Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups among COVID-19 Patients in the Emergency Department

Julián Benito-León, Mª Dolores del Castillo, Alberto Estirado, Ritwik Ghosh, Souvik Dubey, J. Ignacio Serrano

Submitted to: Journal of Medical Internet Research
on: November 23, 2020

Table of Contents

Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups among COVID-19 Patients in the Emergency Department

Julián Benito-León[1] MD, PhD; Mª Dolores del Castillo[2] PhD; Alberto Estirado[3]; Ritwik Ghosh[4] MD; Souvik Dubey[5] MD, DM; J. Ignacio Serrano[2] PhD

[1]Department of Neurology, University Hospital "12 de Octubre" Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED) Department of Medicine, Complutense University Madrid ES
[2]Neural and Cognitive Engineering group, Automation and Robotics Center (CAR), CSIC-UPM Arganda del Rey ES
[3]HM Hospitals Madrid ES
[4]Department of General Medicine, Burdwan Medical College & Hospital, Burdwan Burdwan IN
[5]Department of Neuromedicine, Bangur Institute of Neurosciences Kolkata IN

Corresponding Author:
Julián Benito-León MD, PhD
Department of Neurology, University Hospital "12 de Octubre"
Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED)
Department of Medicine, Complutense University
Avenida de Córdoba km. 5,400
Madrid
ES

Abstract

Background: Early detection and intervention are the key factors for improving outcomes in COVID-19.

Objective: To detect severity subgroups among COVID-19 patients, based only on clinical data and standard laboratory tests obtained during the assessment at the emergency department.

Methods: We applied unsupervised machine learning to a dataset of 853 COVID-19 patients from HM hospitals in Spain.

Results: From a total of 850 variables, four tests, the serum levels of aspartate transaminase (AST), lactate dehydrogenase (LDH) and C-reactive protein (CRP), and the number of neutrophils, were enough to segregate the entire patient pool into three separate clusters. Further, the percentage of monocytes and lymphocytes and the levels of alanine transaminase (ALT) distinguished the cluster 3 from the other two clusters. The cluster 1 was characterized by the higher mortality rate and higher levels of AST, ALT, LDH, CRP and number of neutrophils, and low percentage of monocytes and lymphocytes. The cluster 2 included patients with a moderate mortality rate and medium levels of the previous laboratory determinations. The cluster 3 was characterized by the lower mortality rate and lower levels of AST, ALT, LDH, CRP and number of neutrophils, and higher percentage of monocytes and lymphocytes. Age, sex, comorbidities, and vital signs did not allow us to separate the three clusters. An online cluster assignment tool can be found at https://g-nec.car.upm-csic.es/COVID19-severity-group-assessment/.

Conclusions: A few standard laboratory tests, deemed to be available in all emergency departments, have shown far discriminative power for characterization of severity subgroups among COVID-19 patients.

(JMIR Preprints 23/11/2020:25988)
DOI: https://doi.org/10.2196/preprints.25988

Preprint Settings

1) Would you like to publish your submitted manuscript as preprint?

✔ **Please make my preprint PDF available to anyone at any time (recommended).**
Please make my preprint PDF available only to logged-in users; I understand that my title and abstract will remain visible to all users.
Only make the preprint title and abstract visible.
No, I do not wish to publish my submitted manuscript as a preprint.

2) If accepted for publication in a JMIR journal, would you like the PDF to be visible to the public?

✔ **Yes, please make my accepted manuscript PDF available to anyone at any time (Recommended).**

Yes, but please make my accepted manuscript PDF available only to logged-in users; I understand that the title and abstract will remain v

Yes, but only make the title and abstract visible (see Important note, above). I understand that if I later pay to participate in <a href="http

Original Manuscript

Using Unsupervised Machine Learning to Identify Age- and Sex-Independent Severity Subgroups among COVID-19 Patients in the Emergency Department

Julián Benito León;[1,2,3*] Mª Dolores del Castillo;[4*] Alberto Estirado;[5]

Ritwik Ghosh;[6] Souvik Dubey;[7] J. Ignacio Serrano[4*]

Department of Neurology, University Hospital “12 de Octubre”,[1] Madrid, Spain; Centro de Investigación Biomédica en Red sobre Enfermedades Neurodegenerativas (CIBERNED),[2] Madrid, Spain; Department of Medicine, Complutense University,[3] Madrid, Spain; Neural and Cognitive Engineering group, Automation and Robotics Center (CAR),[4] CSIC-UPM, Arganda del Rey, Spain; HM hospitals,[5] Madrid, Spain; Department of General Medicine,[6] Burdwan Medical College & Hospital, Burdwan, West Bengal, India; Department of Neuromedicine,[7] Bangur Institute of Neurosciences, Kolkata, India.

