



# Article Early Stage Identification of COVID-19 Patients in Mexico Using Machine Learning: A Case Study for the Tijuana General Hospital

Cristián Castillo-Olea <sup>1,\*</sup>, Roberto Conte-Galván <sup>1</sup>, Clemente Zuñiga <sup>2</sup>, Alexandra Siono <sup>3</sup>, Angelica Huerta <sup>4</sup>, Ornela Bardhi <sup>5</sup> and Eric Ortiz <sup>6</sup>

- <sup>1</sup> Ensenada Center for Scientifc Research and Higher Education, Ensenada 22860, Mexico; conte@cicese.mx
- <sup>2</sup> Tijuana General Hospital, Tijuana 22000, Mexico; drclementezuniga@gmail.com
- <sup>3</sup> Faculty of Engineering, CETYS University, Mexicali 21259, Mexico; alexandra.siono@cetys.edu.mx
- <sup>4</sup> Faculty of Medicine and Psychology, Autonomous University of Baja California, Mexicali 21100, Mexico; angelica.huerta.d@gmail.com
- <sup>5</sup> Independent Researcher, 1001 Tirana, Albania; alenroidhrab@gmail.com
- <sup>6</sup> comeMed Teleconsulting, Colonia Roma, Mexico City 6700, Mexico; dr.ericortiz.oncomed@gmail.com
- Correspondence: cristian.castillo2@gmail.com; Tel.: +52-5574302237



Keywords: machine learning; COVID-19; identification

# 1. Introduction

A novel coronavirus, known as Severe Acute Respiratory Syndrome (SARS-CoV-2), was identified in December 2019 as the cause of a respiratory illness called Coronavirus Disease 2019, or COVID-19 [1]. The origin of this virus is not yet confirmed, but an analysis of its genetic sequence suggests it is phylogenetically related to bat viruses similar to SARS (severe acute respiratory syndrome), making bats a possible key reservoir [2]. Symptoms of COVID-19 infection appear after an incubation period of approximately 5.2 days [3]. The period from the onset of COVID-19 symptoms to death ranges from 6 to 41 days with a median of 14 days [4]. This period depends largely on the age and the state of the patient's immune system [4].



Citation: Castillo-Olea, C.; Conte-Galván, R.; Zuñiga, C.; Siono, A.; Huerta, A.; Bardhi, O.; Ortiz, E. Early Stage Identification of COVID-19 Patients in Mexico Using Machine Learning: A Case Study for the Tijuana General Hospital. *Information* **2021**, *12*, 490. https:// doi.org/10.3390/info12120490

Academic Editors: Sidong Liu, Cristián Castillo Olea and Shlomo Berkovsky

Received: 13 October 2021 Accepted: 22 November 2021 Published: 24 November 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 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/). The infection is transmitted through droplets generated by symptomatic patients when coughing or sneezing, but it can also occur through asymptomatic patients and even before the onset of symptoms [5].

The clinical features of COVID-19 are diverse, from an asymptomatic state to acute respiratory distress syndrome and multiorgan failure [5]. The most common early symptoms of COVID-19 illness are fever, cough, and fatigue, while other symptoms include headache, sputum production, hemoptysis, dyspnea, diarrhea, and lymphopenia [6]. Advanced age, cardiovascular disease, diabetes, chronic respiratory disease, hypertension, and cancer are said to increase the risk of death for people diagnosed with COVID-19.

Regarding COVID-19, as of 15 August 2021, there were 207,784,507 confirmed cases (410,464 new cases) and 4,370,424 deaths, while 4,462,336,040 vaccine doses have been reported worldwide [7,8]. Most estimates of fatality ratios have been based on cases detected through surveillance and calculated using crude methods, giving rise to widely variable estimates of CFR depending on the country—from less than 0.1% to over 25% [9].

Currently, in Mexico (August 2021), there are 3,310,989 estimated positives, with 261,384 estimated deaths, and 133,866 estimated actives. However, there are 3,108,438 confirmed cases, 5,527,343 negative, 477,811 suspected and 248,652 accumulated deaths. Of the confirmed cases, 53.56% have been women and 46.44% men. A total 5.6% of patients have been hospitalized, and 94.4% have been outpatients. Among the main comorbidities are hypertension (10.34%), Obesity (8.989%), diabetes (7.31%) and smoking (8.03%), with information updated on 16 August 2021. On the same date, the state of Baja California had 913 active cases, with 54,453 accumulated cases, and 8979 deaths. The state's capital, Mexicali, is the city with the highest number of cases in the state, with 21,778 accumulated cases, followed by Tijuana, our case study [10].

# 2. Background

Several authors have addressed the issue of SARS-CoV-2 from a technological point of view, with the development of artificial intelligence algorithms. There are models to predict the mortality rate [11]. Some studies present the detection of severely ill patients with COVID-19 from those with mild symptoms using clinical information and data from blood and urine tests [12]. Several artificial intelligence models have been used with Machine Learning and Deep Learning methods which have been used intensively for COVID-19. Although Machine Learning and Deep Learning methods show successful results in the COVID-19 cases tested, there are accounting challenges that can be considered to improve the quality of the research in that direction [13].

Table 1 summarizes the articles included in this minireview, together with some characteristics of these studies.

Authors.	Year	Objective	Learners	Metrics	Novelties
Li WT et al. [14].	2020	Classification	XGBoost	sensitivity of 92.5% and a specificity of 97.9%	Novel associations between clinical variables, including correlations between being male and having higher levels of serum lymphocytes and neutrophils. We found that COVID-19 patients could be clustered into subtypes based on serum levels of immune cells, gender, and reported symptoms.
Guan X et al. [15]	2020	Prediction	XGBoost	>90% precision and >85% sensitivity, as well as F1 scores >0.90	Proposed disease severity, age, serum levels of hs-CRP, LDH, ferritin, and IL-10 as significant predictors for death risk of COVID-19, which may help to identify the high-risk COVID-19 cases.

Table 1. Minireview of papers.

