 2).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 8 shows the evaluation score achieved by each clustering method.

Table 7. Clustering results

Clustering methods	Cluster information	Clusters		
		Cluster 0	Cluster 1	Cluster 2
Mini Batch K-Means	Set	[-24, -4]	[-3, 4]	[5, 25]
	Number of elements	144	468	132
Spectral Clustering	Set	Unable to create continuous sets		
	Number of elements	485	230	29
Ward	Set	[-24, -7]	[-6, 1]	[2, 25]
	Number of elements	66	418	260
Average Linkage	Set	[-24, -7]	[-6, 4]	[5, 25]
	Number of elements	66	546	132
Birch	Set	[-24, -6]	[-5, 8]	[9, 25]
	Number of elements	80	608	56
Jenks	Set	[-24, -5]	[-4, 3]	[4, 25]
	Number of elements	106	469	169

Table 8. Evaluation of clustering methods.

Clustering method	Evaluation Method Silhouette Coefficient	Calinski Harabasz Index	Davies Boulding Index	Cumulative normalized score
Mini Batch K-Means	0.55	1106.78	0.60	2.94
Spectral Clustering	0.12	24.95	14.79	0.00
Ward	0.54	989.18	0.58	2.80
Average Linkage	0.57	1048.06	0.52	2.94
Birch	0.55	784.60	0.49	2.64
Jenks	0.56	1112.73	0.58	**2.96**

3.2.2 Feature selection

Table 9 shows the 40 most significant features of our dataset derived from ReliefF, while Figure 5 illustrates the number of features per feature category.

Table 9. Results from feature selection with the categories of the 40 first features.

Features	Category	Features	Category
1st feature	Social life	21st feature	Daily activities
2nd feature	Behavioral effects	22nd feature	Behavioral effects
3rd feature	Medical diagnosis / rehabilitation	23rd feature	Behavioral effects
4th feature	Social life	24th feature	Social life
5th feature	Personal life	25th feature	Daily activities
6th feature	Medical diagnosis / rehabilitation	26th feature	Daily activities
7th feature	Demographics	27th feature	Medical diagnosis / rehabilitation
8th feature	Family life	28th feature	Demographics
9th feature	Family life	29th feature	Behavioral effects
10th feature	Social life	30th feature	Health concerns
11th feature	Social life	31st feature	Sleeping habits
12th feature	Daily activities	32nd feature	Social life
13th feature	Daily activities	33rd feature	Demographics
14th feature	Health concerns	34th feature	Social life
15th feature	Daily activities	35th feature	Medical diagnosis / rehabilitation
16th feature	Health concerns	36th feature	Social life
17th feature	Demographics	37th feature	Sleeping habits
18th feature	Behavioral effects	38th feature	Sleeping habits
19th feature	Social life	39th feature	Sleeping habits
20th feature	Health concerns	40th feature	Demographics

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

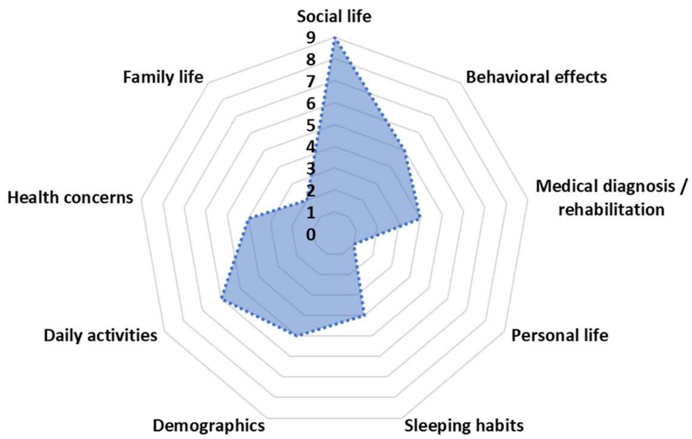


Figure 5. Spider plot of the number of features that belongs to each feature category for the first 40 features where the best performance was achieved.

3.2.3 Classification and calibration

Figure 6 illustrates the accuracy of the comparative prediction models per number of features. Table 10 shows the maximum achieved accuracy of each prediction model used in the experimental evaluation and the number of features where the maximum accuracy was reached.

To increase the performance of XG Boost model, we perform calibration with Isotonic Regression and Platt's methods. We use the logistic regression loss (Log-loss) and the accuracy to evaluate the models. Table 11 shows the results after XG Boost classifier calibration with Isotonic Regression and Platt's methods. Figure 7a,b depict the change of predicted probabilities on test samples after calibration with Isotonic Regression and Platt's (sigmoid) methods, respectively. The red, green and blue colors of an arrow represent the true classes 0, 1 and 2 respectively. Class 0, class 1 and class 2 represent the patients with negative, neutral and positive change on their mood state, respectively. Figure 8a,b depicts the learned calibration maps. The learned calibration map consists of a grid of possible uncalibrated probabilities over the 2-simplex by computing the corresponding calibrated probabilities and plot arrows for each. The arrows are colored according to the highest uncalibrated probability. Figure 9, Figure 10 and Figure 11 illustrate the calibration plots for each class over the others.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

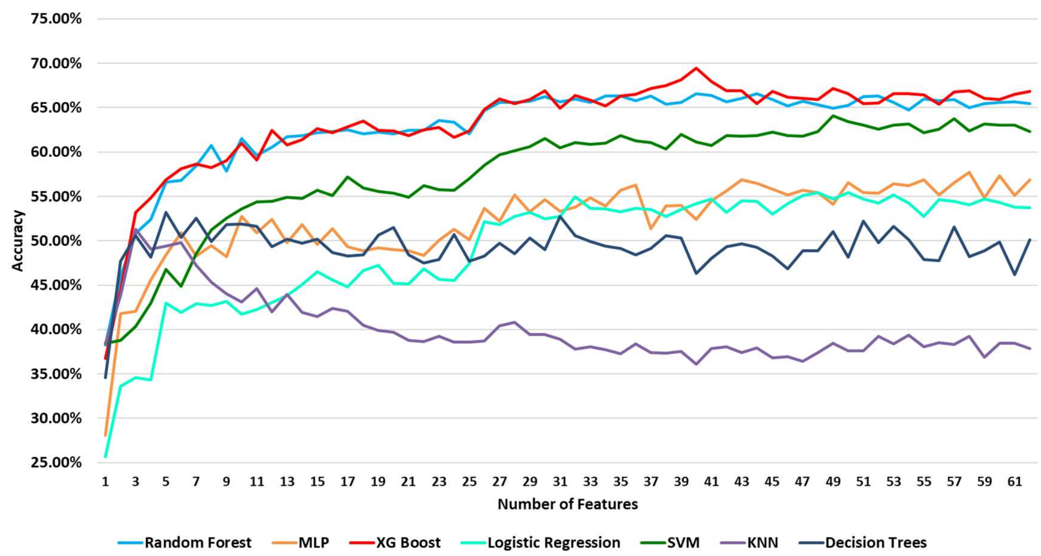


Figure 6. Classification results.

