

## Article

# Application of C5.0 Algorithm for the Assessment of Perceived Stress in Healthcare Professionals Attending COVID-19

Juan Luis Delgado-Gallegos <sup>1</sup>, Gener Avilés-Rodríguez <sup>2</sup>, Gerardo R. Padilla-Rivas <sup>1</sup>,  
María De los Ángeles Cosío-León <sup>3</sup>, Héctor Franco-Villareal <sup>4</sup>, Juan Iván Nieto-Hipólito <sup>5</sup>,  
Juan de Dios Sánchez López <sup>5</sup>, Erika Zuñiga-Violante <sup>5</sup>, Jose Francisco Islas <sup>1</sup>,  
and Gerardo Salvador Romo-Cardenas <sup>5,\*</sup>

<sup>1</sup> Departamento de Bioquímica y Medicina Molecular, Facultad de Medicina, Universidad Autónoma de Nuevo León, Monterrey 64260, Mexico

<sup>2</sup> Escuela de Ciencias de la Salud, Universidad Autónoma de Baja California, Ensenada 22890, Mexico

<sup>3</sup> Universidad Politécnica de Pachuca, Carretera, Carretera Ciudad Sahagún-Pachuca Km. 20, Ex-Hacienda de Santa Bárbara, Zempoala 43830, Mexico

<sup>4</sup> Althian Clinical Research, Calle Capitán Aguilar Sur 669, Col. Obispedo, Monterrey 64060, Mexico

<sup>5</sup> Facultad de Ingeniería, Arquitectura y Diseño, Universidad Autónoma de Baja California, Carr. Transpeninsular 391, Ensenada 22860, Mexico

\* Correspondence: romo.gerardo@uabc.edu.mx



**Citation:** Delgado-Gallegos, J.L.; Avilés-Rodríguez, G.; Padilla-Rivas, G.R.; De los Ángeles Cosío-León, M.; Franco-Villareal, H.; Nieto-Hipólito, J.I.; de Dios Sánchez López, J.; Zuñiga-Violante, E.; Islas, J.F.; Romo-Cardenas, G.S. Application of C5.0 Algorithm for the Assessment of Perceived Stress in Healthcare Professionals Attending COVID-19. *Brain Sci.* **2023**, *13*, 513. <https://doi.org/10.3390/10.3390/brainsci13030513>

Academic Editors: Giovanni Martinotti, Mauro Pettorruso, Domenico De Berardis and Drozdstoi Stoyanov

Received: 20 February 2023

Revised: 14 March 2023

Accepted: 15 March 2023

Published: 20 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Coronavirus disease (COVID-19) represents one of the greatest challenges to public health in modern history. As the disease continues to spread globally, medical and allied healthcare professionals have become one of the most affected sectors. Stress and anxiety are indirect effects of the COVID-19 pandemic. Therefore, it is paramount to understand and categorize their perceived levels of stress, as it can be a detonating factor leading to mental illness. Here, we propose a computer-based method to better understand stress in healthcare workers facing COVID-19 at the beginning of the pandemic. We based our study on a representative sample of healthcare professionals attending to COVID-19 patients in the northeast region of Mexico, at the beginning of the pandemic. We used a machine learning classification algorithm to obtain a visualization model to analyze perceived stress. The C5.0 decision tree algorithm was used to study datasets. We carried out an initial preprocessing statistical analysis for a group of 101 participants. We performed chi-square tests for all questions, individually, in order to validate stress level calculation ( $p < 0.05$ ) and a calculated Cronbach's alpha of 0.94 and McDonald's omega of 0.95, demonstrating good internal consistency in the dataset. The obtained model failed to classify only 6 out of the 101, missing two cases for mild, three for moderate and one for severe (accuracy of 94.1%). We performed statistical correlation analysis to ensure integrity of the method. In addition, based on the decision tree model, we concluded that severe stress cases can be related mostly to high levels of xenophobia and compulsive stress. Thus, showing that applied machine learning algorithms represent valuable tools in the assessment of perceived stress, which can potentially be adapted to other areas of the medical field.

**Keywords:** decision tree; COVID-19 stress; healthcare professionals in Mexico; explainable artificial intelligence for healthcare

## 1. Introduction

With the global spread of the COVID-19, both medical and allied healthcare professionals have become the most highly affected sectors [1–3]. In developing democracies, the public health system became engulfed by the overwhelming levels of stress [4,5]. In addition, the situation becomes even more taxing for attending personnel as they not only deal with the burdened system [6] but also with the enemy (COVID-19) upfront. It is here, where they can also become prey to the disease [7]. Recently in Mexico, reports for the period of late February to 23 August showed that over 97,600 healthcare professionals had

become infected with COVID-19 [8]. Hence, Mexico showed atop of all Latin America countries in infection-to-death rate (>10%) [9]. The number of “total confirmed”, possible, active cases, and mortality of COVID-19 amongst physicians, almost doubled during the period of 16 August up to 3 November, potentially generating high levels of stress on them. This is of particular interest when we consider stress as a potential trigger to lose focus during procedures or while attending to patients; therefore, enabling conditions for COVID-19 infection, or making costly mistakes [10].

According to the Pan American Health Organization (PAHO), Mexico has the highest number of healthcare workers infected with COVID-19 in Latin America [11]. In 28 December 2020, the number of health care professionals affected by COVID-19, as reported by the National health ministry, was just over 182,200 [12]. Reports show that both physicians and nurses have similar levels of burnout and emotional fatigue [3,13–15]. Physicians typically work in a more independent manner. This, along with their long shift hours, high-sense of duty, work ethics, and the fact that they partake in multiple jobs normally of low wages, becomes a source of additional stress [8]. With the data being generated while facing the disease, it is important to apply rapid methods that allow study of this scenario and allow development of policies or strategies. For this purpose, machine learning algorithms have proved efficient in the analysis of stress in working employees [16–18]. Still, for medical applications, it is important for the algorithm to provide explainability for computer-aided diagnosis [19]. Therefore, in this study, we propose the use of the C5.0 algorithm to assess perceived stress in healthcare workers exposed to COVID-19, generating an explainable classification diagram that contributes to the understanding of mental health in pandemic scenarios.