# Table 1. Cont.

Authors.	Year	Objective	Learners	Metrics	Novelties
Delafiori J et al. [16]	2021	Diagnosis and risk assesment	gradient tree boosting (GDB) ADA tree boosting	96.0% of specificity and 83.1% of sensitivity 80.3% of specificity and 85.4% of sensitivity	Propose machine learning techniques to determine from databases the five main challenges in responding COVID-19 and how to overcome these challenges to save lives.
Allam M et al. [17]	2020	Diagnosis and prediction	Neural Networks	100% sensitivity and 99.9% specificity	The Abbott antibody test (SARS-CoV-2 IgG assay) has shown 100% sensitivity and 99.9% specificity thus far. The Abbott test finds whether the patient has IgG antibodies for COVID-19, which can stay for months to years after a person has recovered.
Assaf D et al. [18]	2020	Prediction	Classification and Regression Tree (CRT) model	Sensitivity, specificity, PPV, NPV and accuracy of 88.0%, 92.7%, 68.8%, 97.7% and 92.0%, respectively, with ROC AUC of 0.90.	The analysis of the database in this study found that most contributory variables to the models were APACHE II score, white blood cell count, time from symptoms to admission, oxygen saturation and blood lymphocytes count. Machine-learning models demonstrated high efficacy in predicting critical COVID-19 compared to the most efficacious tools available.
Naseem M et al. [19]	2020	Detection	Neural Networks	sensitivity of 90% and specificity of 96% respectively	Results were synthesized and reported under 4 themes. (a) The need of AI during this pandemic: AI can assist to increase the speed and accuracy of identification of cases and through data mining to deal with the health crisis efficiently, (b) Utility of AI in COVID-19 screening, contact tracing, and diagnosis: Efficacy for virus detection can a be increased by deploying the smart city data network using terminal tracking system along-with prediction of future outbreaks, (c) Use of AI in COVID-19 patient monitoring and drug development:
Arga KY [20]	2020	Prediction and diagnosis	Apache, Gleason and PASI	-	Machine learning is considered to help reduce diagnostic errors and unnecessary use of diagnostic tools through the development of rational algorithms. Indeed, the COVID-19 pandemic showed that digital health is invaluable, feasible, and not too far.
Majhi R et al. [21]	2020	Prediction	Nonlinear Regression (NLR), Decision Tree (DT) based regression, and random forest (RF) models	Evaluation metrics obtained using the Mean Absolute Percentage Error (MAPE). NLR = 0.24% DT = 0.18% RF = 0.02%	The algorithm predict the number of positive cases in India. In essence, the paper proposes a machine learning model that can predict the number of cases well in advance very effectively and also suggest some key inputs.

Authors.	Year	Objective	Learners	Metrics	Novelties
van der Schaar M et al. [22]	2020	Prediction	-		This paper summarizes the use of machine learning techniques in different studies to manage limited healthcare results, developing personalized treatment, informing policies and able effective collaboration and expediting clinical trials.
Das AK et al. [23]	2020	Prediction	Linear Regression	For liner regression (area under ROC curve = 0.830), calibration (Matthews Correlation Coefficient = 0.433; Brier Score = 0.036).	In this study, according to the random forest algorithm, age was the most important predictor followed by exposure, sex and province, whereas this order was sex, age, province and exposure as per logistic regression
Swapnarekha H et al. [24]	2020	Prognosis	Support Vector Machine (SVM), Rannom Forest RF, K-Means, XGBoost and linear regression	0.933 true positive rate, 0.74 true negative rate and 0.875 accuracy.	This article obtained good metrics for COVID-19 prediction. On the other hand, mentioned machined learning techniques used for classification and prediction to reduce the spread of coronavirus and understand the limitation of machine learning analysis, being: lack of information, accuracy of predictions, usage of advanced approaches, providing feasible solutions for developing countries and necessity of advance intelligent systems on symptom based identification of COVID-19.

#### Table 1. Cont.

#### 3. Materials and Methods

This article is based on a study of COVID-19 patients at the Tijuana General Hospital, a public hospital that serves a very particular low-income population. Tijuana is a border city in northern Mexico next to San Diego, California, in the United States. Including neighboring Rosarito, the greater Tijuana region has a population of around 1,900,000 inhabitants, the majority Mexican nationals. However, there are also many migrants from other Central American, South American, and Caribbean countries living in unaccounted shantytowns, seeking to enter the United States in any way, either legal or illegally. While doing so, they temporarily live in Tijuana without a permanent job, a fixed salary, a steady place of work, or a regular postal address, and therefore, do not have access to social security or health services, and eventually fall ill, often due to a myriad of causes. The range of pathologies and diseases found in the city of Tijuana is much broader than those found in other cities with more homogeneous or steady populations, which further complicates medical and healthcare services for this specific segment of people. Tijuana General Hospital is one of the few public health institutions that serve this marginalized segment of the Tijuana population, hence the size and complexity of the challenges faced by its medical staff every day, as well as the diversity of pathologies met by the health professionals who treat them. In this article, we evaluate a group of COVID-19-diagnosed patients, who were treated at this Tijuana hospital during 2020.

# 3.1. Sample Size

The required sample size for this study was estimated considering the expected prevalence in studies carried out using bioimpedance analysis, 17% [25], assuming a

margin of error of 5% and a confidence interval of 95%. According to these criteria, a total of 185 patients were needed in order to obtain the desired results. The average age of the studied population was 55 years, while the average hospital length of stay duration was six days.

Patients arrived mainly from the Tijuana and Rosarito urban and suburban areas. This research included patient medical history, pharmacology, PCR testing and biochemical data.

#### 3.2. Database

Information on 185 patients with 99 variables was collected for each of them to create the database, the description of the variables can be found in Appendix A. Table 1 shows the gender criteria for evaluating patients at the Tijuana General Hospital, which serves a low-income population in the Tijuana and Rosarito areas of Baja California, with a higher percentage of men than women as seen from the Table 2.

Table 2. Evaluation Criteria.

Gender	%	Kg-m <sup>2</sup>	SpO
Women	39.46%	$<6.1 \text{ kg/m}^2$	>95%
Ivien	60.54%	<8.5 kg/m-	>93%

#### 3.3. Bedford's Law

Benford's Law validation method was used in order to make sure the data was consistent, in order to develop an efficient study. Benford's Law, or the Law of First Digits, is a tool used in different fields of science, with a method to suggest a mathematical pattern in the distribution of the first digits in a dataset that does not display a uniform distribution, but rather are arranged in such a way that the digit "1" is the most frequent, followed by "2", then "3", and so on, down to "9". This model suggests that, within a random set of data, the first digit of approximately 30.10% of the numbers will be "1". Several studies have used this technique to validate and evaluate veracity in databases with information about COVID-19 [26,27].

By using Benford's Law as a validation method, it was demonstrated that there is consistency in the data collected from the Tijuana General Hospital, as the curve of our current information is close to the curve generated by the percentages established by Benford's Law. For this comparison, "length of stay" data were used, as within the database that was used, most information is described by binary numbers (0 and 1), while the variable "length of stay" is a defined variable. For this, in addition to the graph function, two Excel functions were used:

*Left*: This function allowed the first digit to be taken to the left of the number within the "length of stay" column.

*Countif*: This function allows to count the frequency of each of the digit numbers without considering 0.

Figure 1 shows the comparison results between Bedford's Law and "length of stay" which suggest that the data is consistent.



Figure 1. Benford's Law curve comparison with the curve generated by the actual data from "length of stay".

The distribution by first digit numbers in the data can be seen in Table 3. Invalidating the data, the comparison suggests that according to Benford's law, the results obtained from this analysis for the Tijuana Hospital database are accurate.

Benford's Law	Actual	<b>R.E.</b> *	
	Tijuana cases <sup>1</sup>		
30.10%	25.57%	-15.01%	
17.61%	14.20%	-19.30%	
12.49%	14.00%	9.12%	
9.69%	12.50%	22.89%	
7.92%	10.23%	29.16%	
6.69%	7.95%	18.83%	
5.80%	6.25%	7.75%	
5.12%	6.82%	33.20%	
4.58%	2.84%	1.73%	

Table 3. Data compared with Bedford's Law.