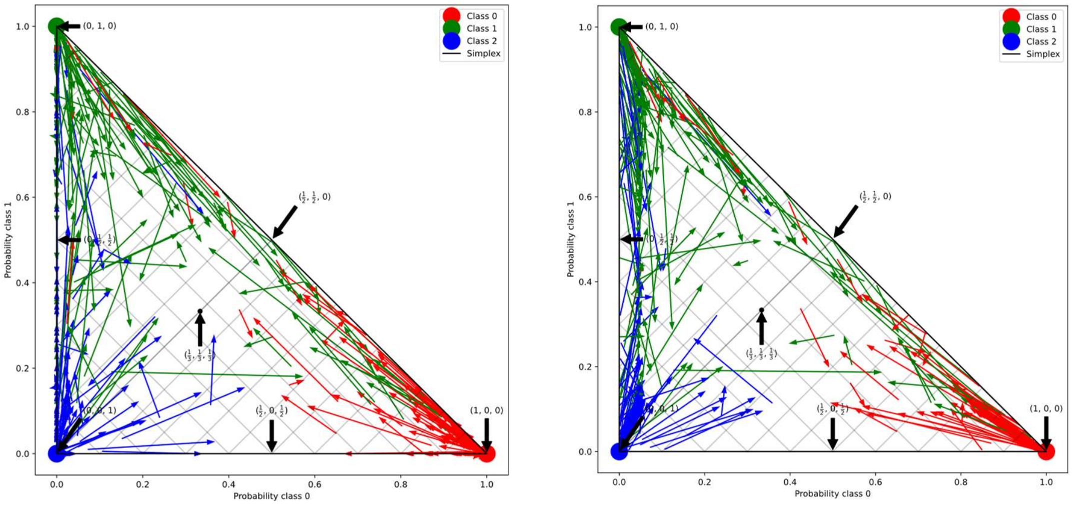
Table 10. The maximum accuracy achieved from the classification models.

Models	Maximum Accuracy (%)	Number of features for maximum accuracy
Random Forest	66.60	44
MLP	57.73	58
XG Boost	**69.47**	**40**
Logistic Regression	55.44	50
SVM	64.05	49
KNN	51.28	3
Decision Trees	53.23	5

Table 11. Results after XG Boost classifier calibration with Isotonic Regression and Platt's methods.

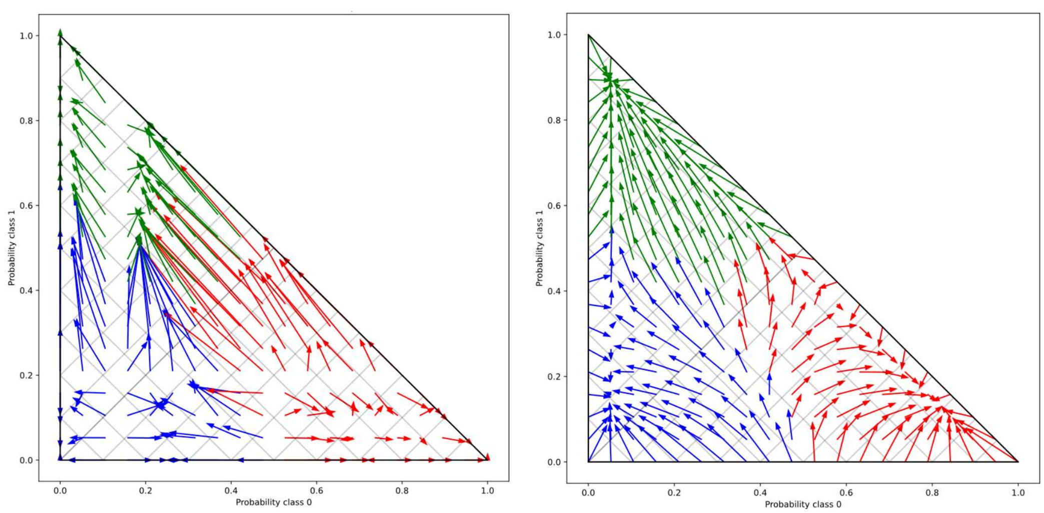
Models	Log-loss	Accuracy (%)
XG Boost	1.195	69.47
XG Boost + Isotonic	0.513	72.03
XG Boost + Platt	**0.489**	**76.52**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1



(a) (b)

Figure 7. Change of predicted probabilities on test samples after calibration with: (a) Isotonic Regression method; (b) Platt's (sigmoid) method.



(a) (b)

Figure 8. Learned calibration map with: (a) Isotonic Regression method; (b) Platt's (sigmoid) method.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

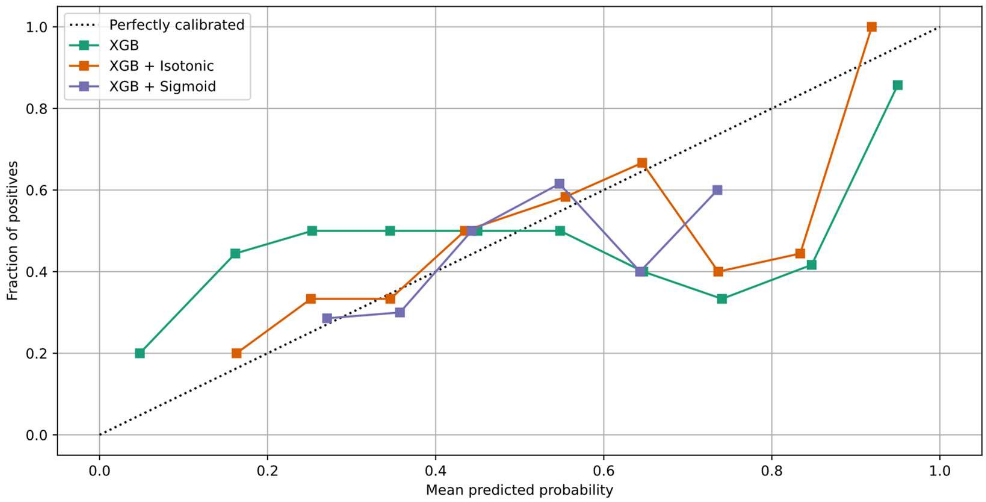


Figure 9. Calibration plot of XG Boost classifier for class 0.

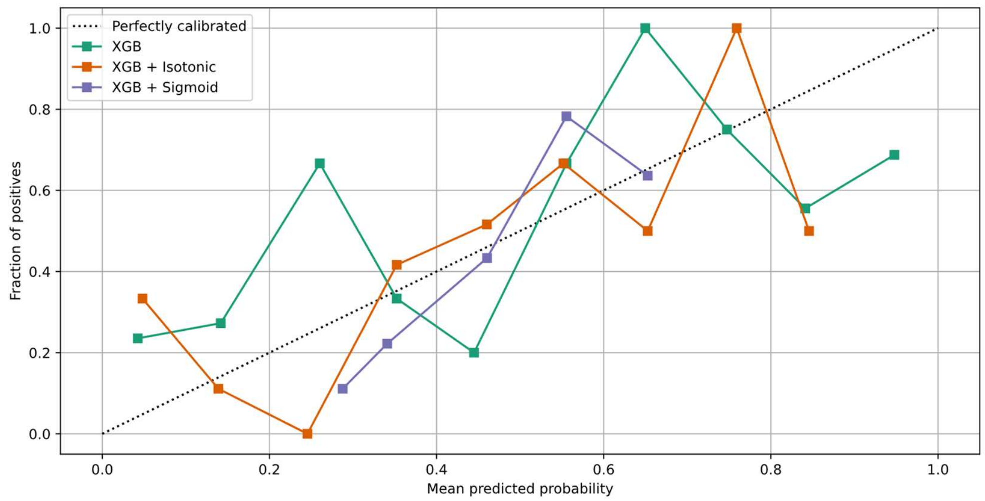


Figure 10. Calibration plot of XG Boost classifier for class 1.