* These authors contributed equally to the manuscript

Correspondence

Julián Benito León MD. PhD. jbenitol67@gmail.com

Keywords: COVID-19; Machine Learning; Outcome

References: 29; **Tables:** 3; **Figures:** 1

Word Count: Abstract 269; Body text 3,156.

Authors’ contributions

Dr. Benito-León (jbenitol67@gmail.com) collaborated in: 1) the conception, organization and execution of the research project; 2) the writing of the manuscript first draft and the review and critique of the manuscript.

Dr. Del Castillo (md.delcastillo@csic.e) collaborated in: 1) the conception, organization of the research project; 2) the statistical analyses; and 3) the writing of the manuscript first draft and the review and critique of the manuscript.

Mr. Estirado (aestirado@hmhospitales.com) collaborated in: 1) the organization of the research project, and; 2) the review and critique of the manuscript.

Dr. Ghosh (alvarosferro@hotmail.com) collaborated in: 1) the organization of the research project, and; 2) the review and critique of the manuscript.

Dr. Dubey (drsouvik79@gmail.com) collaborated in: 1) the organization of the research project, and; 2) the review and critique of the manuscript.

Dr. Serrano (jignacio.serrano@csic.es) in: 1) the conception, organization of the research project; 2) the statistical analyses; and 3) the writing of the manuscript first draft and the review and critique of the manuscript.

Abstract

Background: Early detection and intervention are the key factors for improving outcomes in the 2019 coronavirus infectious disease (COVID-19). **Objective:** Our aim was to identify severity subgroups (clusters) among COVID-19 patients, based exclusively on clinical data and standard laboratory tests, obtained during the assessment in the emergency department. **Methods**: We applied unsupervised machine learning to a dataset of 853 COVID-19 patients from HM hospitals in Madrid (Spain). Age and sex were not considered while building the clusters as these variables could introduce biases in machine learning algorithms and raise ethical implications or discriminations in triage protocols. **Results:** From 850 clinical and laboratory variables, four tests, the serum levels of aspartate transaminase (AST), lactate dehydrogenase (LDH) and C-reactive protein (CRP), and the number of neutrophils, were enough to segregate the entire patient pool into three separate clusters. Further, the percentage of monocytes and lymphocytes and the levels of alanine transaminase (ALT) distinguished the cluster 3 from the other two clusters. The highest mortality rate and the highest levels of AST, ALT, LDH, CRP and number of neutrophils, and low percentage of monocytes and lymphocytes, characterized the cluster 1. The cluster 2 included patients with a moderate mortality rate and medium levels of the previous laboratory tests. The lowest mortality rate and the lowest levels of AST, ALT, LDH, CRP and number of neutrophils, and the highest percentage of monocytes and lymphocytes, characterized the cluster 3. An online cluster assignment tool can be found at https://g-nec.car.upm-csic.es/COVID19-severity-group-assessment/.

Conclusion: A few standard laboratory tests, deemed available in all emergency departments, have shown far discriminative power for characterization of severity subgroups among COVID-19 patients.

1. INTRODUCTION

The 2019 coronavirus infectious disease (COVID-19) pandemic has brought the scarcity of healthcare resources worldwide to light.[1] One of the main challenges faced by the healthcare systems while tackling this pandemic is the lack of affordable, accurate and simple information that can allow clinicians to predict the evolution of the patients sooner,

upon admission to the hospital. This information might help the clinicians to take early decisions regarding arrangement and organization of medical resources, as well as early interventions to improve health outcomes in these patients.