<sup>1</sup> Considering "length of stay" for validation. \* Relative Error.

# 3.4. Machine Learning Analysis

For the evaluation of each dataset, the following classifiers were used: Decision Tree (DT), Support Vector Machine (SVM), Random Forest (RF), Multi-Layer Perceptron Neural Network (MLPNN), Naive Bayes (NB), Logistic Regression (LR) and AdaBoost (AB). The configuration for each algorithm were suggested by Orange. These configurations are shown in Table 4. The algorithms were evaluated by stratified k-fold cross-validation, where the data was iteratively examined by ten folds, using nine folds for training and 1-fold for testing; with this method, the data can be divided into equal parts, which gives better results, avoiding generalization which means possible errors when new data is used for predictions with the trained models.

Learner	Configuration for Learners
RF	A number of trees: 10, minimum subsets split: 5, maximum tree depth: unlimited.
kNN	Number of neighbors: 3, metric: Euclidean, weight: uniform
SVM	Type: SVM Regression, C = 1, $\varepsilon$ = 0.1, Kernel= Radial Basis Function (RBF), exp (-auto   x - y   <sup>2</sup> ), numerical tolerance: 0.001
MLPNN	Hidden layers: 100, activation: ReLu, solver: Adam, alpha: 0.0001, maximum iterations: 200, replicable training: True.
NB	
AB	Base estimator: tree, number of estimators: 50, algorithm (classification): Samme. r, loss (regression): Linear
DT	Type: binary tree, internal nodes < 5, maximum depth: 100, splitting: 95%.

Table 4. Configurations used for each learner in the analysis.

The datasets for each analysis were created by feature selection; for this, HAS, DM, and Obesity were selected as targets due to their importance in the development of COVID-19. After selecting the targets, the complete database was analyzed iteratively with each of the chosen targets. The process of selecting the most relevant features was done by using the DT and comparing the results with the rank, using the following scoring methods: gain ratio, Gini, X<sup>2</sup>, ReliefF and Fast Correlation Based Filter (FCBF). The algorithm followed for the scoring techniques is shown in Table 5.

Table 5. Rank Scoring Algorithms.

Method	Algorithm
Gain Ratio	$IG(Ex,a) = H(Ex) - \sum_{v \in values (a)} \left( \frac{ \{x \in Ex   value(x,a) = v\} }{ Ex } * H(\{x \in Ex   value(x,a) = v\}) \right)$
Gini	$G = \frac{\sum_{i=1}^{n} * \sum_{j=1}^{n} *  x_i - x_j }{2\sum_{i=1}^{n} * \sum_{i=1}^{n} * x_i}$
X <sup>2</sup>	$X = \frac{m - Np}{\sqrt{Npq}}$
ReliefF	$W_i = W_i - (x_i - nearHit_i)^2 + (x_i - nearMiss_i)^2$
FCBF	$H(X) = -\sum_{i} P(x_i) \log_2(P(x_i))$

# 4. Results

Given that the comorbidities prevalent in patients of the analyzed database are SAH with 49% of patients, DM with 34% of patients, and Obesity with 11% of patients, it was decided to take these three diseases independently as a target to find the variables that were most related to these diseases present in patients with COVID-19, determining the factors involved where the information had to be pre-processed. The Decision Tree and ranking methods were used to determine the variables with the most significant impact. Tables 6–8 show the resulting datasets after the analysis. Dataset 1 in each table was determined by using all of the existing variables in the database; then, for dataset 2, the selection of features was determined by choosing the ones closer to the root within the first ten levels; finally, to assess dataset 3, ranking results were considered, regardless of the variables "Tos" (cough), "Fiebre" (fever), "Disnea" (dyspnea), and "Dolor de cabeza/Cefalea (headache)" as these were established by the World Health Organization as official COVID-19 symptoms, and by using them for analysis they performed better obtaining an increase in the dataset scores, therefore by dismissing them the research found other significant variables.

Dataset	Target HAS
Dataset 1	Edad, género, grupo etario, DM, ECV, Hepáticas, SNC, Neumopatía, Cáncer, Inmunosupresión, Obesidad, Otros (1), Comorbilidades, Fiebre, Mialgias/artralgias, Fatiga, Odinogafia/ardor faringeo, Tos, Disnea, Dolor Toracico, Congestión Nasal, Rinorrea, Expectoración Diarrea, Náusea, Anorexia, Vómito, Cefalea, Mareo, Hispomia/Anosmia, Ageusia, Conjuntivitis, Saturación >90, Saturación 80–90%, Saturación < 80%, Leucopenia, Leucocitosis, Neutropenia, Neutrofilia, Linfopenia, Linfocitosis, Eosinopenia, Trombocitopenia, Neutrofilia, Linfopenia, Linfocitosis, Eosinopenia, Trombocitopenia, TrPa alargado, Creatinina normal, Creatinina alta, Ferritina normal, Ferritina alta, Dímero D normal, Dímero D Alto, Fibrinógeno normal, Fibrinógeno Alto, PCR normal, PCR alta, Procalcitonina normal, Procalcitonina alta, Troponina normal, Troponina alta, CPK normal, CPK alta, CK-MB normal, CK-MB alta, Albúmina baja, Albúmina normal, Bilirrubina total normal, Bilirrubina total alta, ALT/TGP normal, ALT/TGP alta, AST/TGO normal, AST/TGO alta, DHL normal, DHL alta, DHL > 1000, Fosfatasa alcalina normal, Fosfatasa alcalina alta, Muestra, POSITIVA, Mayor 50%, Moderado, Grave, Oseltamivir, Ceftriaxona, Claritromicina, Azitromicina, Levofloxacino, Otros (2), Hidroxicloroquina/Cloroquina, Tocilizumab, Esteroides, Pronación, Respondedor, Respondedor parcial, No respondedor, Alta por mejoría, Defunción, Días de estancia hospitalaria.
Dataset 2	Comorbilidades, Edad, Tos, CK-MB Normal, INR Alto, DM, Cefalea, Neutrofilia, Dímero, Leucocitos, Neumopatía, Obesidad, Días de estancia, CPK alta, Saturación 90, Eosinopenia, TP Alargado y Odinofagia
Dataset 3	Comorbilidades, Edad, CK-MB Normal, CK-MB Alto, DM, Neutrofilia, Dímero D Alto, Leucocitosis, Neumopatía, Obesidad, Dias de estancia, Otros (1)

**Table 6.** Datasets for HAS as a target (variable names in Spanish).

 Table 7. Datasets for DM as a target (variable names in Spanish).

Tabl	le	7.	Cont.