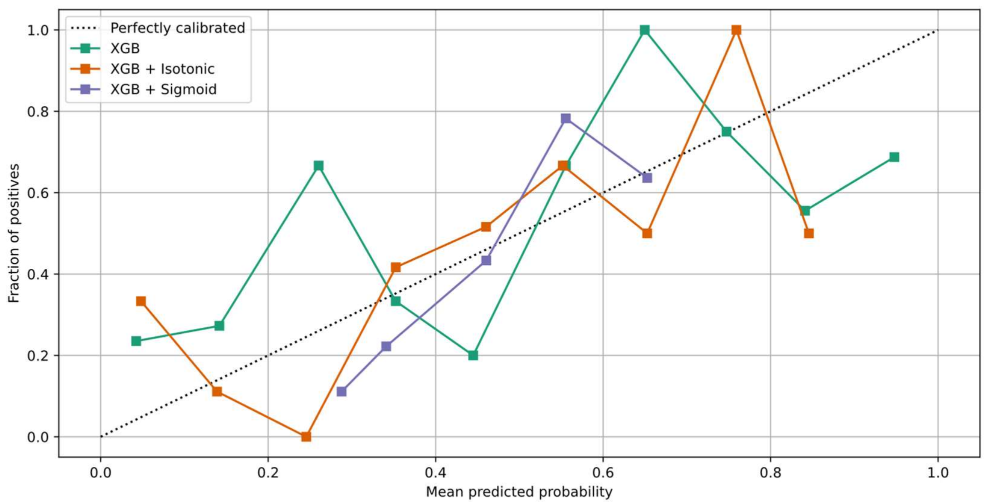


Figure 11. Calibration plot of XG Boost classifier for class 2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

3.2.4. Post-hoc explainability

In Figure 12 the x-axis represents the average magnitude change in model output when a feature is excluded from the model. The higher the value, the higher the importance of this feature is in the prediction outcome of the model. In Figure 13, Figure 14 and Figure 15, the feature names are presented in y-axis based on their importance from top to bottom, while the x-axis indicates the mean SHAP value showing the change in log-odds. Gradient color (red to blue) indicates the original value of that feature. Each point represents a patient from the original dataset. Figure 16, Figure 17 and Figure 18 show the mean SHAP values of each feature that affects the classification of a patient between two groups.

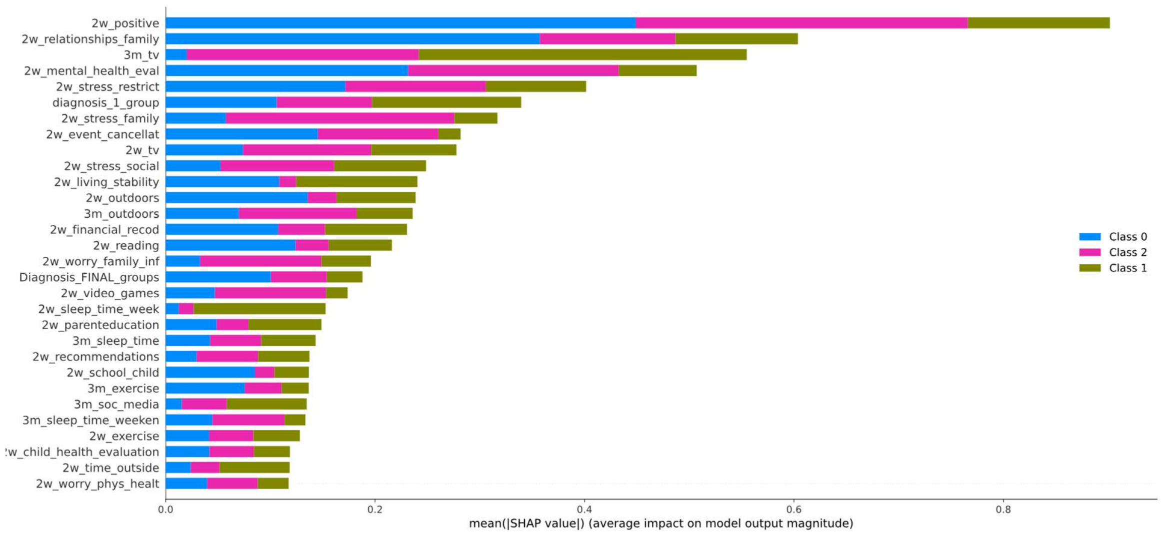


Figure 12. Mean SHAP values.

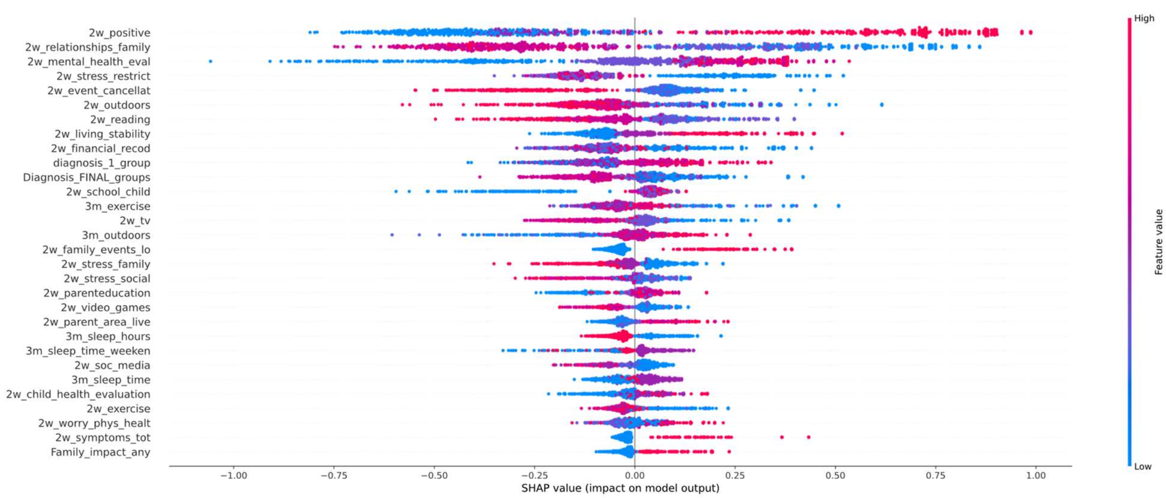


Figure 13. SHAP values of patients from class 0.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

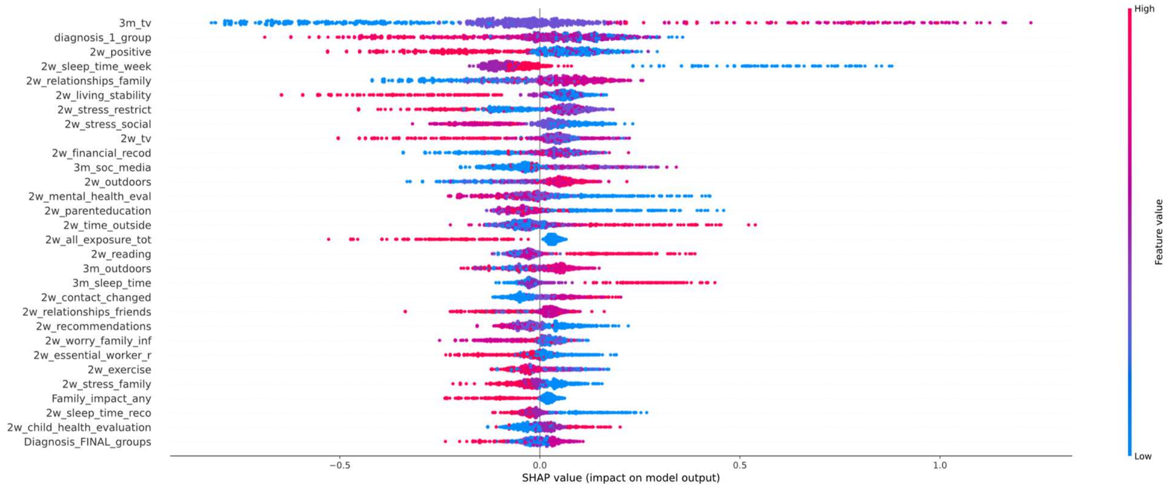


Figure 14. SHAP values of patients from class 1.