The exhaustive and inefficiently structured amount of health data available do not permit parametric modelling in an easy way. To overcome this issue, machine learning techniques have recently been identified as promising tools in data analysis for individual class prediction allowing us to deal with a great number of variables simultaneously and observe inherent disease-related patterns in the data.[2]

Machine learning for healthcare is a key discipline aimed to translate large health datasets into operative knowledge in different medicine fields.[3-7] The methods of this paradigm of artificial intelligence can be classified, on the basis on the underlying strategy used, in supervised and unsupervised.[8] In inductive or supervised machine learning, the method builds, from a set of previously categorized examples, a general class description of the target categories.[8] In general, supervised learning methods are used to design classifiers from labeled samples that predict the class of an unseen new sample.[8] In the field of medicine, these methods have been applied to find prognostic and predictive biomarkers.[9] On the other hand, in unsupervised machine learning, the goal is to find the class or classes that cover the sample.[8] These methods permit to discover the underlying structure and relations among unlabeled samples.[8] Unsupervised clustering techniques can obtain groups of samples, so that the intra-similarity within each group is maximized, while inter-similarity between groups is minimized.[8] They are usually applied in medicine to identify homogeneous groups of patients based on their medical records and relations between clinical manifestations and therapeutic responses, or to detect sets of co-expressed genes, among others.[10, 11]

There are several research reports using COVID-19 datasets, which focus on predicting the patients' mortality or severity by using mainly regression modeling from labeled clinical records.[12-17] Further, in a multi-center study, using supervised machine learning, a personalized COVID-19 mortality risk score for hospitalized patients upon admission has been proposed;[18] however, in that study,[18] it was not explained the reason for choosing only a subset of the recorded clinical variables to build their model. Therefore, the algorithm might have been biased, even by expert's knowledge. In all these studies [12-18] and in a study based on cluster analysis,[19] demographics, such as age and sex, were considered as key variables in their prediction models. By contrast, these variables were deliberately excluded from the training dataset in the present study in which we used an unsupervised machine learning method for data handling.

Health agencies recommend that clinical decisions should be made based on individual's biological age rather than chronological age.[20, 21] There are multiple physiological and molecular markers for estimation of biological aging that can predict lifespan.[22] Beside these markers, the heterogeneity of eating habits, physical and mental conditions, and therapeutics have influence on the overall health state, making biological aging a heterogeneous process too.

Frailty and multi-morbidities, as measures of biological aging, have been found as risk factors for mortality independent of chronological age in patients with COVID-19.[23] New procedures for the therapeutic management of COVID-19 are required regardless chronological age.[24]

Furthermore, reports about case-fatality rate for COVID-19 scheduled by age groups could sentence elderly people not only to social exclusion, but also to healthcare indifference. Considering elderly population as a highly vulnerable group is a simple and

negative stereotyping that may even influence decision making in the clinical resource management.[25]

There is also a different gender-based prevalence and severity of COVID-19, demonstrating men having higher mortality than women.[26] The severity of the disease implies that the person may need hospitalization, intensive care support, and mechanical ventilation. However, the medical treatments scheduled during hospitalization or stay in intensive care is the same for every severe patient of COVID-19, regardless age or sex.[27]

Since chronological age as well as sex cannot be considered as pivotal aspects to determine the individual's health status and resilience,[28] these should not be key determinants for healthcare or resource allocation amid people suffering from COVID-19. Therefore, predictive models based on intelligent data processing that takes into account the age as a major determinant in the access for healthcare may be inappropriate and unethical.[25]

Demographic variables (i.e., age and sex) were not used by the previously published studies, which were based on either artificial intelligence algorithms or regression modeling, for building models on effective treatments based upon sex or age groups or for understanding sex or age differences.[12-19] These predictive models of severity and mortality risk for COVID-19 could be discriminating.[29] For example, consideration of the age of people in the emergency department might discriminate against older people (ageism) in the access to care, since the decision would be based purely upon age of the population rather than their healthcare needs.[30]

Our aim was to identify non-overlapping severity subgroups (clusters) among COVID-19 patients, using exclusively standard laboratory tests and clinical data obtained during the first medical contact in the emergency department, by means of unsupervised

machine learning techniques. Age and sex were not taken into account to build the subgroups by ethical implications. For this purpose, we used the dataset collected by the HM group of hospitals in Madrid (Spain) (HM, 2020).[31]