Dataset	Target DM
Dataset 2	Edad, Neutropenia, Comorbilidades, Cáncer, Claritromicina, HAS, linfocitosis, Ferritina normal, Hepáticas, SNC, Leucopenia, Inmunosupresión, eosinofilia, ferritina alta, Troponina normal, vómito, INR alto, CM-KB alta, Disnea, TTP alargado, Levofloxacino, Fatiga, AST/TGO alta, bilirrubina total alta, fiebre, creatinina alta, INR normal, Diarrea, Augesia.
Dataset 3	Edad, Género, HAS, Obesidad, Otros (1), Comorbilidades, Leucocitosis, Creatinina normal, Creatinina alta, Procalcitonina alta, Levofloxacino, Hidroxicloroquina

Table 8. Datasets for Obesity as a target (variable names in Spanish).

Dataset	Target Obesity
Dataset 1	Edad, género, grupo etario, HAS, DM, ECV, Hepáticas, SNC, Neumopatía, Cáncer, Inmunosupresión, Otros (1), Comorbilidades, Fiebre, Mialgias/artralgias, Fatiga, Odinogafia/ardor faringeo, Tos, Disnea, Dolor Toracico, Congestión Nasal, Rinorrea, Expectoración Diarrea, Náusea, Anorexia, Vómito, Cefalea, Mareo, Hispomia/Anosmia, Ageusia, Conjuntivitis, Saturación > 90, Saturación 80–90%, Saturación < 80%, Leucopenia, Leucocitosis, Neutropenia, Neutrofilia, Linfopenia, Linfocitosis, Eosinopenia, Trombocitopenia, Trombocitosis, TP normal, TP alargado, INR normal, INR Alto, TTPa normal, TTPa alargado, Creatinina normal, Creatinina alta, Ferritina normal, Ferritina alta, Dimero D normal, Dimero D Alto, Fibrinogeno normal, Fibrinogeno Alto, PCR normal, PCR alta, Procalcitonina normal, Procalcitonina alta, Troponina normal, Troponina alta, CPK normal, CPK alta, CK-MB normal, CK-MB alta, Albumina baja, Albumina normal, Bilirrubina total normal, Bilirrubina total alta, ALT/TGP normal, ALT/TGP alta, AST/TGO normal, AST/TGO alta, DHL normal, DHL alta, DHL > 1000, Fosfatasa alcalina normal, Fosfatasa alcalina alta, Muestra, POSITIVA, Mayor 50%, Moderado, Grave, Oseltamivir, Ceftriaxona, Claritromicina, Azitromicina, Levofloxacino, Otros (2), Hidroxicloroquina/Cloroquina, Tocilizumab, Esteroides, Pronación, Respondedor, Respondedor parcial, No respondedor, Alta por
Dataset 2	Saturación 80–90, Edad, Cefalea, Género, Levoflaxina, GRAVE, Hisponia/asmonia, Linfopeni, a Neumopatía, Fosfatasa alcalina normal, Creatinina normal, días de estancia, PCR alta
Dataset 3	Saturación 80–90, Edad, Género, Levoflaxina, GRAVE, Hisponia/asmonia, Linfopenia Neumopatía, Fosfatasa alcalina normal

After the analysis of each dataset, the obtained results are shown in Tables 9–11; where the best scores are shown in bold font for each method.

Model	AUC	CA	F1	Precision	Recall
Tree Decision	0.814	0.784	0.784	0.784	0.784
SVM	0.867	0.762	0.762	0.762	0.762
Random Forest	0.866	0.784	0.784	0.784	0.784
Neural Network	0.876	0.773	0.773	0.773	0.773
Naive Bayes	0.832	0.757	0.755	0.761	0.757
Logistic Regression	0.910	0.800	0.800	0.801	0.800
AdaBoost	0.860	0.811	0.811	0.811	0.811

Table 9. Highest scores from a dataset with HAS as a target.

Table 10. Highest scores from a dataset with DM as a target.

Model	AUC	CA	F1	Precision	Recall
Tree	0.867	0.881	0.878	0.882	0.881
SVM	0.934	0.886	0.885	0.886	0.886
Random Forest	0.877	0.849	0.847	0.847	0.849
Neural Network	0.871	0.816	0.812	0.813	0.816
Naive Bayes	0.849	0.838	0.838	0.838	0.838
Logistic Regression	0.912	0.892	0.888	0.897	0.892
AdaBoost	0.827	0.843	0.844	0.844	0.843

Best model: Logistic Regression with 0.934 of AUC, 0.886 of CA, 0.885 of F1, 0.886 of Precision, and 0.886 of Recall.

Table 11. Highest scores from a dataset with Obesity as a target.

Model	AUC	CA	F1	Precision	Recall
Tree	0.700	0.876	0.869	0.864	0.876
SVM	0.643	0.838	0.831	0.824	0.838
Random Forest	0.807	0.903	0.872	0.912	0.903
Neural Network	0.834	0.914	0.899	0.906	0.914
Naive Bayes	0.793	0.865	0.845	0.833	0.865
Logistic Regression	0.778	0.881	0.856	0.851	0.881
AdaBoost	0.697	0.881	0.879	0.876	0.881

Best model: Neural Network with 0.834 of AUC, 0.914 of CA, 0.899 of F1, 0.906 of Precision and 0.914 of Recall.

As shown in Table 8, the highest score using as target HAS was obtained with the Logistic Regression algorithm, using the default parameters with a cost strength of 1.00 (C-1) and Ridge-type regularization (L2), which shows 91.0% Area under ROC (AUC), 80% Classification Accuracy (CA), 80% F1, and 80.1%, Precision and Recall for selected variables in Dataset 3 in Table 6 as possible variables essential to consider for a more accurate determination of vulnerability in people with HAS.

On the other hand, when targeting DM patients, the best-functioning model using the default parameters already mentioned above was found to be Logistic Regression, obtaining 91.2% AUC, 89.2% CA, 88.8% F1, 89.7% precision, and recall 89.2% for selected variables from dataset 3 in Table 7.

While the Neural Network algorithm showed better results for patients with Obesity, obtaining 83.4% AUC, 91.4% CA, 89.9% F1, 90.6% Accuracy, and 91.4% Recall (See Table 11).

The prevalent symptoms present in people with different comorbidities, and statistical analysis was performed without considering cough, fever, dyspnea, and headache as part of the symptoms for the same reason mentioned above. In people with HAS, DM, and Obesity, the presence of fatigue and myalgias/arthralgias was greater; while the third dominant symptom in people with HAS and DM was odynophagia, instead of people with Obesity, this symptom was positioned with the eighth, occupied the third position with chest pain (See Figures 2–4).



Quantity of patients

Figure 2. Frequency of symptoms in patients with DM.



Quantity of patients

Figure 3. Frequency of symptoms in patients with HAS.



Figure 4. Frequency of symptoms in patients with Obesity.

# 5. Discussion

Obesity, diabetes, and hypertension are high-prevalence comorbidities in the Mexican population. In 2018, approximately 1/4 of the Mexican people had high blood pressure, while 71.2% of people over the age of 20 had a prevalence of obesity and overweight. These two conditions are considered to be significant risk factors for developing other diseases such as diabetes since 90% of people with diabetes in Mexico are overweight and obese [28].