Recent developments in computational modeling have led to the ever-evolving field of artificial intelligence, which, when combined with neuro- and behavioral science, has created the new field of computational psychiatry [20,21]. Computational psychiatry helps to model and understand underlying mental illness, allowing the prediction of potential behavioral patterns, improving classification, and assisting the physician to provide a faster and personalized medical attention [22]. Nowadays, machine learning algorithms are promising technologies used by various healthcare providers, as they result in better scale-up, speed-up, processing power, and reliability, which translates into a more efficient performance of the clinical team [23–25]. Therefore, a trend is to use these techniques to better understand, and fight against the current pandemic and other chronic diseases, especially when the resulting model could have a graphical-based explanation [26]. Using well-known machine learning algorithms, such as decision trees, for establishing classification systems are but one of the many features of their application [27]. Typically, it is possible to classify a population into branch-like segments that generate an inverted tree [27,28]. These algorithms can efficiently deal with large, complicated datasets without imposing a complicated parametric structure [28]. Researchers have reported the use of these types of algorithms for applications in the study of behavioral and mental health [29] and on the use of computational based methods to classify stress from data generated by sensor devices [30]. Thus, it is possible to use these tools to better understand disease and propose different clinical paths, and to classify subgroups of patients for different diagnostic tests, treatment strategies, and assessment of mental health-related conditions [31,32].

Several approaches on machine learning-based stress assessments have been reported. A common method considers the use of bio-marker data to stratify stress on several levels [33,34], since these algorithms are able to assess not only stress, but depression and anxiety as well [18,35]. For COVID-19-related stress evaluation, the use of these type of algorithms has been previously explored for general population studies, based on distributed questionnaires data [36,37], which allows for exploration of data acquired from these kinds of questionnaires for clinical applications [38].

Specifically, decision tree algorithms have been able to obtain 92% accuracy, providing not only a reliable stress categorization [39], but they also generate a visual model that

allows to analysis of the actual scenario of the problem, which is not common with machine learning algorithms.

In machine learning, a common strategy used for data analytics is the cross-industry standard process for data mining method (CRISP-DM). This method defines six steps for data-based knowledge projects. This strategy begins with defining problems and objectives (business understanding), followed by data insights (data understanding). Next, defining a dataset and its analysis (data preparation), and results from this analysis generates a model (modeling). Once generated, it is evaluated (evaluation), and if the goal is achieved, it can be implemented [40].

Given that it is possible to use decision tree algorithms to identify prominent features that influence stress [16], it is feasible to apply this type of algorithms to obtain an explicative model of the studied scenario. Additionally, the proven efficiency of the C5.0 algorithm as a biomedical decision support tool for assisted diagnosis makes it a likely tool for the case [41–43]. In the current work, we studied the application of a C5.0 decision tree algorithm, as proposed in the literature, to locate the combination of factors needed to classify, correctly, healthcare professionals attending to COVID-19 patients, by the category of perceived stress. This provides a graphical tool that allows a better understanding of the mental health of healthcare professionals at the beginning of the COVID-19 pandemic in northeast Mexico.

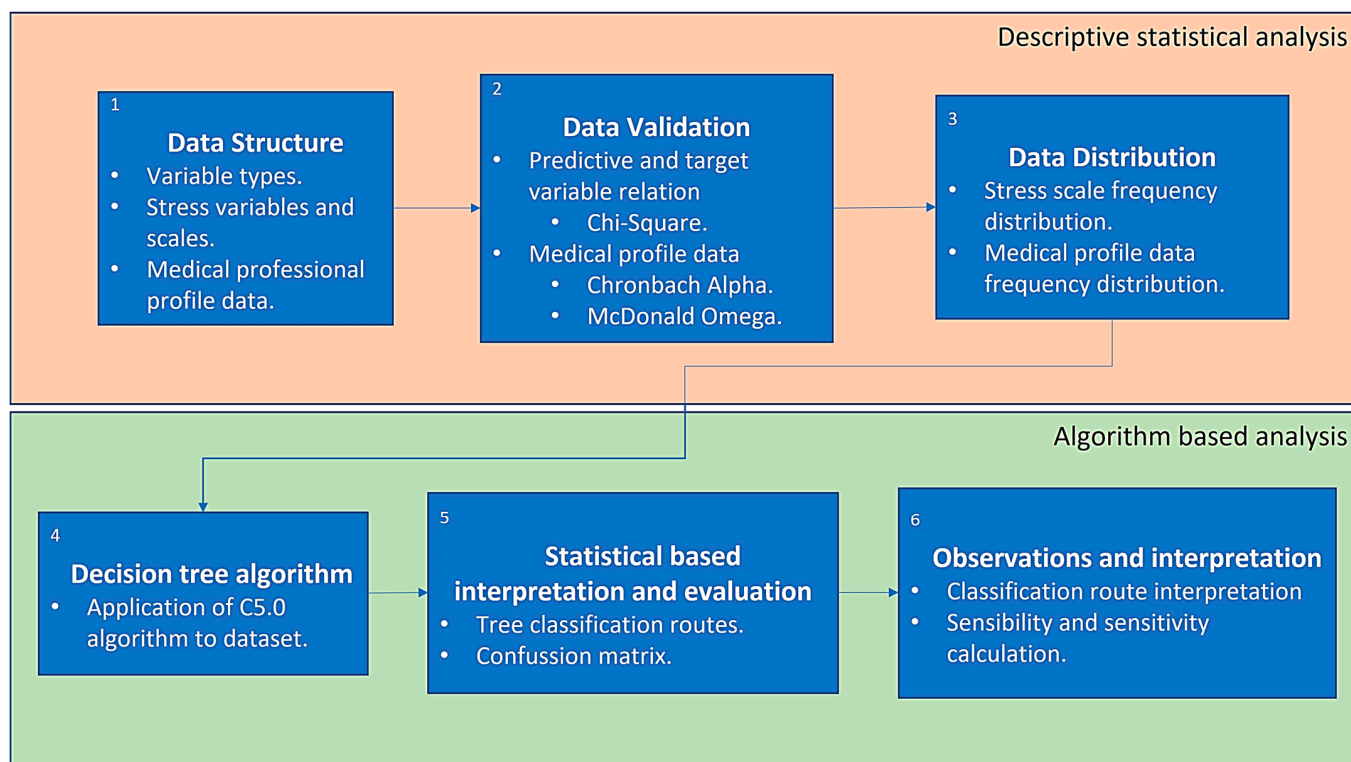
## 2. Materials and Methods

Some other work on stress perception during the COVID-19 pandemic has been reported regarding healthcare workers [44,45]. Our work is based on previously reported adapted COVID-19 stress scale (ACSS) data [1], at the beginning of the pandemic, in healthcare workers in northeast Mexico. The dataset was previously classified into different categories of perceived stress for healthcare professionals attending to COVID-19 patients: five variables were defined (danger and contamination, xenophobia, traumatic stress, compulsive checking, and social economical) and four different results were defined, with scores per area: 0–6 absent, 7–23 mild, 13–18, moderate, 19–24 severe. A total tallied score of all the areas was obtained, and further analyzed and correlated in accordance to job-specific characteristics [1]. We adapted the analysis method using the CRISP-DM model, commonly used in data analytics [40], having the same number of stages and sequences, as shown in Figure 1. Initially, we performed a data structure study from the data analytics scope to consider the type of variables from the ACSS. This was to establish the type of variables and how they contributed to the context of the ACSS, along with the four categories of stress defined from the scores as outcomes: absent, mild, moderate and severe.