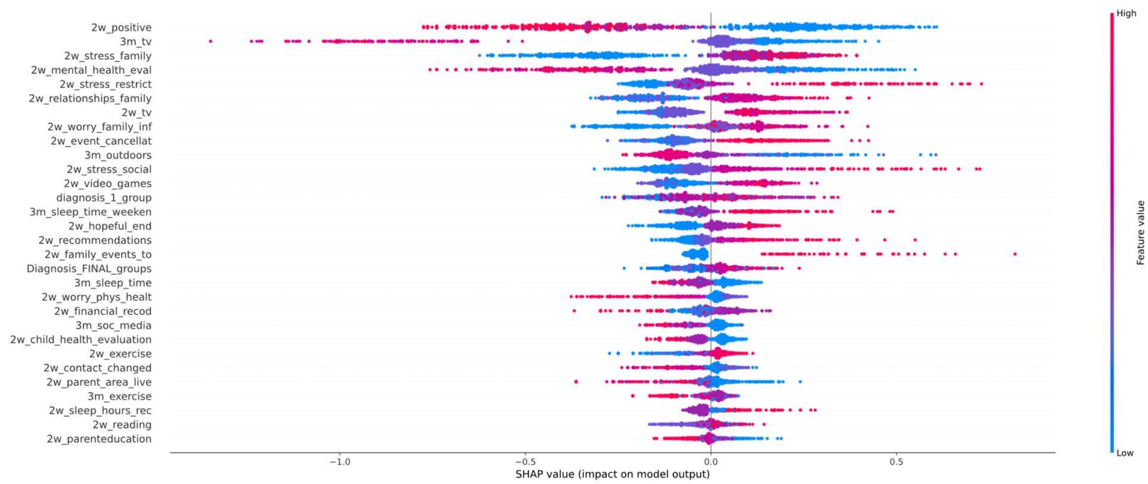


Figure 15. SHAP values of patients from class 2.

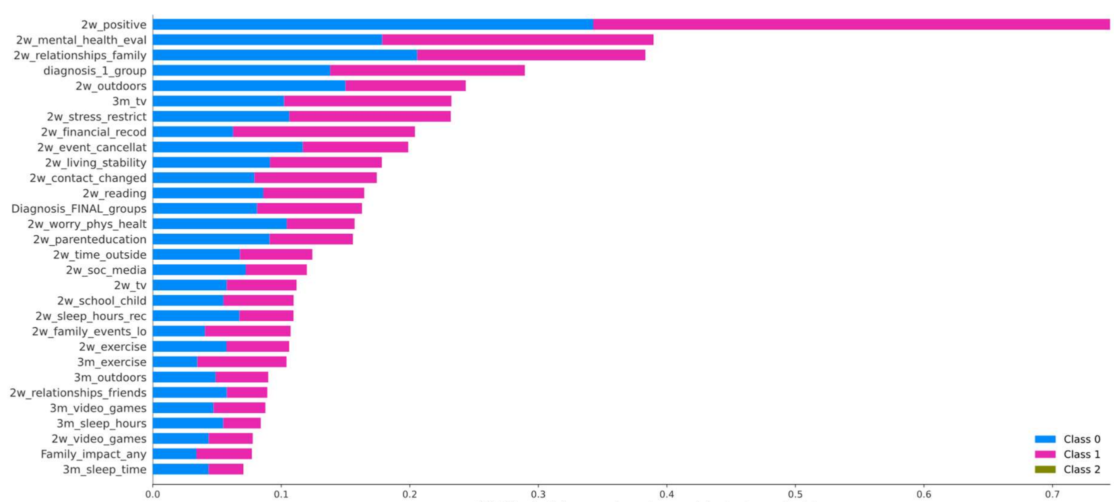


Figure 16. Mean SHAP values of patients from class 0 and class 1.

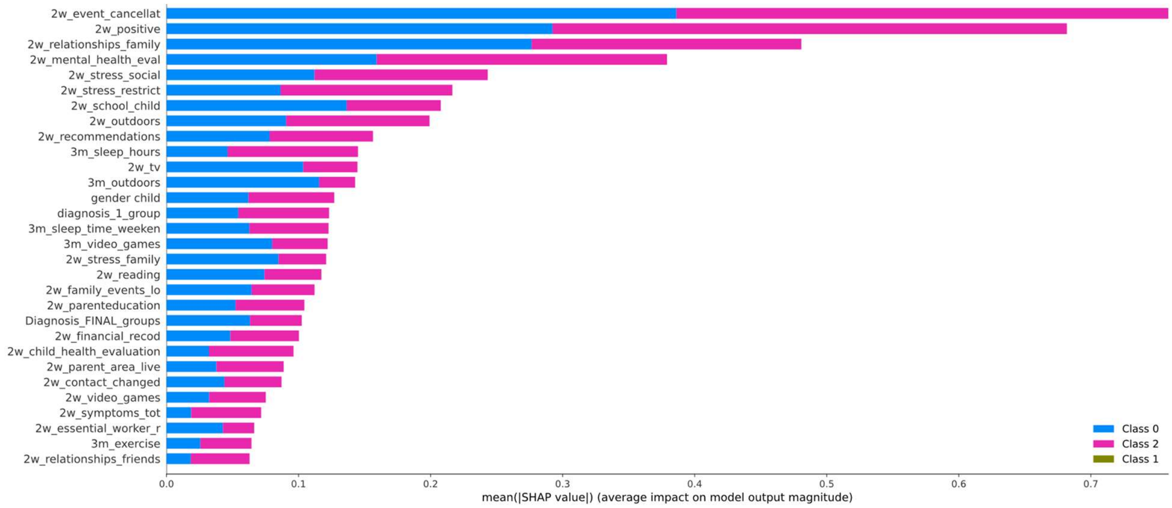


Figure 17. Mean SHAP values of patients from class 0 and class 2.

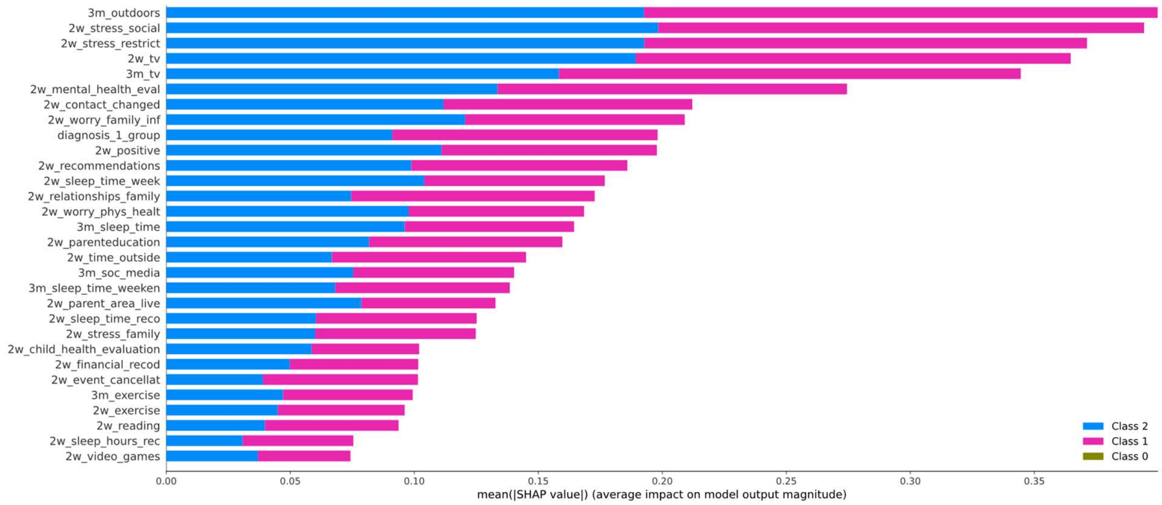


Figure 18. SHAP values patients from class 1 and class 2.

4. Discussion

4.1 Clustering

The clustering results indicated that the Jenks method is the most suitable to be adopted in our study, reaching the highest evaluation score (

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

Table 8). The clusters derived from Jenks method indicate that most of the individuals that participated in this study (469 out of 744, 63.04%) did not have any significant alteration to their mood state (Table 7). Also, it is important to mention that the 1st lockdown in Greece had negative impact on more individuals (169, 22.71%) than positive (106, 14.25%).