In processing the information in this study, the prevalence of these three diseases considered as risk factors for COVID-19 infection could be noted. For this reason, it was decided to carry out an analysis using Artificial Intelligence (Machine Learning) techniques, aiming for each of these diseases, looking for any relationship between the symptoms and conditions of people suffering from these comorbidities.

After reviewing the medical literature, it was found that COVID-19 causes chronic inflammation in patients with obesity, along with other diseases considered as risk factors, such as lack of vitamin D and intestinal dysbiosis, which result in deficiencies in the functioning of the immune system in the face of infections. On the other hand, Obesity has a negative impact on respiratory mechanics as it is affected due to the resistance generated by lack of elasticity in the chest box [29,30]. Both diabetes and high blood pressure are diseases more present in the elderly, one of the reasons why these diseases are thought to be closely related to COVID-19, due to the fact that people with advanced age are highly vulnerable to COVID-19. The reason diabetes could have a high impact on the condition of patients with COVID-19 may be due to a disruption generated in the endocrine system where the COVID-19 virus affects angiotensin-converting Enzyme 2 (ECA2), which is responsible for anti-inflammatory regulation, vasodilators, and the process of releasing sodium into the urine. Another function of ECA2 is to offer protection to different organs, including those that are part of the pulmonary and cardiovascular systems, which is why the absence of this enzyme is related to the involvement of the lungs and the development of hypertension [31,32].

Different authors mentioned in a report of the analysis carried out with a database of 3894 patients in Italy [33], obtaining values for a death risk target with an accuracy rating of 83.4%, F1 value of 90.4%, specificity of 30.8%, and Recall of 95.2% when using Random Forest over a dataset with variables such as Glomerular Filtration Rate (eGFR), C- Reactive Protein (CRP), Age, Diabetes, Sex, Hypertension, Smoking, Lung Disease, Myocardial Infection, Obesity, Heart Failure, and Cancer, demonstrating this as vulnerable to those who exhibit the above-mentioned comorbidities. To achieve a comparison between the variables established with the results of this analysis, the scores of the three targets used

(HAS, Diabetes, and Obesity) were taken into account and the average was calculated, obtaining values of 84.50% for CA, 83.43% of F1, 84.76% Precision, and 84.53% of Recall.

On the other hand, a study of Brazilian patients shows their scores at 86% for CA, 92% for AUC, 28% of Precision, 86% of Recall, and 42% of F1 for the Logistic Regression model, where the results show that there is a high relationship between patients over 60 years of age with breathing difficulties, fever, cough, rhinorrhea, odynophagia, diarrhea, headache, heart disease, pneumopathies, kidney disease, diabetes, smoking, Obesity, and hospitalization with those who come down with COVID-19 [34].

This analysis provides a complete and accurate perspective of the current situation in Tijuana since having information from a single source allows us to know the current situation of those served at the Tijuana General Hospital, who are mostly part of low-income or economically disadvantaged family groups. Since 41.9% of the Mexican population is economically disadvantaged [35], it is important to understand the comorbidities and symptoms present in this group, which results in more vulnerable people and increases expenditures in public sector hospitals regarding space and medical equipment, which often is extremely limited. Since the study was done considering typical human patients, regardless of their geographical and ethnic origin, a similar genomic situation can be assumed with migrant patients from both parts of Mexico and regions such as Central America, South America, and the Caribbean.

# 6. Conclusions

The values of the prevalent comorbidity found in this study were as follows: 49% in patients with HAS; 34% in patients with DM; and 11% in patients with Obesity, corresponding to the population segment that was served at the General Hospital of Tijuana during the aforementioned dates.

The deaths reported during the evaluated period of the total number of patients evaluated in this study were 42 subjects, from a population universe of 185 patients evaluated, with a fatality rate of 22%, a non-representative sample but within the data that places Mexico with a fatality rate (Case fatality) of 8.8% compared to other fatality rates in Latin America, such as Peru (3.5%), Colombia (2.6%), Chile (2.5%), Brazil (2.4%), and 2–3% globally [36].

A total of 52 different medicines were prescribed, where Steroids (16.1%), azithromycin (10.9%), enoxaparin (8.2%), levofloxacin (7.3%), hydroxyquinone (6.6%), omeprazole (6.2%), and acetaminophen (6.2%) were most commonly used. The rest, presenting themselves with a <6.0% per drug, correspond as a whole to 38.5% and are classified as "other".

The values obtained for each dataset evaluating Obesity, DM, and HAS in this study using Artificial Intelligence, specifically with various Machine Learning techniques, have been compared with values presented in other similar academic/medical publications, particularly those for HAS were obtained with the Logistic Regression algorithm obtaining 91.0% AUC, CA, F1 and Recall 80%, and accuracy 80.1%. For Diabetes, the Logistic Regression algorithm obtained 91.2% AUC, 89.2% CA, 88.8% F1, 89.7% accuracy, and recall 89.2%, obtaining similar results for Hypertension and Diabetes, while improved results were obtained in the case of Obesity using the Neural Network algorithm, obtaining 83.4% AUC, 91.4% CA, 89.9% F1, 90.6% Accuracy and 91.4% Recall.

The results presented here confirm the convenience of using Logistic Regression algorithms in the dataset presented when evaluating HAS as a target; Similarly, Logistic Regression's algorithms were the most successful in evaluating DM as a target; while finally, Neural Network algorithms showed the best results for the case of Obesity as a target, in the specific sample case of 185 patients with limited or non-existing financial resources, who suffered these medical conditions, and who were served at the Tijuana General Hospital during the year 2020.

With these results, we can conclude that the use of Artificial Intelligence using Machine Learning techniques has been effectively used to identify the early stages of COVID-19 in patients in Baja California.

Author Contributions: Conceptualization, C.C.-O., C.Z., R.C.-G., E.O.; Data curation, C.C.-O., A.S.; Methodology, C.C.-O., A.S., R.C.-G., A.H.; Project administration, C.C.-O., R.C.-G., E.O., C.Z.; Supervision, R.C.-G., E.O., O.B., C.Z.; Validation, C.C.-O., O.B., E.O., C.Z. and R.C.-G.; Writing—original draft, C.C.-O., A.S., R.C.-G., O.B.; Writing—review & editing, C.C.-O., O.B., C.Z. and E.O. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by a postdoctoral fellowship from the Consejo Nacional de Ciencia y Tecnología (National Council of Science and Technology).

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by the Ethics Committee of Hospital General de Tijuana protocol code 171101-20 on 12 April 2020.

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

Data Availability Statement: Not applicable.

**Acknowledgments:** The authors would especially like to express their gratitude to the Mexican National Council of Science and Technology (CONACYT) for the Postdoctoral scholarship 334844.

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

# Appendix A

Table A1. Description of variables.