Next, we performed a data validation analysis considering statistical tests to confirm relations between variables from the scales and classification outcomes from the raw data, and to confirm internal consistency [46]. This was completed by obtaining both Cronbach's alpha and McDonald's omega from the raw ACSS responses and a Pearson chi-square statistic applied to the ACSS and the resulting stress scale. We followed the validation process with a data distribution analysis to study stress components for model selection and interpretation. This measured the central tendency of the professional profile, which included the profession and work area from the healthcare workers who participated in the study, as well as for the ACSS and the resulting stress class.

Given that the approach of this work is to provide an AI-based method that could become a tool for clinical decision making, we selected a decision tree (DT) model to study the relations and classification routes for stress level according to data from its respective scales.

We carried out an accuracy analysis based on the results from algorithm training, as well as a sensitivity and specificity analysis by splitting the categories defined for stress into different subgroups for healthy and disease states.



**Figure 1.** Methods for machine learning-based analysis on stress scales of healthcare workers.

### 2.1. Descriptive Statistical Analysis

We performed both the statistical and algorithm performance analysis in R language to obtain behavior patterns and understanding of data distribution. For data preparation and preprocessing, we also carried out a descriptive statistical study to understand data structure and distribution. To obtain valuable information for model interpretation, measures of central tendency were obtained from the professional profile data of the healthcare workers who participated in the study, as well as from the ACSS.

For the instrument validation purposes, we estimated the value of Cronbach's alpha considering the numerical values from all participant responses [47]. Finally, we applied Pearson chi-square statistics using SPSS (ver. 21) to the ACSS areas to show results robustness [48].

### 2.2. Application of C5.0 Algorithm

Following the statistical analysis on the instrument results, we developed a DT to behave as a computational supportive scaffold for the study of mental illness. We opted to use a C5.0 algorithm to analyze and classify the stress level from the dataset and for construction of a classification tree, as used in previous health-related scenarios [49]. This algorithm uses information gain as its splitting criteria and the binomial confidence limit method for the pruning technique, improving the feature selection and reducing error pruning. These methods have been reported useful to build efficient classifying models having small datasets, given the mathematical background of the model [50]. Additionally, DT outperforms other algorithms with smaller datasets, as in this case.

Following both the statistical and computational analysis of the instrument and dataset, we analyzed the performance based on sensitivity and accuracy on the generated model [51]. Given the size of the dataset, the confusion matrix obtained from the algorithm training was used to define the accuracy of the obtained model. Then, sensitivity and specificity calculations were completed using the results of the confusion matrix. Given that there are four different levels of stress defined as outcome, three different combination subgroups were used to define healthy and disease states. Conclusions were drawn from the results

of the analysis, as well as routes defined by the tree model branches considering initial statistical analysis. The application of this algorithm is not intended as a classification tool but as a computer-aided tool that provides a wider scope of stress in healthcare workers. For this, the whole dataset was used to train the algorithm and to obtain the DT with the use of R and RStudio.

### 2.3. Dataset

As mentioned before, the study considers a dataset obtained from 106 participants from which information related to medical or healthcare education, work field and experience. Then, the data is built into a stress concept conformed by five components, which are: danger + fear of contamination, socioeconomical, xenophobia, traumatic stress and compulsive checking. Danger + fear of contamination refers to perceived stress related to the probability of being exposed and contracting the disease. The socioeconomical factor refers to financial-related stress that is associated with the chance of losing their job and the financial burden of becoming unemployed. Xenophobia is a scale that refers to the fact that the disease comes from abroad and it might not be possible to stop it. Traumatic stress refers to the emotional burden related to work with COVID-19 patients, and compulsive checking it related to compulsive behavior around the need to look for information about the disease.

### 3. Results

We applied an initial preprocessing statistical analysis to the 106 entries dataset. After eliminating missing data entries for statistical and algorithm-based analysis preparation, we used a group of 101 entries for the study. Besides explainability from the graphical output, decision trees have proved useful for small datasets [52]. Still, the dataset is greater than the minimal size of 62 required for decision tree models [50].

From the total entries, we counted the frequency of the profession and work area variables, as shown in Table 1.

**Table 1.** Frequency count on participant profession and work area.

Participant Profession	Counts	Participant Work Area	Counts
Medical Student	2	Front line health professional	29
Nursing Staff	10	Others	34
Physician	69	COVID-19 designated area	11
Physician in community service *	4	Surgical	11
Resident	15	ER	9
Technician	2	Internal medicine	8

\* Physician in community service. A medical student who has finished the required medical school training in Mexico and is doing a compulsory one-year internship at a local community hospital or healthcare facility.

We then built upon the five areas of the ACSS, calculating the central tendency metrics for each of these components based on the cumulative result of each participant, as shown in Table 2.

**Table 2.** Central tendency metrics for the adapted COVID-19 stress scale features.

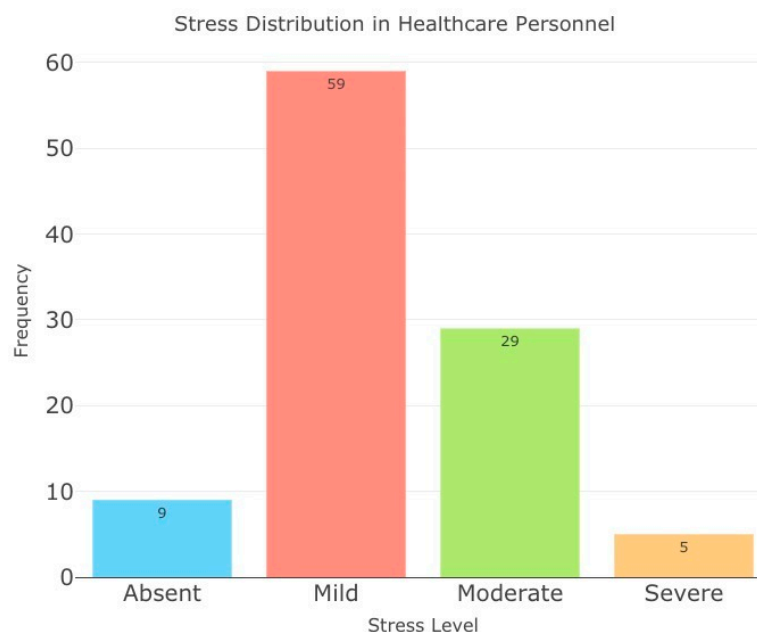
Stress Scale Feature	Min	1st Quartile	Median	Mean	3rd Quartile	Max
Danger + fear of contamination	5	23	25	25.2	33.75	48
Socioeconomical	4	14	17	16.27	19	24
Xenophobia	1	7	10.5	10.9	14	24
Traumatic stress	0	2	6	7.37	12	22
Compulsive checking	0	5	8	9.38	13.75	24