4.2 Feature selection

The results revealed that social life aspects play significant role in the prediction output (Table 9, Figure 5). Furthermore, daily activities is the second most important category with 6 features in the feature selection subset. Last behavioral effects and demographics contribute with 5 features each. The rest features belong to the categories of medical diagnosis / rehabilitation, sleeping habits, health concerns, family life and personal life (Figure 5).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

4.3 Classification and calibration

The results on Table 10 showed that the XG Boost model presented a more stable performance compared to the other models achieving the maximum accuracy (69.47%) at 40 features. A comparable performance (66.60%) was also achieved by Random Forest at 44 features.

The calibration results showed that the calibrated XG Boost with Isotonic Regression achieved lower log-loss but also slightly lower accuracy compared to the calibrated XG Boost with Platt's method (Table 11). In Figure 7a,b the vertexes of the simplex represent the perfectly predicted classes (e.g., 0, 0, 1). The middle point $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$ inside the simplex represents the prediction of the 3 classes with equal probability $(\frac{1}{3},\frac{1}{3},\frac{1}{3})$. The start of an arrow is at the uncalibrated probabilities while the head of an arrow shows the calibrated probability. For a lower over-confident model, the arrows point away from the edges where the probabilities of a class is zero. This can be better observed to the calibrated XG Boost with Platt's method which produces more accurate predicted probabilities incurring a lower log-loss.

The learned calibration maps showed that the Platt's method succeeded on calibrating the model better compared to Isotonic Regression method. Indeed, this can also be observed on Figure 9, Figure 10 and Figure 11 where the calibration plots for each class over the others are illustrated. In all cases the XG Boost model calibrated with Platt's (sigmoid) method verges more to the perfectly calibrated line compared to non-calibrated model or XG Boost model calibrated with Isotonic Regression method.

4.4 Post-hoc explainability

The results showed that the mood state of the population was highly linked with their positive or negative changes in the child's life due to the COVID-19 crisis, the relationships among the family and the evaluation of the mental health before the COVID-19 crisis (Figure 12). Also, important contribution was proved to have the increase of the child's time spent on watching TV or digital means among 3 months before and 2 weeks after the lockdown. Therefore, we can observe that there was a negative impact on children who did not use to spend much time on TV but their time was increased due to lockdown. It is important to mention that the first diagnosis defined by the medical expert was played a significant role in the change of children's mood state.

Regarding local exploration, Figure 15 shows that the most important features that contribute to classifying an individual to the group with negative change of mood state include the lack of positive changes to their life, the increase on watching tv, the stress derived from the restrictions, and the stress caused to the child by changes in family contacts. When it comes to the individuals that have not been affected by the 1st lockdown imposed in Greece, the following features found to contribute most to this category: 3m_tv, diagnosis_1_group, 2w_positive and 2w_sleep_time_week. Based on Figure 14, responses that indicate a neutral attitude towards these features, led to the classification of an individual as a child without mood state alteration. For instance, a child's time spent on watching TV was not affected significantly during the lockdown or their sleeping schedule was acceptable for a child (sleeping time at 20:00-22:00 o'clock) could lead to a more stable mood state. On the other hand, from the beeswarm in Figure 13, it is shown that more positive changes to their lives due to COVID-19 and better relationships with the family members can lead to a more positive behavior during the lockdown. Family cohesion and continuity in functional routines are protective factors that enhance mental resilience, involving a balance between adversity and availability of support. Protective factors act as a buffer against stress and moderate its impact on emotional well-being, as they enable children to cope with significant life events. Resilient family function provides children a sense of connectedness, healthy family attachments and stability. Supportive parenting and family warmth facilitate stress exposure and thus result in positive emotional development [54].

When it comes to the pair-wise comparison among the groups, the Figure 16 indicates that the main features that contributed to the distinction among the individuals that got better during the 1st lockdown and those whose mood state was not significantly affected. These features are 2w_positive, 2w_mental_health_eval, and 2w_relationships_family. The most contributed features among the groups of children that had positive (class 0) or negative (class 2) change to their mood state are 2w_event_canellat, 2w_positive, 2w_relationships_family, and 2w_mental_health_eval (Figure 17). Last, the main features that contributed to the classification output among class 1 and class 2 are the 2w_event_canellat, 2w_positive, 2w_relationships_family, and 2w_mental_health_eval (Figure 18).

Overall, we can conclude that if the 1st lockdown did not lead to positive changes or impacted negatively the daily activities and family relationships of the child then a deterioration on the mood state of a child was noticed. On the other hand, if the COVID-19 restrictions did not affect the daily life and habits of the child, *i.e.* time spent on TV, sleeping schedule, then no significant change on the mood state was noticed. Indeed, the stability on the functional routines constitutes a critical factor for the management of stressful events, such as a pandemic [55]. Last, if during the 1st lockdown, children managed to change their life habits in a positive way, improved their relationships with family members and did not get affected by the cancellation of social events, then the change on their mood state was positive. Based on these conclusions, we can generalize that more outgoing and active children that did not use to spend more time at home watching TV prior to the pandemic were the most affected by the lockdown. On the other hand, children, whose habits and daily life schedule did not alter significantly, were the least affected by the COVID-19 restrictions.

Apart from the features that have been included in the analysis, another perspective that should be considered and could probably explain the significant larger size of the class 1 compared to the others (class 0 and 2) is the resilience in children and youth. Based on [56], resilience is defined as the capacity of a dynamic system to adapt successfully to challenges that threaten the function, survival, or development of the system. Various studies in the literature have highlighted the ability of children to adapt and benefit from their strengths and protective factors to succeed despite biological and environmental influences, such as poverty, illness, violence, disasters, and family dissonance, among others [57–59] while few of them have focused on the case of COVID-19 [60]. Protective factors mainly include individual characteristics, environmental supports and family conditions. Indeed, in Figure 12, 6 factors are directly related to family conditions, such as relationships with family members, parental education and financial stress, 9 factors are indirectly related to family and parental control, such as sleeping schedule and time dedicated to social media and TV. Also, 8 factors are related to the ability of the child or youth to adapt to COVID-19 changes, such as changes to school attendance and social contacts, etc., while the rest factors are linked with environmental supports, such as outdoor activities.

5. Conclusions

In this study an explainable machine learning pipeline was proposed to identify and interpret the most important features that contributed to the changes on the mood state of children and youths during the 1st lockdown in Greece. To identify the change on the mood state of the individuals under examination, the problem was formulated as a three-class classification problem. The classes include the individuals with positive (class 0) and negative (class 2) change on their mood state and the individuals without significant change on their mood state (class 1). A thorough comparative evaluation was conducted to identify the best performed clustering method and prediction model for this problem. Jenks method was selected as the clustering method, following by a feature selection performed by ReliefF. Then, the best performed prediction model, XG Boost, was used for calibration and a post-hoc explainability analysis to justify the main features that

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

contribute to the prediction output of the model. Also, insights were given about the influence of each feature among the classes.

Overall, we can conclude that the positive changes to child's life due to the 1st lockdown, the relationships among the family members, the time spent on TV and parental evaluation of the child's mental health and the stress caused by COVID-19 restrictions could play crucial role to the change of the mood state of the child. These results are aligned with the results of relevant studies found on the literature that incorporated pre-pandemic clinical samples or population-based cohorts of children at high risk for transition from subclinical to clinically significant levels of psychopathology [61–63]. Moreover, the finding that that most of the children and youths managed to maintain stable mood (63.04%: 469 out of 744) or even have positive mood change (14.25%: 106 out of 744) may be related to the concept of resilience. This is aligned to the psychological approach and perspectives on resilience in children and youth [57–59] and specifically on COVID-19 [60]. Specifically, these children seem to maintain their capacity for resilience, even under these difficult restrictive conditions. People may experience conditions of loss or high anxiety, but they may have little effect on their mental health or even experience positive aspects [64]. In a recent meta-analysis conducted by Prati & Mancini (2021), which includes also studies of children and adolescents, the psychological impact of COVID-19 lockdowns was small in magnitude, highlighting that most people are psychologically resilient to their effects [65]. There can be a positive adjustment of children after an acute life event and the factors that contribute to it are both intra-individual and contextual factors (e.g. supportive relations) [66], as well as relationships with parents or the school's ability to respond to the emergency [67]. Also, it seems that stability in functional routines is a key factor in managing stressful events. In accordance with this are the results of Giuntella et al. (2020) who found that disruptions in physical activity, sleep and screen time among young adults at the onset of the pandemic are more closely linked to depression during the pandemic [68].