Variable	Description
Edad	Age
Género	Gender (Female/Male)
Grupo etario	Grupo etario
HAS	Defined by two levels according to the 2017 American College of Cardiology/American Heart Association (ACC/AHA) guidelines: (1) Elevated blood pressure, with a systolic pressure (SBP) between 120—129 mm Hg and diastolic pressure (DBP) less than 80 mm Hg, and (2) stage 1 hypertension, with a SBP of 130 to 139 mm Hg or a DBP of 80 to 89 mm Hg
DM	Diagnosis by meeting any of the criteria: Fasting glucose ≥126 mg/dL (7.0 mmol/L). Fasting is defined as the absence of caloric intake for at least 8 h OR 2 h postprandial glucose ≥200 mg/dL (11.1 mmol/L). The test should be performed as described by the WHO, using a glucose load containing the equivalent of 75 g of anhydrous glucose dissolved in water. OR Glycated hemoglobin ≥6.5% (48 mmol/mol). The test must be performed in a laboratory using a method that is certified by NGSP and standardized for the DCCT assay. OR In a patient with classic symptoms of hyperglycemia or hyperglycemic crisis, a random plasma glucose ≥ 200 mg/dL (11.1 mmol/L).
ECV	Neurological alteration is characterized by its sudden onset, generally without warning, with symptoms lasting 24 h or more, causing sequelae and death.
Hepaticas	Primary or secondary diseases that affect liver tissue.
SNC	Central Nervous System.

Table	A1.	Cont.	
lable	A1.	Com.	

Variable	Description
Neumopatía	Lung disease is a generic term to describe diseases that affect the lungs. It should not be confused with the term pneumonia, which specifically refers to infection of the lung by a virus or bacteria.
Cancer	Any of a large number of diseases characterized by the development of abnormal cells that divide uncontrollably and have the ability to infiltrate and destroy normal body tissue.
Inmunosupresión	Suppression or reduction of immune reactions. It may be due to the deliberate administration of immunosuppressive drugs used in the treatment of autoimmune diseases or in recipients of transplanted organs to avoid rejection. It can also be secondary to pathological processes such as immunodeficiencies, tumors, or malnutrition.
Obesidad	A condition characterized by the excessive accumulation and storage of fat in the body and which in an adult is typically indicated by a body mass index of 30 or more.
Otros	Other types are of diseases not classifiable in the previous variables.
Comorbilidades	A concomitant but unrelated disease process or disease; is commonly used in epidemiology to indicate the coexistence of two or more disease processes.
Otros Especificar	Other types are of diseases not classifiable in the previous variables, where the type is specified.
Fiber	Temperature above the normal range due to an increase in the body temperature set point. There is no agreed upper limit for normal temperature with sources using values between 37.2 and 38.3 °C (99.0 and 100.9 °F) in humans.
Mialgias/arthralgias	Muscle or joint pain.
Fatigue	Difficulty starting or maintaining physical or mental activity voluntarily.
Odinofagia/ardor faringeo	Feeling of pain when swallowing.
Tos	Sudden and acute expulsion of air from the lungs that acts as a protective mechanism to clear the airways or as a symptom of a pulmonary disorder.
Disnea	Difficulty breathing.
Dolor toracico	Localized chest pain, regardless of its etiology.
Congestión nasal	A feeling of blockage or obstruction in the nasal cavity and/or sinuses due to inflammation of the mucous lining of the nose.
Rinorrea	Flow or abundant emission of fluid from the nose.
Expectoración	Expulsion through coughing or sputum or other secretions formed in the respiratory tract.
Diarrhea	It consists of the expulsion of three or more liquid stools, with or without blood, in 24 h, which adopt the shape of the container that contains them.
Nausea	Feeling sick or sick in the stomach that may appear with an urgent need to vomit.
Anorexia	It is used to denote lack of appetite or lack of appetite that can occur in very different circumstances.
Vómito	Violent expulsion through the mouth of what is contained in the stomach.

Variable	Description
Cefalea	They are painful and disabling primary disorders such as migraine, tension headache, and cluster headache.
Mareo	Feeling of vertigo and instability in the head and discomfort in the stomach that can lead to the urge to vomit and loss of balance.
Hyposmia/Anosmia	Decreased or absent sense of smell.
Ageusia	Decreased or absent sense of taste.
Conjunctivitis	Inflammation or irritation of the conjunctiva.
Saturación >90	Oxygen saturation in ambient air >90%.
Saturación 80–90%	Oxygen saturation in ambient air of 80–90%.
Saturación <80%	Oxygen saturation in ambient air <80%.
Leucopenia	Reduction in circulating white blood cell count <4000/mcL.
Leukocytosis	A white blood cell count greater than 11,000/mm <sup>3</sup> ,
Neutropenia	When the neutrophil numbers are below 1500–1800 per mm3.
Neutrophilia	Neutrophil blood values equal to or less than 7700/microL.
Linfopenia	Total lymphocyte count <1000/mcL.
Linfocitosis	When the lymphocyte count is greater than 4000 per microliter.
Eosinopenia	Reduction in circulating eosinophils $<0.01 \times 10^9$ /L.
Eosinophilia	A count of more than 500 eosinophils per microliter of blood.
Thrombocytopenia	Decrease in the absolute number of platelets in the peripheral blood below 150,000 per $\mu$ L.
Trombocitosis	Platelet count greater than 600,000 per $\mu$ L.
TP normal	TP in blood with a range of 11 to 13.5 s.
TP alargado	TP in blood >13.5 s.
INR normal	INR with a value between 0.9 to 1.3.
INR Alto	INR with a value >1.3.
TTP normal	APTT in blood in a range of 25 to 35 s.
TTP alrgado	APTT in blood >35 s.
Creatinine normal	The normal range for creatinine is 0.7 to 1.3 mg/dL (61.9 to 114.9 μmol/L) for men and 0.6 to 1.1 mg/dL (53 to 97.2 μmol/L) for women.
Creatinine alta	Values >1.3 mg/dL for men and >1.1 mg/dL for women.
Ferritin normal	The range of normal values for ferritin are: Men: 12 to 300 nanograms per milliliter (ng/mL) Women: 12 to 150 ng/mL.
ferritin alta	Ferritin values in Men of >300 nanograms per milliliter (ng/mL), in women >150 ng/mL.
Dimero D normal	The normal range for D-dimer is less than 0.5 micrograms per milliliter.
Dimero D Alto	D-dimer >0.5 micrograms per milliliter.
Fibrinogen normal	The normal range for fibrinogen is 200 to 400 mg/dL (2.0 to $4.0 \text{ g/L}$ ).
Fibrinogeno Alto	Fibrinogen value >400 mg/dL.
PCR normal	0 y 5 mg/dl
PCR alta	above 5 mg/dl
Procalcitonina normal	Normal blood procalcitonin values are less than 0.5 ng/mL