2. MATERIAL AND METHODS

2.1. Dataset

We used a dataset collected by the HM group of hospitals in Madrid (Spain) in the context of the project COVID DATA SAVE LIVES (HM, 2020).[31] The information of this dataset comes from the electronic health records data system of the seven HM hospitals, placed at the Madrid community in Spain.[31] This dataset contains the anonymized records of 2,310 patients, admitted to any of the seven HM hospitals, with a diagnosis of COVID-19 from March 1st, 2020 to April 24th, 2020. The dataset collects the different interactions in the COVID-19 treatment processes, including detailed information on diagnoses, treatments, intensive care unit (ICU) admission, discharge or death, among many other variables. They also include diagnostic imaging and laboratory tests or records of previous medical care, if any. It also includes the drugs administered to each patient during admission (more than 60,000 records) with the dates corresponding to the first and last administration of each drug identified by their brand name and classification in the Anatomical Therapeutic Chemical codes (ATC5/ATC7). Moreover, laboratory data are also included (398.884 records). Finally, it contains the records of the diagnostic and procedural information, coded according to the international ICD-10 classification in its latest distributed version, for the patients referred, both for episodes of hospital admission (more than 1,600) and for the emergency (more than 1,900) prior to those episodes, if any.

2.2. Data preprocessing

We collected the information for each patient identifier in one record. This included age, sex, vital constants in the emergency department and the need or not of ICU. COVID-19 symptoms, ICD-10 codes of previous and current conditions, as well as different laboratory tests performed in the emergency department, were also recorded. We also calculated for each patient the duration (in days) of the hospital stay, including ICU admission and the days from hospitalization to ICU admission. We also considered the first laboratory tests obtained from the emergency department and grouped all the ICD-10 codes under the first three characters (first letter and two subsequent numbers) of the code to reduce the number of variables and provide generalization. We codified each ICD-10 feature grouped as “present in emergency department admission”, “not present in emergency department admission” or “developed during hospital stay”.

Only the patients with discharge reason equals to “Death” or “Recovered” were included for the analyses. The patients with discharge reason equals to “Transferred to another hospital” or “Transferred back to the nursing home” (about 3.6% of the total dataset) were excluded since no additional information was available after they left the hospital. We only selected the records (patients) without missing values on clinical data and laboratory tests, which left a final sample of 853 (37.2% women) patients, who were included in our analyses. Their mean age was of 67.2 ± 15.7 years (range: 21-106). Each one of these patients had 850 variables, including eight regarding demographics, hospitalization stay and outcome measures; one about COVID-19 symptoms; 10 about vital signs (temperature, heart rate, oxygen saturation, and systolic and diastolic blood pressure) in the emergency department; 29 laboratory tests in the emergency department (see Table 1); 168 ICD-10 codes in the emergency department; and 634 ICD-10 codes during hospital stay.

The final sample of 853 was similar to the excluded sample (N=1,457) in terms of

age (66.2 ± 15.7 vs. 67.1 ± 17.0, F(1.2308) = 1.508, p = .22); discharge reason (selected deceased: 15.6% vs. excluded deceased: 18.2%, F(1.2308) = 2.474, p = .116); ICU admission (6.8% vs. 7.3%; F(1.2308) = 0.003, p = .957); or admission date (March 27, 2020 ± 8.3 days vs. March 28, 2020 ± 11.6 days). However, there were significant differences in terms of sex (37.2% women vs 42.2% women; F(1.2308) = 5.768, p = .016) and days in hospital (9 ± 6 vs 8 ± 7; F(1.2308) = 4.786, p = .029). Notwithstanding, the effect size was small for both differences (η2 = 0.003 and η 2= 0.002, respectively).

Table 1. Laboratory tests used to characterize the patients.

Code	Description	Unit
RDW	Red cell distribution width	%
BAS	Basophils	x10e^3/µL
BAS%	% of Basophils	%
MCHC	Mean corpuscular hemoglobin concentration	g/dL
CREA	Creatinine	mg/dL
EOS	Eosinophils	x10e^3/µL
EOS%	% of Eosinophils	%
GLU	Glucose	mg/dL
AST	Aspartate transaminase	U/L
ALT	Alanine transaminase	U/L
MHC	Mean cell hemoglobine	pg
HCT	Hematocrit	%
RBC	Red blood cells	x10e^6/µL
HB	Hemoglobin	g/dL
K	Potasium	mmol/L
LDH	Lactate dehydrogenase	U/L
LEUC	Leucocytes	x10e^3/µL
LYM	Lymphocytes	x10e^3/µL
LYM%	% of Lymphocites	%
MONO	Monocytes	x10e^3/µL
MONO%	% of Monocytes	%
NA	Sodium	mmol/L
NEU	Neutrophils	x10e^3/µL
NEU%	% of Neutrophils	%
CRP	C-reactive protein	mg/L
PLAT	Platelet count	x10e^3/µL
BUN	Blood urea nitrogen	mg/dL
MCV	Mean cell volume	fL
MPV	Mean platelet volume	fL