Main limitations of this work that should be taken into account are the unexpected end of therapies by some children and the fact that parents answered the questionnaires on the behalf of their children considering different time periods. Also, the large diversity of clinical diagnoses in combination with the small number of children falling into separate specifically defined diagnostic codes, imposed the necessity to use broader diagnostic categories and therefore to not succeed on observing the relation among the impact of COVID-19 related restrictions to children and the diagnostic criteria from a specific disorder (e.g. ADHD). Future work includes a within subject analysis of the data from the longitudinal study of the 1st and 2nd lockdowns. It remains to be seen whether the second prolonged lockdown (six months) had a greater impact on the clustering of the population.

Author Contributions: Conceptualization, C.N.; methodology, C.N.; software, C.N.; validation, C.N.; formal analysis, C.N., I.R., I.G., As.S. and E.L.; data curation, C.N., D.P., K.M., I.G., K.K., As.S., E.T., A.G., K.L., I.K., N.S., G.O'C.; writing—original draft preparation, C.N., D.P., I.R., As.S., E.L. and M.C.-T.; writing—review and editing, C.N., I.R., I.G. and M.C.-T.; visualization, C.N.; supervision, Ar.S. and E.L.; project administration, Ar.S. and E.L.; All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Each CAMHS contacted the parents of all children and adolescents who attended the service from March 1st, 2019 to March 1st, 2020. All parents interested in taking part in the survey were sent an email containing information about the study, along with a unique identification code number and the link to log into Google Forms Survey app. After reading the information about the goals of the study, the process of data collection and confidentiality, and providing informed consent online they proceeded answering the questionnaire. The study was approved by the Ethics Committee of each hospital, with which the service is affiliated. The study was performed in line with the principles of the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Acknowledgments: The authors would like to thank all the respondents to this study who took the time to complete the questionnaire. We would also like to thank the Hellenic cOvid-19 imPact survEy (HOPE) Consortium for their contribution during the data collection process:

Lagakou E., First Psychiatric Department, Eginition Hospital, National and Kapodistrian University of Athens, Athens, Greece; elagakou@gmail.com

Mamaki, E., Mental Health Center, General Hospital "G. Hatzikosta", Ioannina, Greece; g.vottis@yahoo.gr

Neou, E., Hellenic Centre for Mental Health and Research, Athens, Greece; evaneou@gmail.com

Polaki, O., Community Mental Health Center for Children and Adolescents in N.Smyrni, Division of Psychiatry, "Asklepieion Voulas' General Hospital, Attica, Greece; olympiapolaki@yahoo.gr

Priftis D., University Mental Health Research Institute; icedale@gmail.com

Triantafyllou, G., Second Psychiatric Department, "Attikon" University Hospital, National and Kapodistrian University of Athens, Athens, Greece; g_triantafillou@yahoo.gr

Valvi E., Athens Child and Adolescent Mental Health Centre, General Children's Hospital "Pan. & Aglaia Kyriakou", Athens, Greece

Vassara, V., Community Mental Health Center for Children and Adolescents, Department of Psychiatry, University Hospital of Ioannina, Ioannina, Greece; vasilikivassara@gmail.com

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chakraborty, I.; Maity, P. COVID-19 Outbreak: Migration, Effects on Society, Global Environment and Prevention. *Science of The Total Environment* **2020**, *728*, 138882, doi:10.1016/j.scitotenv.2020.138882.
2. World Health Organization Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). **2020**.
3. Bonardi, J.-P.; Gallea, Q.; Kalanoski, D.; Lalive, R. Fast and Local: How Did Lockdown Policies Affect the Spread and Severity of COVID-19? 148.
4. Ma, Z.; Idris, S.; Zhang, Y.; Zewen, L.; Wali, A.; Ji, Y.; Pan, Q.; Baloch, Z. The Impact of COVID-19 Pandemic Outbreak on Education and Mental Health of Chinese Children Aged 7–15 Years: An Online Survey. *BMC Pediatrics* **2021**, *21*, doi:10.1186/s12887-021-02550-1.
5. Abas, M.A.; Weobong, B.; Burgess, R.A.; Kienzler, H.; Jack, H.E.; Kidia, K.; Musesengwa, R.; Petersen, I.; Collins, P.Y.; Nakimuli-Mpungu, E. COVID-19 and Global Mental Health. *Lancet Psychiatry* **2021**, *8*, 458–459, doi:10.1016/S2215-0366(21)00155-3.
6. Vindegaard, N.; Benros, M.E. COVID-19 Pandemic and Mental Health Consequences: Systematic Review of the Current Evidence. *Brain, Behavior, and Immunity* **2020**, *89*, 531–542, doi:10.1016/j.bbi.2020.05.048.
7. Xiong, J.; Lipsitz, O.; Nasri, F.; Lui, L.M.W.; Gill, H.; Phan, L.; Chen-Li, D.; Iacobucci, M.; Ho, R.; Majeed, A.; et al. Impact of COVID-19 Pandemic on Mental Health in the General Population: A Systematic Review. *Journal of Affective Disorders* **2020**, *277*, 55–64, doi:10.1016/j.jad.2020.08.001.
8. Vizheh, M.; Qorbani, M.; Arzaghi, S.M.; Muhidin, S.; Javanmard, Z.; Esmaeili, M. The Mental Health of Healthcare Workers in the COVID-19 Pandemic: A Systematic Review. *J Diabetes Metab Disord* **2020**, *19*, 1967–1978, doi:10.1007/s40200-020-00643-9.
9. Rens, E.; Smith, P.; Nicaise, P.; Lorant, V.; Van den Broeck, K. Mental Distress and Its Contributing Factors Among Young People During the First Wave of COVID-19: A Belgian Survey Study. *Frontiers in Psychiatry* **2021**, *12*, doi:10.3389/fpsyt.2021.575553.
10. Zhou, S.-J.; Zhang, L.-G.; Wang, L.-L.; Guo, Z.-C.; Wang, J.-Q.; Chen, J.-C.; Liu, M.; Chen, X.; Chen, J.-X. Prevalence and Socio-Demographic Correlates of Psychological Health Problems in Chinese Adolescents during the Outbreak of COVID-19. *European Child and Adolescent Psychiatry* **2020**, *29*, 749–758, doi:10.1007/s00787-020-01541-4.