Given that we based each feature on the addition of the responses from the survey, we considered all the values from each question and participant for the calculation of Cronbach's alpha, which shows a good internal consistency (0.94) for the whole survey instrument and data, and a similar result for McDonald's omega (0.95). In addition, Supplementary Table S1 shows chi-square tests to each question in order to define significance in the relationship of the variables. Table 3 shows the result of the test for each scale area and each question, and for the cumulative ACSS.

**Table 3.** Analysis per general area.

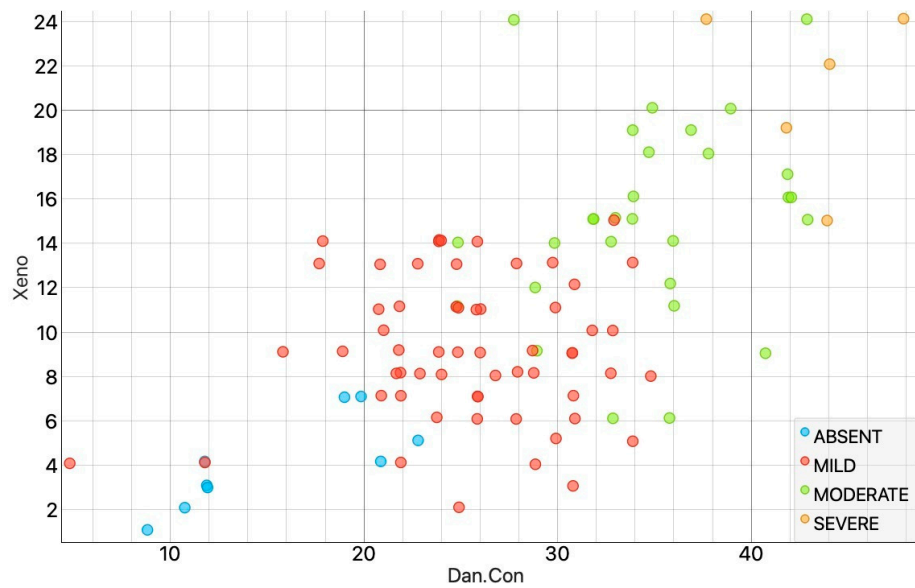
COVID Areas	Absent	Mild	Moderate	Severe	Xi2	Sig
Danger + fear of contamination	3	23	58	17	64.98	<0.001
Socioeconomical	30	35	24	12	11.673	<0.009
Xenophobia	15	45	29	12	27.119	<0.001
Traumatic stress	47	25	21	8	31.238	<0.001
Compulsive checking	26	43	22	10	22.129	<0.001
CSS general score	9	59	28	5	72.109	<0.001
Xi2 = chi-square test						

Both results from Cronbach's alpha and the chi-square test show internal consistency of the data and validate the dependence for stress level calculation, ensuring the dataset quality for algorithm-based analysis. Distribution for the stress level classification in healthcare personnel calculated from the ACSS is shown in Figure 2.



**Figure 2.** Stress level distribution in healthcare personnel. (Left to right) Absent, Mild, Moderate, Severe.

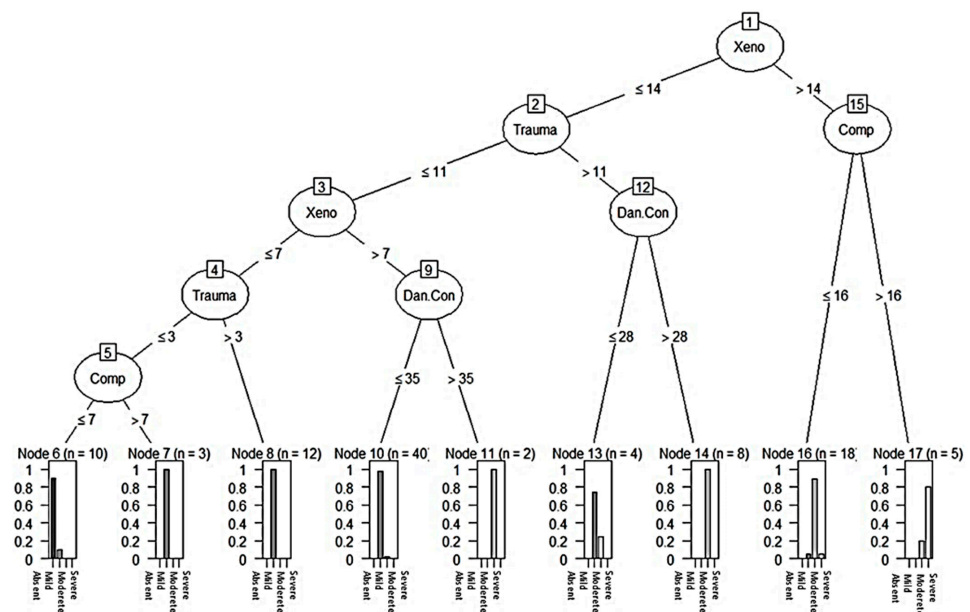
Figure 3 shows a scatter plot from the intersection from the xenophobia and danger + fear of contamination scales from the ACSS, allowing to observe the distribution of the stress levels based on these two variables in some areas of the graph.



**Figure 3.** Stress level distribution in healthcare personnel from the intersection of xenophobia and danger + fear of contamination scales from the ACSS.

Stress scale distribution showed in Figure 2 shows the general incidence of the stress level in healthcare professionals at the beginning of the pandemic. Although imbalanced, commonly in medical data, correlation distribution showed in Figure 3 confirms the feasibility to use the dataset, despite the size and imbalance, for the purpose to decipher medical context [53].