2.3. Clustering

The unsupervised automatic clustering X-Means,[32] concretely the implementation

in RapidMiner Studio 9.7 Community Edition (RapidMiner, Inc., Boston, MA, USA), was applied to the preprocessed dataset previously described (section Data preprocessing). The algorithm determines the optimum number of clusters so that the intra-cluster distance of patients is minimum, and the inter-cluster distance of patients is maximum. The X-Means algorithm was used instead of the most common K-Means algorithm to overcome the three major shortcomings of the latter:[32] poor computational scaling, manual selection of the number of clusters and tendency to local minima. X-Means determines the optimal number of clusters by the Bayesian Information Criterion (BIC), also known as Schwarz criterion, which is used to maximize the explained variance by the clusters and minimize the number of parameters (k).[32] The latter is also an improvement over K-means since this tends to create clusters formed by only one sample to minimize inertia.[32] Moreover, the later use of the Davies Bouldin index to evaluate the cluster distributions is also intended to overcome this issue since it considers a mix of both inertia and distortion to quantitatively asses the cluster models (see below). In addition, the automatic selection of the number of clusters by X-Means avoids the possible bias in the manual selection of k. [32] This bias is also present in Hierarchical Agglomerative Clustering (HAC), where a threshold must be set to obtain the ultimate clusters after the hierarchy is built. Despite X-Means is not completely deterministic, it is certainly very stable with minimum variations between different runs,[32] and significantly more stable than K-means. X-Means introduces a bias though. Since it uses the BIC to evaluate the cluster models in each iteration, this criterion purposely favors the models with a lesser number of clusters. This means that an alternative cluster model with a better Davies Boulding index and a higher number of clusters might have been discarded. However, more number of clusters with better Davies Boulding index usually implies clusters with small number of samples (notice that the best index would be obtained by a model of one cluster per sample), which is not desirable at all for the clinical

stratification purpose aimed in this study.

Patients were considered here as vectors with several dimensions equal to the number of variables. In this case, the number of variables taken to apply the clustering algorithm was 842. None of the eight variables regarding demographics, hospital stay and outcome measures were included. They were removed from the clustering formation because the potential ethical controversies and biasing (demographics) or prospective information (hospitalization stay and outcome measures). The algorithm was applied using several similarity or distance metrics between patients:[33] the Euclidean distance, the Camberra distance, the Chebychev distance, the correlation similarity, the cosine similarity, the Dice similarity, the Inner Product similarity, the Jaccard similarity, the Kernel Euclidean distance, the Manhattan distance, the Max Product similarity, the overlap similarity, the Generalized divergence, the Itakura Saito distance, the KL divergence, the logarithmic loss, the logistic loss, the Mahalanobis distance, the Squared Euclidean distance and the Squared loss. In spite of we could have had *a priori* good similarity measure candidates, based on dataset characteristics, such as dimensionality, the best practice was the selection based on empirical evaluation.[34] To avoid any *a priori* biases, we just empirically tested all measures available in the software and just kept the one yielding the best results.

To assess the fitness of the cluster distributions from the algorithm executions with the above metrics, the Davies Bouldin index was calculated for each one of them.[35] The Davies Bouldin index is a common measure, which evaluates cluster models.[35] It quantifies the average maximum ratio of the within-cluster scatter to the between-cluster separation for every pair of clusters in a cluster model.[35] In other words, it provides a trade-off between inter-cluster similarity and intra-cluster distance.[35] With this definition, the lower the Davies Bouldin index the lower the within-cluster scatter and the higher the between-cluster separation, which is the best desirable property of a cluster model.[35] The

Davies Bouldin index allowed us to quantitatively select the best cluster model among those created, one for each similarity measure considered.

2.4. Cluster validation

From the 1,457 patients excluded due to missing values (not used to obtain the clusters), we performed a validation analysis with the patients who presented no missing values in the variables that statistically differed between the three clusters obtained. Subsequently, these patients were assigned to one of the clusters previously obtained by using the best distance metric determined in the clustering process described above.

2.5. Statistical analysis

The difference in the 850 variables between all the clusters obtained was tested by a multivariate analysis of variance (one-way MANOVA). Pairwise post-hoc comparisons between clusters were analyzed by Bonferroni test. Significance was accepted at the 5% level ($\alpha = 0.05$). The observed power and effect size (as partial eta squared) were reported for statistically significant differences.