11. Tamarit, A.; de la Barrera, U.; Mónaco, E.; Schoeps, K.; Montoya-Castilla, I. Psychological Impact of COVID-19 Pandemic in Spanish Adolescents: Risk and Protective Factors of Emotional Symptoms. *Revista de Psicologia Clinica con Ninos y Adolescentes* **2020**, *7*, 73–80, doi:10.21134/rpcna.2020.mon.2037.
12. Chen, X.; Qi, H.; Liu, R.; Feng, Y.; Li, W.; Xiang, M.; Cheung, T.; Jackson, T.; Wang, G.; Xiang, Y.-T. Depression, Anxiety and Associated Factors among Chinese Adolescents during the COVID-19 Outbreak: A Comparison of Two Cross-Sectional Studies. *Translational Psychiatry* **2021**, *11*, doi:10.1038/s41398-021-01271-4.
13. Tavolacci, M.P.; Wouters, E.; Van de Velde, S.; Buffel, V.; Déchelotte, P.; Van Hal, G.; Ladner, J. The Impact of Covid-19 Lockdown on Health Behaviors among Students of a French University. *International Journal of Environmental Research and Public Health* **2021**, *18*, doi:10.3390/ijerph18084346.
14. Fu, W.; Yan, S.; Zong, Q.; Anderson-Luxford, D.; Song, X.; Lv, Z.; Lv, C. Mental Health of College Students during the COVID-19 Epidemic in China. *Journal of Affective Disorders* **2021**, *280*, 7–10, doi:10.1016/j.jad.2020.11.032.
15. Xiao, H.; Shu, W.; Li, M.; Li, Z.; Tao, F.; Wu, X.; Yu, Y.; Meng, H.; Vermund, S.H.; Hu, Y. Social Distancing among Medical Students during the 2019 Coronavirus Disease Pandemic in China: Disease Awareness, Anxiety Disorder, Depression, and Behavioral Activities. *International Journal of Environmental Research and Public Health* **2020**, *17*, 1–13, doi:10.3390/ijerph17145047.
16. Wathelet, M.; Duhem, S.; Vaiva, G.; Baubet, T.; Habran, E.; Veerapa, E.; Debien, C.; Molenda, S.; Horn, M.; Grandgenèvre, P.; et al. Factors Associated With Mental Health Disorders Among University Students in France Confined During the COVID-19 Pandemic. *JAMA Network Open* **2020**, *3*, e2025591–e2025591, doi:10.1001/jamanetworkopen.2020.25591.
17. Ren, Z.; Xin, Y.; Ge, J.; Zhao, Z.; Liu, D.; Ho, R.C.M.; Ho, C.S.H. Psychological Impact of COVID-19 on College Students After School Reopening: A Cross-Sectional Study Based on Machine Learning. *Frontiers in Psychology* **2021**, *12*, doi:10.3389/fpsyg.2021.641806.
18. Cost, K.T.; Crosbie, J.; Anagnostou, E.; Birken, C.S.; Charach, A.; Monga, S.; Kelley, E.; Nicolson, R.; Maguire, J.L.; Burton, C.L.; et al. Mostly Worse, Occasionally Better: Impact of COVID-19 Pandemic on the Mental Health of Canadian Children and Adolescents. *European Child and Adolescent Psychiatry* **2021**, doi:10.1007/s00787-021-01744-3.
19. Zhou, S.-J.; Wang, L.-L.; Yang, R.; Yang, X.-J.; Zhang, L.-G.; Guo, Z.-C.; Chen, J.-C.; Wang, J.-Q.; Chen, J.-X. Sleep Problems among Chinese Adolescents and Young Adults during the Coronavirus-2019 Pandemic. *Sleep Medicine* **2020**, *74*, 39–47, doi:10.1016/j.sleep.2020.06.001.
20. Yeasmin, S.; Banik, R.; Hossain, S.; Hossain, M.N.; Mahumud, R.; Salma, N.; Hossain, M.M. Impact of COVID-19 Pandemic on the Mental Health of Children in Bangladesh: A Cross-Sectional Study. *Children and Youth Services Review* **2020**, *117*, doi:10.1016/j.childyouth.2020.105277.
21. Garcia de Avila, M.A.; Hamamoto Filho, P.T.; Jacob, F.L.D.S.; Alcantara, L.R.S.; Berghammer, M.; Jenholt Nolbris, M.; Olaya-Contreras, P.; Nilsson, S. 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. *International journal of environmental research and public health* **2020**, *17*, doi:10.3390/ijerph17165757.
22. Liang, L.; Ren, H.; Cao, R.; Hu, Y.; Qin, Z.; Li, C.; Mei, S. The Effect of COVID-19 on Youth Mental Health. *Psychiatric Quarterly* **2020**, *91*, 841–852, doi:10.1007/s11126-020-09744-3.
23. Ma, Z.; Zhao, J.; Li, Y.; Chen, D.; Wang, T.; Zhang, Z.; Chen, Z.; Yu, Q.; Jiang, J.; Fan, F.; et al. Mental Health Problems and Correlates among 746 217 College Students during the Coronavirus Disease 2019 Outbreak in China. *Epidemiology and Psychiatric Sciences* **2020**, doi:10.1017/S2045796020000931.
24. Sciberras, E.; Patel, P.; Stokes, M.A.; Coghill, D.; Middeldorp, C.M.; Bellgrove, M.A.; Becker, S.P.; Efron, D.; Stringaris, A.; Faraone, S.V.; et al. Physical Health, Media Use, and Mental Health in Children and Adolescents With ADHD During the COVID-19 Pandemic in Australia. *J Atten Disord* **2020**, 1087054720978549, doi:10.1177/1087054720978549.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 22 December 2021 doi:10.20944/preprints202112.0374.v1

25. Ge, F.; Zhang, D.; Wu, L.; Mu, H. Predicting Psychological State Among Chinese Undergraduate Students in the COVID-19 Epidemic: A Longitudinal Study Using a Machine Learning. *Neuropsychiatr Dis Treat* **2020**, *16*, 2111–2118, doi:10.2147/NDT.S262004.
26. Prout, T.A.; Zilcha-Mano, S.; Aafjes-van Doorn, K.; Békés, V.; Christman-Cohen, I.; Whistler, K.; Kui, T.; Di Giuseppe, M. Identifying Predictors of Psychological Distress During COVID-19: A Machine Learning Approach. *Front Psychol* **2020**, *11*, doi:10.3389/fpsyg.2020.586202.
27. Cabitza, F.; Campagner, A. The Need to Separate the Wheat from the Chaff in Medical Informatics: Introducing a Comprehensive Checklist for the (Self)-Assessment of Medical AI Studies. *Int J Med Inform* **2021**, *153*, 104510, doi:10.1016/j.ijmedinf.2021.104510.
28. Peng, K.; Leung, V.C.M.; Huang, Q. Clustering Approach Based on Mini Batch Kmeans for Intrusion Detection System Over Big Data. *IEEE Access* **2018**, *6*, 11897–11906, doi:10.1109/ACCESS.2018.2810267.
29. von Luxburg, U. A Tutorial on Spectral Clustering. *Stat Comput* **2007**, *17*, 395–416, doi:10.1007/s11222-007-9033-z.
30. Ward, J.H. Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association* **1963**, *58*, 236–244, doi:10.1080/01621459.1963.10500845.
31. Murtagh, F.; Legendre, P. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? *J Classif* **2014**, *31*, 274–295, doi:10.1007/s00357-014-9161-z.
32. Sokal, R.R.; Michener, C.D. A Statistical Method of Evaluating Systematic Relationships. *The University of Kansas Science Bulletin* **1958**, *38*, 1409–1438.
33. Yim, O.; Ramdeen, K.T. Hierarchical Cluster Analysis: Comparison of Three Linkage Measures and Application to Psychological Data. *TQMP* **2015**, *11*, 8–21, doi:10.20982/tqmp.11.1.p008.
34. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: An Efficient Data Clustering Method for Very Large Databases. *SIGMOD Rec.* **1996**, *25*, 103–114, doi:10.1145/235968.233324.
35. Zhang, T.; Ramakrishnan, R.; Livny, M. BIRCH: A New Data Clustering Algorithm and Its Applications. **1997**, 42.
36. Anchang, J.Y.; Ananga, E.O.; Pu, R. An Efficient Unsupervised Index Based Approach for Mapping Urban Vegetation from IKONOS Imagery. *International Journal of Applied Earth Observation and Geoinformation* **2016**, *50*, 211–220, doi:10.1016/j.jag.2016.04.001.
37. North, M.A. A Method for Implementing a Statistically Significant Number of Data Classes in the Jenks Algorithm. In Proceedings of the 2009 Sixth International Conference on Fuzzy Systems and Knowledge Discovery; August 2009; Vol. 1, pp. 35–38.
38. Zhang, L.; Zhang, X.; Yuan, S.; Wang, K. Economic, Social, and Ecological Impact Evaluation of Traffic Network in Beijing–Tianjin–Hebei Urban Agglomeration Based on the Entropy Weight TOPSIS Method. *Sustainability* **2021**, *13*, 1862, doi:10.3390/su13041862.
39. Robnik-Šikonja, M.; Kononenko, I. Theoretical and Empirical Analysis of ReliefF and RReliefF. *Machine Learning* **2003**, *53*, 23–69, doi:10.1023/A:1025667309714.
40. Spolaôr, N.; Cherman, E.A.; Monard, M.C.; Lee, H.D. ReliefF for Multi-Label Feature Selection. In Proceedings of the 2013 Brazilian Conference on Intelligent Systems; October 2013; pp. 6–11.
41. Alelyani, S. Stable Bagging Feature Selection on Medical Data. *Journal of Big Data* **2021**, *8*, 11, doi:10.1186/s40537-020-00385-8.
42. Huang, Y.; McCullagh, P.J.; Black, N.D. An Optimization of ReliefF for Classification in Large Datasets. *Data & Knowledge Engineering* **2009**, *68*, 1348–1356, doi:10.1016/j.datak.2009.07.011.
43. Kilicarslan, S.; Adem, K.; Celik, M. Diagnosis and Classification of Cancer Using Hybrid Model Based on ReliefF and Convolutional Neural Network. *Medical Hypotheses* **2020**, *137*, 109577, doi:10.1016/j.mehy.2020.109577.
44. Kononenko, I. Estimating Attributes: Analysis and Extensions of RELIEF. In *Lecture Notes in Computer Science*; Springer, 1994; Vol. 784, pp. 171–182.