Following the descriptive statistical analysis, we trained a decision tree model with the preprocessed dataset ( $n = 101$ ) using the C5.0 algorithm [28,49], considering the stress level to be the target variable. We used all areas of the ACSS including participant profession and work area as the predictive variables to find any relationship between them to predict stress level. Figure 4 shows the decision tree obtained from the dataset.



**Figure 4.** Decision tree applied into healthcare personnel stress scale level dataset. Atop variables influencing stress are xenophobia (Xeno) and compulsive checking (Comp), which leads to severe stress. Traumatic stress (Trauma) and danger + contamination (Dan Con) also influenced the perception of stress. The socioeconomical variable did not influence the outcome of the decision tree.

At the class level, a set of boxes with all four levels of stress are observed. In each box, the extreme right bar corresponds to the severe level indicator, followed, to the left, by moderate, mild and absent levels, respectively. Despite declaring the features related to the participant profession and work area, these variables did not provide valuable information gain to be considered in the model. Table 4 shows the confusion matrix from the obtained decision tree model, where only 6 out of 101 entries were incorrectly classified, missing two cases for mild level, three for moderate and one for severe. All these errors were classified only in neighboring levels, giving the model an accuracy of 94.1%.

**Table 4.** Confusion matrix of obtained decision tree model for stress level classification.

		Classified as				
		(a)	(b)	(c)	(d)	Actual Class
	9					(a) Absent
	1		57	1		(b) Mild
			2	26	1	(c) Moderate
				1	4	(d) Severe

To analyze model performance, a sensitivity and specificity calculation were carried out. For these, three different scenarios were considered based on the stress classification outcome from the dataset, dividing entries into healthy and disease groups. Calculation was completed with the figures from the confusion matrix. Results are shown in Table 5.

**Table 5.** Decision tree sensitivity and specificity calculation from stress scales dataset.

	Healthy: Absent + Mild + Moderate Disease: Severe	Healthy: Absent + Mild Disease: Moderate + Severe	Healthy: Absent Disease: Mild + Moderate + Severe
Sensitivity	0.8	0.91	0.989
Specificity	0.989	0.98	0.9

#### 4. Discussion

Our purpose was to define a statistical and computational framework algorithm to analyze and understand stress levels in healthcare professionals for the impact of the COVID-19 pandemic and to potentially define a graphical self-explainable clinical tool, which can be further used as a severity predictor of stress.

A dataset related to the ACSS, as defined by Delgado-Gallegos et al., was studied with a calculated Cronbach's alpha of 0.94, which shows a good internal consistency; stress levels were calculated as a geometrical result from the addition of five scales from the survey defined as danger + fear of contamination, socioeconomic stress, xenophobia, traumatic stress and compulsive checking. Chi-square tests were carried out for all questions individually, looking to validate stress level calculation. Statistical significance ( $p < 0.05$ ) was found in most of the questions, considering the answers of all participants, except one question for the traumatic scale, and four for the compulsive checking scale (all shown in Supplementary Table S1). However, all scales showed statistical significance when the test was applied to the accumulated value for each of these scales, as seen in Table 3; thus, validating, the use of the ACSS in a population [1,54]. Therefore, the use of this model can be re-adapted to help in correctly assessing and providing a faster diagnosis and opportune treatment.

From the central tendency metrics statistical analysis, no relation was observed between participant profession and work area, similar analysis was done for the stress scales which showed an exception for danger + fear of contamination joint scale, all other areas



had a similar maximum value but with different means. Therefore, considering the results from the preprocessing stage, the dataset shows good quality, independence, and internal consistency for algorithm analysis. All 101 entries from the dataset were used to train a decision tree model by the C5.0 algorithm, where stress level was defined as the target variable, with participant profession, work area, and cumulative stress scales as predictors. The resulting model showed an accuracy of 94%, adding a more precise assessment to the initial stress classification. Nonetheless, the algorithm did not find enough information gain from the participant profession, work area, and the socioeconomic scale. Neglecting these variables from the resulting model allows to understand that experience and day-to-day work routine are not a factor on how healthcare professionals perceive stress. Resilience could help explain this pattern, as it is an adaptation mechanism in which a person, overtime, can handle stress in overwhelming situations [15,55].

Computational psychiatry states the similarity between the brain and a computer and proposes the use of computational terminology for the study of mental illness [56]. Our results show interesting data denoting hypothetical tendencies based on the purity of the resulting branches of the decision tree, where severe stress cases can be related mostly to high levels of xenophobia and compulsive stress, as shown by the relation of the threshold values from the extreme right route of the decision tree, which are above the 3rd quartile for xenophobia and compulsive stress scales, and from the measures of central tendency shown in Table 2. In a similar manner, absent stress level comes from the scenario of combined thresholds below the 1st quartile from xenophobic, compulsive and traumatic stress scales. It is interesting to note that the danger + fear of contamination scale can be used to find both mild and moderate cases, despite being a larger joint scale.

Even though there are various classification algorithms, such as K-Nearest Neighbors, Support Vector Machines, Naive Bayes, Random Forest, Radial Basis Function or Adaptive Boosting (AdaBoost) that are used for classification process with prominent accuracy and performance, it has been previously reported that with the use of decision tree algorithms, it is possible to rely on a few variables from a health-related problem to stratify patients with a visual tool that empowers clinical decision [41–43]. Given the size of our dataset ( $n = 101$ ), this constitutes an efficient input for the C5.0 algorithm, which was further confirmed with the sensitivity and specificity analysis. In addition, the sensitivity and specificity analysis showed acceptable results despite the few severe stress cases. Supplementary content shows the analysis of the studied dataset with the algorithms mentioned above.

Currently, machine learning and decision tree algorithms are still in their initial stages of application in the medical field. Recently, Yu et al. published a retrospective study on the conditions to predict metabolic syndrome [57] and Peng et al. published a recent study on the prediction of exacerbation of chronic obstructive pulmonary disease using key indicators, the result had an overall accuracy of 80.3% with a confidence level of 95% [58]. Machine learning has also been used to identify complex patterns in emergency hospital services, which implies intelligence data-driven decisions even under overwhelming circumstances [59].