Variable	Description
Procalcitonina alta	Procalcitonin values in blood >0.5 ng/mL
Troponina normal	Troponin in blood, within the reference limit up to 0.04 ng/mL.
Tropoonina alta	Troponin in blood >0.04 ng/mL.
CPK normal	Normal values for creatine phosphokinase (CPK) are between 32 and 294 U/L for men and 33 to 211 U/L for women.
CPK alta	CPK values greater than 294 U/L for men and greater than 211 U/L for women
CK-MB normal	CK-MB blood values within a range of 5 to 25 IU/L.
CK-MB alta	CK-MB blood values >25 IU/L.
Albumina baja	Albumin in blood <3.4 g/dL.
Albumina normal	Albumin in blood with a range of 3.4 to 5.4 g/dL.
Bilirrubina total normal	Total blood bilirubin values of 3–1.9 mg/dL
Bilirrubina total alta	Total blood bilirubin values >1.9 mg/dL
ALT/TGP normal	ALT blood values in a range of 10–40 IU/L.
ALT/TGP alta	ALT blood values >40 IU/L.
AST/TGO normal	AST blood values in a range of 10–34 IU/L.
AST/TGO alta	AST blood values >34 IU/L.
DHL normal	DHL blood value in a range of 105–333 IU/L
DHL alta	DHL blood value >333 IU/L.
DHL > 1000	DHL > 1000
Fostasa alcalina normal	Alkaline phosphatase blood value in a range of 44–147 IU/L
Fosfatasa alcalina alta	Alkaline phosphatase blood value >147 IU/L.
POSITIVA	PCR sample for COVID-19 positive.
MODERADO	Clinical or radiographic evidence of lower respiratory tract disease, with an oxygen saturation greater than or equal to 94%.
GRAVE	Oxygen saturation <94%, respiratory rate > or equal to 30 breaths/minute, pulmonary infiltrates >50%.
Oseltamivir	A drug that selectively inhibits the neuraminidase enzyme found in influenza A and B viruses, preventing infected cells from releasing viral particles. Its action is greater against influenza A viruses.
Ceftriaxone	Antibiotic of the third generation cephalosporin class, which has broad-spectrum actions against Gram-negative and Gram-positive bacteria.
Claritromicina	Macrolide antibiotic active against gram positives, gram negatives, it is also active against spirochetes, Chlamydophila and several intracellular pathogens.
Azitromicina	Broad-spectrum antibiotic from the group of Macrolides that act against various gram-positive and gram-negative bacteria.
Levofloxacin	Antibacterial fluoroquinolone, used to treat infections caused by sensitive germs.
Otros	Other medications
Especificar	Specify

Variable	Description
Hidroxicloroquina/ Cloroquina	Commonly prescribed aminoquinoline for the treatment of uncomplicated malaria, rheumatoid arthritis, chronic discoid lupus erythematosus, and systemic lupus erythematosus.
Tocilizumab	Humanized monoclonal antibody that inhibits interleukin 6 receptors.
Esteroides	A group of chemicals classified by a specific carbon structure. Steroids include drugs used to relieve inflammation, such as prednisone and cortisone.
Pronación	Anatomical position of the human body characterized by: Body position lying face down and head on its side.
Responder	That responds to a stimulus.
Responder parcial	That responds little to a stimulus.
No respondedor	Non responder
Alta por mejoría	Discharge for improvement
Defunción	Death
Días de estancia hospitalaria	Days of hospital stay.

# References

- 1. Helmy, Y.A.; Fawzy, M.; Elaswad, A.; Sobieh, A.; Kenney, S.P.; Shehata, A.A. The COVID-19 pandemic: A comprehensive review of taxonomy, genetics, epidemiology, diagnosis, treatment, and control. *J. Clin. Med.* **2020**, *9*, 1225. [CrossRef]
- 2. Shereen, M.A.; Khan, S.; Kazmi, A.; Bashir, N.; Siddique, R. COVID-19 infection: Origin, transmission, and characteristics of human coronaviruses. *J. Adv. Res.* 2020, 24, 91–98. [CrossRef]
- 3. Guan, W.J.; Ni, Z.Y.; Hu, Y.; Liang, W.H.; Ou, C.Q.; He, J.X.; Liu, L.; Shan, H.; Lei, C.L.; Hui, D.S.; et al. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *N. Engl. J. Med.* **2020**, *382*, 1199–1207.
- 4. Wang, W.; Tang, J.; Wei, F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China. J. Med. Virol. 2020, 92, 441–447. [CrossRef] [PubMed]
- 5. Singhal, T. A Review of Coronavirus Disease-2019 (COVID-19). Indian J. Pediatrics 2020, 87, 281–286. [CrossRef]
- 6. Rothan, H.A.; Byrareddy, S.N. The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak. *J. Autoimmun.* **2020**, *109*, 102433. [CrossRef]
- 7. Wu, Z.; Mcgoogan, J.M. Characteristics of and Important Lessons from the Coronavirus Disease 2019 (COVID-19) Outbreak in China. *JAMA* 2020, 323, 1239. [CrossRef] [PubMed]
- 8. World Health Organization. WHO Coronavirus (COVID-19) Dashboard. 2021. Available online: https://covid19.who.int/ (accessed on 15 August 2021).
- World Health Organization. Estimating Mortality from COVID-19. Scientific Brief. 2020. Available online: https://apps.who. int/iris/bitstream/handle/10665/333642/WHO-2019-nCoV-Sci\_Brief-Mortality-2020.1-eng.pdf?sequence=1&isAllowed=y (accessed on 10 May 2020).
- Secretaría de Salud. COVID-19 Tablero México. COVID-19 Tablero México. 2020. Available online: https://coronavirus.gob.mx/ datos/ (accessed on 16 August 2021).
- 11. Yadaw, A.S.; Li, Y.C.; Bose, S.; Iyengar, R.; Bunyavanich, S.; Pandey, G. Clinical predictors of COVID-19 mortality. *medRxiv* 2020. Available online: https://pubmed.ncbi.nlm.nih.gov/32511520/ (accessed on 11 July 2021).
- Yao, H.; Zhang, N.; Zhang, R.; Duan, M.; Xie, T.; Pan, J.; Peng, E.; Huang, J.; Zhang, Y.; Xu, X.; et al. Severity Detection for the Coronavirus Disease 2019 (COVID-19) Patients Using a Machine Learning Model Based on the Blood and Urine Tests. *Front. Cell Dev. Biol.* 2020, *8*, 683. Available online: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7411005/ (accessed on 11 June 2021). [CrossRef] [PubMed]
- Alyasseri, Z.A.A.; Al-Betar, M.A.; Doush, I.A.; Awadallah, M.A.; Abasi, A.K.; Makhadmeh, S.N.; Alomari, O.A.; Abdulkareem, K.H.; Adam, A.; Damasevicius, R.; et al. *Review on COVID-19 Diagnosis Models Based on Machine Learning and Deep Learning Approaches. Expert Systems*; John Wiley and Sons Inc.: Hoboken, NJ, USA, 2021. Available online: https://www.ncbi.nlm.nih.gov/ pmc/articles/PMC8420483/ (accessed on 11 June 2021).
- Li, W.T.; Ma, J.; Shende, N.; Castaneda, G.; Chakladar, J.; Tsai, J.C.; Apostol, L.; Honda, C.O.; Xu, J.; Wong, L.M.; et al. Using machine learning of clinical data to diagnose COVID-19: A systematic review and meta-analysis. *BMC Med. Inform. Decis. Making* 2020, 20, 247. [CrossRef]