45. Kononenko, I.; Robnik-Sikonja, M.; Robnik, M.; Pompe, U. ReliefF for Estimation and Discretization of Attributes in Classification, Regression, and ILP Problems 1996.
46. Ntakolia, C.; Kokkotis, C.; Moustakidis, S.; Tsaopoulos, D. Prediction of Joint Space Narrowing Progression in Knee Osteoarthritis Patients. *Diagnostics* **2021**, *11*, 285, doi:10.3390/diagnostics11020285.
47. Ntakolia, C.; Kokkotis, C.; Moustakidis, S.; Tsaopoulos, D. A Machine Learning Pipeline for Predicting Joint Space Narrowing in Knee Osteoarthritis Patients. In Proceedings of the 2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE); October 2020; pp. 934–941.
48. Liu, M.; Xu, X.; Tao, Y.; Wang, X. An Improved Random Forest Method Based on RELIEFF for Medical Diagnosis. In Proceedings of the 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC); July 2017; Vol. 1, pp. 44–49.
49. Jamshidi, A.; Pelletier, J.-P.; Martel-Pelletier, J. Machine-Learning-Based Patient-Specific Prediction Models for Knee Osteoarthritis. *Nat Rev Rheumatol* **2019**, *15*, 49–60, doi:10.1038/s41584-018-0130-5.
50. Harimoorthy, K.; Thangavelu, M. Multi-Disease Prediction Model Using Improved SVM-Radial Bias Technique in Healthcare Monitoring System. *J Ambient Intell Human Comput* **2021**, *12*, 3715–3723, doi:10.1007/s12652-019-01652-0.
51. Lundberg, S.M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In Proceedings of the Advances in Neural Information Processing Systems; Curran Associates, Inc., 2017; Vol. 30.
52. SMOTE | Overcoming Class Imbalance Problem Using SMOTE Available online: https://www.analyticsvidhya.com/blog/2020/10/overcoming-class-imbalance-using-smote-techniques/ (accessed on 16 August 2021).
53. Douzas, G.; Bacao, F.; Last, F. Improving Imbalanced Learning through a Heuristic Oversampling Method Based on K-Means and SMOTE. *Information Sciences* **2018**, *465*, 1–20, doi:10.1016/j.ins.2018.06.056.
54. Hornor, G. Resilience. *J Pediatr Health Care* **2017**, *31*, 384–390, doi:10.1016/j.pedhc.2016.09.005.
55. Koome, F.; Hocking, C.; Sutton, D. Why Routines Matter: The Nature and Meaning of Family Routines in the Context of Adolescent Mental Illness. *Journal of Occupational Science* **2012**, *19*, 312–325, doi:10.1080/14427591.2012.718245.
56. Masten, A.S. Resilience Theory and Research on Children and Families: Past, Present, and Promise. *Journal of Family Theory & Review* **2018**, *10*, 12–31, doi:10.1111/jftr.12255.
57. Zolkoski, S.M.; Bullock, L.M. Resilience in Children and Youth: A Review. *Children and Youth Services Review* **2012**, *34*, 2295–2303, doi:10.1016/j.childyouth.2012.08.009.
58. Masten, A.S. Global Perspectives on Resilience in Children and Youth. *Child Development* **2014**, *85*, 6–20, doi:10.1111/cdev.12205.
59. Masten, A.S.; Barnes, A.J. Resilience in Children: Developmental Perspectives. *Children* **2018**, *5*, 98, doi:10.3390/children5070098.
60. Masten, A.S.; Motti-Stefanidi, F. Multisystem Resilience for Children and Youth in Disaster: Reflections in the Context of COVID-19. *ADV RES SCI* **2020**, *1*, 95–106, doi:10.1007/s42844-020-00010-w.
61. Bouter, D.; Zarchev, M.; de Neve-Enthoven, N.; Ravensbergen, S.; Kamperman, A.M.; Hoogendijk, W.; Grootendorst, N. A Longitudinal Study of Mental Health in Adolescents before and during the COVID-19 Pandemi 2021.
62. Lopez-Serrano, J.; Díaz-Bóveda, R.; González-Vallespí, L.; Santamarina-Pérez, P.; Bretones-Rodríguez, A.; Calvo, R.; Lera-Miguel, S. Psychological Impact during COVID-19 Lockdown in Children and Adolescents with Previous Mental Health Disorders. *Revista de Psiquiatría y Salud Mental* **2021**, doi:10.1016/j.rpsm.2021.04.002.
63. Penner, F.; Hernandez Ortiz, J.; Sharp, C. Change in Youth Mental Health During the COVID-19 Pandemic in a Majority Hispanic/Latinx US Sample. *J Am Acad Child Adolesc Psychiatry* **2021**, *60*, 513–523, doi:10.1016/j.jaac.2020.12.027.
64. Bonanno, G.A. Loss, Trauma, and Human Resilience: Have We Underestimated the Human Capacity to Thrive after Extremely Aversive Events? *Am Psychol* **2004**, *59*, 20–28, doi:10.1037/0003-066X.59.1.20.
65. Prati, G.; Mancini, A.D. The Psychological Impact of COVID-19 Pandemic Lockdowns: A Review and Meta-Analysis of Longitudinal Studies and Natural Experiments. *Psychol Med* **2021**, *51*, 201–211, doi:10.1017/S0033291721000015.

66. Bonanno, G.A.; Diminich, E.D. Annual Research Review: Positive Adjustment to Adversity--Trajectories of Minimal-Impact Resilience and Emergent Resilience. *J Child Psychol Psychiatry* **2013**, *54*, 378–401, doi:10.1111/jcpp.12021.
67. Masten, A.S. Resilience of Children in Disasters: A Multisystem Perspective. *International Journal of Psychology* **2021**, *56*, 1–11, doi:10.1002/ijop.12737.
68. Giuntella, O.; Hyde, K.; Saccardo, S.; Sadoff, S. Lifestyle and Mental Health Disruptions during COVID-19. *PNAS* **2021**, *118*, doi:10.1073/pnas.2016632118.