## 5. Conclusions

This is only a first approximation based on recent data from healthcare professionals in the northeast part of Mexico [1] and the first study of its kind using the C5.0 decision tree algorithm model on the assessment of stress on self-explainable model basis. Because of the mathematical foundation of these algorithms, it allows not only to obtain a better understanding of a problem, but also to generate accurate predictions. The need of larger datasets and machine learning methodological approaches is well established. Therefore, the impact of applying machine learning algorithms represents a window of opportunity in actual global health and in the decision-making process of developing health policies, based on large-scale studies. For clinical decision-making scenarios, decision trees are specifically useful to simplify assisted diagnosis given the ease of understanding, expanding the scope of computer assisted diagnosis.

This work contributes to mathematical-informed understanding of mental illness and computational psychiatry; thus, forming a diagnostic tool to help in the assessment of patients. In this study, we analyzed healthcare professionals' answers, as they are one of the most affected sectors in the pandemic [60]. In addition, an expansion of this method with the use of algorithm combinations could provide efficient clinical-assisted tools that could apply to scenarios of the internet of medical things; real-time measurement of compounds or metabolites could be analyzed to decipher medical context, as in this work, or even to reach customizable medicine. In addition to uses from the COVID-19 pandemic, it can be used to understand different stress factors and how they can interfere with performance and the social dynamics in different populations.

## 6. Limitation

The main goal of this work is to show that the mathematical-/computer-based analysis applied to a very specific population allowed to identify patterns in behavior and mental health, despite the fact that the sample size could not be big enough for a formal data analytics study. Applying synthetic methods to increase sample size or to balance the target variable could affect the actual scenario of the data from the population analyzed during a small and very specific period of time, making the founded patterns meaningless. The use of a decision tree to the diagnosed population during the COVID-19 pandemic contributes to the understanding of mental health and behavioral patterns within an emblematic event in human history.

A formal analytics study was added as supplementary material. The application for computer-aided diagnosis is suggested for future work.

**Supplementary Materials:** The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/brainsci13030513/s1>, Figure S1: shows the box plot built upon the five areas of the ACSS, de-rived from the calculation of the central tendency metrics for each of these components based on the cumulative result of each participant; Table S1: Performance Evaluation of Algorithms Ex-plored.

**Author Contributions:** Research and writing, J.L.D.-G., J.F.I., G.A.-R., E.Z.-V., G.S.R.-C. and H.F.-V.; algorithm development, M.D.I.Á.C.-L., G.S.R.-C., G.A.-R., J.I.N.-H., J.d.D.S.L., statistical analysis, J.L.D.-G. and G.R.P.-R.; supervision, J.F.I., J.I.N.-H., J.d.D.S.L. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** This study was conducted in accordance to the declaration of Helsinki, and the protocol was approved by the independent ethics committee of Hospital La Mision, Monterrey, Nuevo León, México under protocol number PSY-CSS-ESP-001.

**Data Availability Statement:** Dataset may be downloaded from Kaggle (<https://www.kaggle.com/chepox/css-mexico>, accessed on 19 January 2021). Code may be accessed from github (<https://github.com/Bio-Math/COVID19-Stress-Health-Professionals>) (19 January 2021).

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

## References

1. Delgado-Gallegos, J.L.; de Montemayor-Garza, R.J.; Padilla-Rivas, G.R.; Franco-Villareal, H.; Islas, J.F. Prevalence of stress in healthcare professionals during the COVID-19 pandemic in Northeast Mexico: A remote, fast survey evaluation, using an adapted covid-19 stress scales. *Int. J. Environ. Res. Public Health* **2020**, *17*, 7624. [[CrossRef](#)]
2. Shah, K.; Chaudhari, G.; Kamrai, D.; Lail, A.; Patel, R.S. How essential is to focus on physician's health and burnout in coronavirus (COVID-19) pandemic? *Cureus* **2020**, *12*, e7538. [[CrossRef](#)]
3. Petzold, M.B.; Plag, J.; Ströhle, A. Dealing with psychological distress by healthcare professionals during the COVID-19 pandemia. *Nervenarzt* **2020**, *91*, 417–421. [[CrossRef](#)]
4. Morales, G.; COVID-19 Death Toll in MEXICO. El Universal 2020. Available online: <https://www.eluniversal.com.mx/english/live-updates-covid-19-death-toll-mexico> (accessed on 19 January 2021).