- Guan, X.; Zhang, B.; Fu, M.; Li, M.; Yuan, X.; Zhu, Y.; Peng, J.; Guo, H.; Lu, Y. Clinical and inflammatory features based machine learning model for fatal risk prediction of hospitalized COVID-19 patients: Results from a retrospective cohort study. *Ann. Med.* 2021, 53, 257–266. [CrossRef] [PubMed]
- Delafiori, J.; Navarro, L.C.; Siciliano, R.F.; de Melo, G.C.; Busanello, E.N.B.; Nicolau, J.C.; Sales, G.M.; de Oliveira, A.N.; Val, F.F.A.; de Oliveira, D.N.; et al. COVID-19 Automated Diagnosis and Risk Assessment through Metabolomics and Machine Learning. Analytical Chemistry. *Am. Chem. Soc.* 2021, *93*, 2471–2479. [CrossRef]
- 17. Allam, M.; Cai, S.; Ganesh, S.; Venkatesan, M.; Doodhwala, S.; Song, Z.; Hu, T.; Kumar, A.; Heit, J.; COVID-19 Study Group. COVID-19 Diagnostics, Tools, and Prevention. *Diagnostics* **2020**, *10*, 409. [CrossRef]
- Assaf, D.; Gutman, Y.; Neuman, Y.; Segal, G.; Amit, S.; Gefen-Halevi, S.; Shilo, N.; Epstein, A.; Mor-Cohen, R.; Biber, A.; et al. Utilization of machine-learning models to accurately predict the risk for critical COVID-19. *Intern. Emergency Med.* 2020, 15, 1435–1443. [CrossRef] [PubMed]
- Naseem, M.; Akhund, R.; Arshad, H.; Ibrahim, M.T. Exploring the Potential of Artificial Intelligence and Machine Learning to Combat COVID-19 and Existing Opportunities for LMIC: A Scoping Review. J. Primary Care & Community Health 2020, 11, 215013272096363. [CrossRef]
- 20. Arga, K.Y. COVID-19 and the Futures of Machine Learning. OMICS A J. Integr. Biol. 2020, 24, 512–514. [CrossRef]
- 21. Majhi, R.; Thangeda, R.; Sugasi, R.P.; Kumar, N. Analysis and prediction of COVID-19 trajectory: A machine learning approach. *J. Public Aff.* **2020**, e2537. [CrossRef]
- Van Der Schaar, M.; Alaa, A.M.; Floto, A.; Gimson, A.; Scholtes, S.; Wood, A.; McKinney, E.; Jarrett, D.; Lio, P.; Ercole, A. How artificial intelligence and machine learning can help healthcare systems respond to COVID-19. *Mach. Learn.* 2020, 10, 1–14. [CrossRef]
- 23. Das, A.K.; Mishra, S.; Saraswathy Gopalan, S. Predicting COVID-19 community mortality risk using machine learning and development of an online prognostic tool. *PeerJ* 2020, *8*, e10083. [CrossRef]
- 24. Swapnarekha, H.; Behera, H.S.; Nayak, J.; Naik, B. Role of intelligent computing in COVID-19 prognosis: A state-of-the-art review. *Chaos Solitons Fractals* **2020**, *138*, 109947. [CrossRef]
- 25. Silva, L.; Figueiredo Filho, D. Using Benford's law to assess the quality of COVID-19 register data in Brazil. *J. Public Health* **2020**, 43, 107–110. [CrossRef]
- Lee, K.; Han, S.; Jeong, Y. COVID-19, flattening the curve, and Benford's law. *Phys. A Stat. Mech. Appl.* 2020, 559, 125090. [CrossRef] [PubMed]
- 27. Panorama Epidemiologico. Enfermedades No Transmisibles. Secretaría de Salud. 2018. Available online: https://epidemiologia. salud.gob.mx/gobmx/salud/documentos/pano-OMENT/Panorama\_OMENT\_2018.pdf (accessed on 17 November 2021).
- Petrova, D.; Salamanca-Fernández, E.; Rodríguez Barranco, M.; Navarro Pérez, P.; Jiménez Moleón, J.; Sánchez, M. La obesidad como factor de riesgo en personas con COVID-19: Posibles mecanismos e implicaciones. *Atención Primaria* 2020, 52, 496–500. [CrossRef]
- 29. Monteagudo, D.E. La obesidad: Posibles mecanismos que explican su papel como factor de riesgo de la COVID-19. *Revista Cubana de Alimentación y Nutrición* **2020**, *30*, 12.
- 30. Pérez-Martínez, P.; Carrasco Sánchez, F.J.; Carretero Gómez, J.; Gómez-Huelgas, R. Resolviendo una de las piezas del puzle: COVID-19 y diabetes tipo 2. *Rev. Clin. Esp.* **2020**, 220, 507–510. [CrossRef]
- Giralt-Herrera, A.; Rojas-Velázquez, J.; Leiva-Enríquez, J.; Giralt-Herrera, A.; Rojas-Velázquez, J.; Leiva-Enríquez, J. Relación entre COVID-19 e Hipertensión Arterial. Scielo.sld.cu. Available online: http://scielo.sld.cu/scielo.php?pid=S1729-519X2020000 200004&script=sci\_arttext&tlng=en (accessed on 18 November 2020).
- 32. Di Castelnuovo, A.; Bonaccio, M.; Costanzo, S.; Gialluisi, A.; Antinori, A.; Berselli, N.; Blandi, V.; Bruno, R.; Guaraldi, G. Common cardiovascular risk factors and in-hospital mortality in 3,894 patients with COVID-19: Survival analysis and machine learning-based findings from the multicentre Italian CORIST Study. *Nutr. Metab. Cardiovasc. Dis.* 2020, 30, 1899–1913. [CrossRef]
- 33. De Souza, F.S.H.; Hojo-Souza, N.S.; Dos Santos, E.B.; Da Silva, C.M.; Guidoni, D.L. Predicting the disease outcome in COVID-19 positive patients through Machine Learning: A retrospective cohort study with Brazilian data. *Front. Artif. Intell.* 2021, *4*, 579931. Available online: https://www.medrxiv.org/content/10.1101/2020.06.26.20140764v1 (accessed on 22 February 2021). [CrossRef] [PubMed]
- Comunicado de Prensa No. 10. Coneval.org.mx. 2019. Available online: https://www.coneval.org.mx/SalaPrensa/ Comunicadosprensa/Documents/2019/COMUNICADO\_10\_MEDICION\_POBREZA\_2008\_2018.pdf (accessed on 22 February 2021).
- 35. Hopkins, J. Mortality Analyses-Johns Hopkins Coronavirus Resource Center. Johns Hopkins Coronavirus Resource Center. 2021. Available online: https://coronavirus.jhu.edu/data/mortality (accessed on 21 February 2021).
- Cao, Y.; Hiyoshi, A.; Montgomery, S. COVID-19 case-fatality rate and demographic and socioeconomic influencers: Worldwide spatial regression analysis based on country-level data. *BMJ Open* 2020, *10*, e043560. [CrossRef] [PubMed]