5. Bello-Chavolla, O.Y.; Bahena-López, J.P.; Antonio-Villa, N.E.; Vargas-Vázquez, A.; González-Díaz, A.; Márquez-Salinas, A.; Fermín-Martínez, C.A.; Naveja, J.J.; Aguilar-Salinas, C.A. Predicting mortality due to SARS-CoV-2: A mechanistic score relating obesity and diabetes to COVID-19 outcomes in Mexico. *J. Clin. Endocrinol. Metab.* **2020**, *105*, 2752–2761. [CrossRef]
6. Burki, T. COVID-19 in latin america. *Lancet Infect. Dis.* **2020**, *20*, 547–548. [CrossRef] [PubMed]
7. Shah, K.; Kamrai, D.; Mekala, H.; Mann, B.; Desai, K.; Patel, R.S. Focus on mental health during the coronavirus (COVID-19) pandemic: Applying learnings from the past outbreaks. *Cureus* **2020**, *12*, e7405. [CrossRef] [PubMed]
8. Agren, D. Understanding Mexican health worker COVID-19 deaths. *Lancet* **2020**, *396*, 807. [CrossRef] [PubMed]
9. CONACYT COVID-19 Mexico. Gob. de Mexico. 2020. Available online: <https://coronavirus.gob.mx/datos/> (accessed on 19 January 2021).
10. de Salud, S. Personal de Salud 03 de Noviembre de 2020. Gob. de Mexico. 2020. Available online: [https://www.gob.mx/cms/uploads/attachment/file/590340/COVID-19\\_Personal\\_de\\_Salud\\_2020.11.03.pdf](https://www.gob.mx/cms/uploads/attachment/file/590340/COVID-19_Personal_de_Salud_2020.11.03.pdf) (accessed on 19 January 2021).
11. PAHO Epidemiological Alert: COVID-19 among Health Workers—31 August 2020—PAHO/WHO | Pan American Health Organization. 2020. Available online: <https://www.paho.org/en/documents/epidemiological-alert-covid-19-among-health-workers-31-august-2020> (accessed on 8 July 2021).
12. de Salud, S. Datos Abiertos Dirección General de Epidemiología | Secretaría de Salud | Gobierno | gob.mx. Available online: <https://www.gob.mx/salud/documentos/datos-abiertos-152127> (accessed on 8 July 2021).
13. Hamama, L.; Hamama-Raz, Y.; Stokar, Y.N.; Pat-Horenczyk, R.; Brom, D.; Bron-Harlev, E. Burnout and perceived social support: The mediating role of secondary traumatization in nurses vs. physicians. *J. Adv. Nurs.* **2019**, *75*, 2742–2752. [CrossRef]
14. Lai, J.; Ma, S.; Wang, Y.; Cai, Z.; Hu, J.; Wei, N.; Wu, J.; Du, H.; Chen, T.; Li, R. Factors associated with mental health outcomes among health care workers exposed to coronavirus disease 2019. *JAMA Netw. open* **2020**, *3*, e203976. [CrossRef]
15. Labrague, L.J.; De los Santos, J.A.A. COVID-19 anxiety among front-line nurses: Predictive role of organisational support, personal resilience and social support. *J. Nurs. Manag.* **2020**, *28*, 1653–1661. [CrossRef]
16. Reddy, U.S.; Thota, A.V.; Dharun, A. Machine learning techniques for stress prediction in working employees. In Proceedings of the 2018 IEEE International Conference on Computational Intelligence and Computing Research (ICIC), Madurai, India, 13–15 December 2018; pp. 1–4.
17. Galatzer-Levy, I.R.; Ma, S.; Statnikov, A.; Yehuda, R.; Shalev, A.Y. Utilization of machine learning for prediction of post-traumatic stress: A re-examination of cortisol in the prediction and pathways to non-remitting PTSD. *Transl. Psychiatry* **2017**, *7*, e1070. [CrossRef] [PubMed]
18. Hasanin, T.; Kshirsagar, P.R.; Manoharan, H.; Sengar, S.S.; Selvarajan, S.; Satapathy, S.C. Exploration of Despair Eccentricities Based on Scale Metrics with Feature Sampling Using a Deep Learning Algorithm. *Diagnostics* **2022**, *12*, 2844. [CrossRef] [PubMed]
19. Holzinger, A.; Langs, G.; Denk, H.; Zatloukal, K.; Müller, H. Causability and explainability of artificial intelligence in medicine. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* **2019**, *9*, e1312. [CrossRef] [PubMed]
20. Montague, P.R.; Dolan, R.J.; Friston, K.J.; Dayan, P. Computational psychiatry. *Trends Cogn. Sci.* **2012**, *16*, 72–80. [CrossRef]
21. Mujica-Parodi, L.R.; Strey, H.H. Making Sense of Computational Psychiatry. *Int. J. Neuropsychopharmacol.* **2020**, *23*, 339–347. [CrossRef] [PubMed]
22. Schulz, E.; Dayan, P. Computational Psychiatry for Computers. *Iscience* **2020**, *23*, 101772. [CrossRef] [PubMed]
23. Davenport, T.; Kalakota, R. The potential for artificial intelligence in healthcare. *Future Healthc. J.* **2019**, *6*, 94–98. [CrossRef]
24. Cutillo, C.M.; Sharma, K.R.; Foschini, L.; Kundu, S.; Mackintosh, M.; Mandl, K.D.; MI in Healthcare Workshop Working Group. Machine intelligence in healthcare—Perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit. Med.* **2020**, *3*, 47. [CrossRef]
25. Bhavsar, K.A.; Abugabah, A.; Singla, J.; AlZubi, A.A.; Bashir, A.K. A comprehensive review on medical diagnosis using machine learning. *Comput. Mater. Contin.* **2021**, *67*, 1997. [CrossRef]
26. London, A.J. Artificial intelligence and black-box medical decisions: Accuracy versus explainability. *Hastings Cent. Rep.* **2019**, *49*, 15–21. [CrossRef]
27. Kelleher, J.D.; Mac Namee, B.; D’Arcy, A. Fundamentals of machine learning for predictive data analytics: Algorithms. In *Worked Examples, and Case Studies*; MIT Press: Cambridge, MA, USA, 2015.
28. Song, Y.-Y.; Ying, L.U. Decision tree methods: Applications for classification and prediction. *Shanghai Arch. Psychiatry* **2015**, *27*, 130.
29. Zhu, T.; Ning, Y.; Li, A.; Xu, X. Using decision tree to predict mental health status based on web behavior. In Proceedings of the 2011 3rd Symposium on Web Society, Chicago, IL, USA, 23–25 May 2011; pp. 27–31.
30. Sharma, N.; Gedeon, T. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Comput. Methods Programs Biomed.* **2012**, *108*, 1287–1301. [CrossRef]
31. Li, C.; Glüer, C.-C.; Eastell, R.; Felsenberg, D.; Reid, D.M.; Roux, C.; Lu, Y. Tree-structured subgroup analysis of receiver operating characteristic curves for diagnostic tests. *Acad. Radiol.* **2012**, *19*, 1529–1536. [CrossRef]
32. Magyary, D.; Brandt, P. A decision tree and clinical paths for the assessment and management of children with ADHD. *Issues Ment. Health Nurs.* **2002**, *23*, 553–566. [CrossRef] [PubMed]
33. Nath, R.K.; Thapliyal, H.; Caban-Holt, A.; Mohanty, S.P. Machine learning based solutions for real-time stress monitoring. *IEEE Consum. Electron. Mag.* **2020**, *9*, 34–41. [CrossRef]

34. Subhani, A.R.; Mumtaz, W.; Saad, M.N.B.M.; Kamel, N.; Malik, A.S. Machine learning framework for the detection of mental stress at multiple levels. *IEEE Access* **2017**, *5*, 13545–13556. [[CrossRef](#)]
35. Kumar, P.; Garg, S.; Garg, A. Assessment of anxiety, depression and stress using machine learning models. *Procedia Comput. Sci.* **2020**, *171*, 1989–1998. [[CrossRef](#)]
36. Flesia, L.; Monaro, M.; Mazza, C.; Fietta, V.; Colicino, E.; Segatto, B.; Roma, P. Predicting perceived stress related to the Covid-19 outbreak through stable psychological traits and machine learning models. *J. Clin. Med.* **2020**, *9*, 3350. [[CrossRef](#)]
37. Li, H.; Zheng, E.; Zhong, Z.; Xu, C.; Roma, N.; Lamkin, S.; Von Visger, T.T.; Chang, Y.-P.; Xu, W. Stress prediction using micro-EMA and machine learning during COVID-19 social isolation. *Smart Health* **2021**, *23*, 100242. [[CrossRef](#)] [[PubMed](#)]
38. Padilla-Rivas, G.R.; Delgado-Gallegos, J.L.; Montemayor-Garza, R.D.J.; Franco-Villareal, H.; Coiser-León, M.D.L.Á.; Avilés-Rodríguez, G.; Zuñiga-Violante, E.; Romo-Cardenas, G.S.; Islas, J.F. Dataset of the adapted COVID STRESS SCALES for Healthcare professionals of the Northeast region of Mexico. *Data Br.* **2021**, *34*, 106733. [[CrossRef](#)]
39. Stewart, R.W.; Tuerk, P.W.; Metzger, I.W.; Davidson, T.M.; Young, J. A decision-tree approach to the assessment of posttraumatic stress disorder: Engineering empirically rigorous and ecologically valid assessment measures. *Psychol. Serv.* **2016**, *13*, 1–9. [[CrossRef](#)] [[PubMed](#)]
40. Wirth, R.; Hipp, J. CRISP-DM: Towards a standard process model for data mining. In Proceedings of the 4th International Conference on the Practical Applications of Knowledge Discovery and Data Mining, Crowne Plaza Midland Hotel, Manchester, UK, 11–13 April 2000; Springer: London, UK, 2000; Volume 1.
41. Rafe, V.; Farhoud, S.H.; Rasoolzadeh, S. Breast cancer prediction by using C5. 0 Algorithm and BOOSTING Method. *J. Med. Imaging Health Inform.* **2014**, *4*, 600–604. [[CrossRef](#)]
42. Ahmadi, E.; Weckman, G.R.; Masel, D.T. Decision making model to predict presence of coronary artery disease using neural network and C5. 0 decision tree. *J. Ambient Intell. Humaniz. Comput.* **2018**, *9*, 999–1011. [[CrossRef](#)]
43. Pashaei, E.; Ozen, M.; Aydin, N. Improving medical diagnosis reliability using Boosted C5. 0 decision tree empowered by Particle Swarm Optimization. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 7230–7233.
44. Ruiz-Fernández, M.D.; Ramos-Pichardo, J.D.; Ibáñez-Masero, O.; Cabrera-Troya, J.; Carmona-Rega, M.I.; Ortega-Galán, Á.M. Compassion fatigue, burnout, compassion satisfaction and perceived stress in healthcare professionals during the COVID-19 health crisis in Spain. *J. Clin. Nurs.* **2020**, *29*, 4321–4330. [[CrossRef](#)]
45. Bareeqa, S.B.; Ahmed, S.I.; Samar, S.S.; Yasin, W.; Zehra, S.; Monese, G.M.; Gouthro, R.V. Prevalence of depression, anxiety and stress in china during COVID-19 pandemic: A systematic review with meta-analysis. *Int. J. Psychiatry Med.* **2021**, *56*, 210–227. [[CrossRef](#)]
46. Moret, L.; Mesbah, M.; Chwalow, J.; Lellouch, J. Internal validation of a measurement scale: Relation between principal component analysis, Cronbach's alpha coefficient and intra-class correlation coefficient. *Rev. Epidemiol. Sante Publique* **1993**, *41*, 179–186. [[PubMed](#)]
47. Tavakol, M.; Dennick, R. Making sense of Cronbach's alpha. *Int. J. Med. Educ.* **2011**, *2*, 53. [[CrossRef](#)]
48. Sharpe, D. Chi-square test is statistically significant: Now what? *Pract. Assess. Res. Eval.* **2015**, *20*, 8.
49. Yao, Z.; Liu, P.; Lei, L.; Yin, J. R-C4. 5 Decision tree model and its applications to health care dataset. In Proceedings of the ICSSM'05. 2005 International Conference on Services Systems and Services Management, Chongqing, China, 13–15 June 2005; Volume 2, pp. 1099–1103.
50. van der Ploeg, T.; Austin, P.C.; Steyerberg, E.W. Modern modelling techniques are data hungry: A simulation study for predicting dichotomous endpoints. *BMC Med. Res. Methodol.* **2014**, *14*, 137. [[CrossRef](#)]
51. Zhu, W.; Zeng, N.; Wang, N. Sensitivity, specificity, accuracy, associated confidence interval and ROC analysis with practical SAS implementations. *NESUG Proc. Health Care Life Sci. Baltim. Md.* **2010**, *19*, 67.
52. Priyam, A.; Abhijeeta, G.R.; Rathee, A.; Srivastava, S. Comparative analysis of decision tree classification algorithms. *Int. J. Curr. Eng. Technol.* **2013**, *3*, 334–337.
53. Ramyachitra, D.; Manikandan, P. Imbalanced dataset classification and solutions: A review. *Int. J. Comput. Bus. Res.* **2014**, *5*, 1–29.
54. Taylor, S.; Landry, C.A.; Paluszek, M.M.; Fergus, T.A.; McKay, D.; Asmundson, G.J.G. Development and initial validation of the COVID Stress Scales. *J. Anxiety Disord.* **2020**, *72*, 102232. [[CrossRef](#)]
55. Wang, C.; Pan, R.; Wan, X.; Tan, Y.; Xu, L.; Ho, C.S.; Ho, R.C. Immediate psychological responses and associated factors during the initial stage of the 2019 coronavirus disease (COVID-19) epidemic among the general population in China. *Int. J. Environ. Res. Public Health* **2020**, *17*, 1729. [[CrossRef](#)]
56. Huys, Q.J.M.; Maia, T.V.; Frank, M.J. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat. Neurosci.* **2016**, *19*, 404–413. [[CrossRef](#)]
57. Yu, C.-S.; Lin, Y.-J.; Lin, C.-H.; Wang, S.-T.; Lin, S.-Y.; Lin, S.H.; Wu, J.L.; Chang, S.-S. Predicting metabolic syndrome with machine learning models using a decision tree algorithm: Retrospective cohort study. *JMIR Med. Inform.* **2020**, *8*, e17110. [[CrossRef](#)]
58. Peng, J.; Chen, C.; Zhou, M.; Xie, X.; Zhou, Y.; Luo, C.-H. A machine-learning approach to forecast aggravation risk in patients with acute exacerbation of chronic obstructive pulmonary disease with clinical indicators. *Sci. Rep.* **2020**, *10*, 3118. [[CrossRef](#)] [[PubMed](#)]

- 
59. Gopinath, M.P.; Satyam, S.C.; Jenil, S.M.; Shashank, P. Predictive Analysis of COVID-19 Pandemic in India Based on SIR-F Model. *Res. Sq.* **2021**.
  60. Krystal, J.H. Responding to the hidden pandemic for healthcare workers: Stress. *Nat. Med.* **2020**, *26*, 639. [[CrossRef](#)] [[PubMed](#)]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.