3. RESULTS

Table 2 shows the number of clusters and the corresponding David Bouldin index of the cluster distribution of patients obtained by the X-Means clustering algorithm for each of the similarity measures tested. Note that the lower the David Bouldin index the better the cluster distribution (higher inter-cluster distance and lower intra-cluster distance). The best cluster distribution (lowest David Bouldin index) was obtained by using the Manhattan distance, which grouped the patients into three clusters.

Table 2. Number of clusters and the corresponding David Bouldin index.

Similarity measure	David Bouldin index	Number of clusters
Euclidean distance	0.948	3
Camberra distance	-	1
Chebychev distance	0.966	3
Correlation similarity	1.400	3
Cosine similarity	1.629	3
Dice similarity	-	1
Inner Product similarity	-	1
Jaccard similarity	1.387	3
Kernel Euclidean distance	1.440	3
Manhattan distance	***0.701***	***3***
Max Product similarity	-	1
Overlap similarity	5.099	4
Generalized divergence	3.445	3
Itakura Saito distance	5.919	4
KL divergence	5.677	4
Logarithmic loss	4.595	4
Logistic loss	3.445	3
Mahalanobis distance	4.595	4
Squared Euclidean distance	3.445	3
Squared loss	3.659	3

Demographic and clinical characteristics of the patients in the three clusters are shown in Table 3. Notice that this table also shows the values of the eight variables (demographics, hospital stay and outcome measures), which were not used in the construction of the clusters (marked with '†' in Table 3). The patients in cluster 1 had a significantly higher mortality rate (46.6%) than patients in cluster 2 (18.0%) and cluster 3 (10.0%). No significant difference in the percentage of ICU admission was found between clusters. However, the patients who were admitted to ICU in cluster 1 stayed a significantly shorter time than patients in cluster 3. No significant difference in sex was found between clusters. Patients in cluster 3 were significantly younger than those ones in cluster 1. In addition, clusters 1 and 2 presented a significantly higher heart rate in the emergency department than cluster 3. The average oxygen saturation in the emergency department

was significantly different between all clusters, showing the cluster 1 the lowest oxygen saturation and the cluster 3 the highest. With respect to the previous diseases and surgical procedures, cluster 1 presented a significantly higher percentage of epilepsy and emphysema than clusters 2 and 3. In addition, cluster 2 presented a higher percentage of previous surgical procedures as well as previous thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders than cluster 3, and a significantly lower percentage of disorders of purine and pyrimidine metabolism than clusters 1 and 3. Finally, the percentage of patients who developed surgical operations during the current hospitalization was significant higher in cluster 1 than in clusters 2 and 3.

Regarding the laboratory tests, the patients in cluster 1 showed significantly higher levels of serum creatinine, potassium, and blood urea nitrogen than clusters 2 and 3, and a significantly higher value of red cell distribution width than cluster 2. In addition, patients in cluster 2 presented significantly higher values of lymphocytes and serum levels of sodium, and significantly lower platelet count than patients in cluster 3. In addition, cluster 3 showed lower values of mean corpuscular hemoglobin concentration, leucocytes, serum levels of alanine transaminase (ALT), and percentage of neutrophils than clusters 1 and 2, and significantly higher values of eosinophils (and percentage) and percentage of lymphocytes than clusters 1 and 2. Finally, the laboratory tests that showed significant differences between all clusters were the serum levels of aspartate transaminase (AST) (cluster 1 > cluster 2 > cluster 3), lactate dehydrogenase (LDH) (cluster 1 > cluster 2 > cluster 3), C-reactive protein (CRP) (cluster 1 > cluster 2 > cluster 3) and the number of neutrophils (cluster 1 > cluster 2 > cluster 3).

Table 3. Demographic and clinical characteristics of the patients (N= 853) in the three clusters. Variables marked with the '†' superscript were not used for the cluster construction.

		Cluster 1	Cluster 2	Cluster 3	Statistics
		(N=58)	(N=300)	(N=495)	
Demographics					
Age†		71.1 (13.7)$_{a}$	67.0 (15.1)$_{a,b}$	65.1 (16.2)$_{b}$	F(2,850)=3.457;p=.032;η^2=0.009;1-β=0.648
Sex (men)†		70.7%$_{a}$	60.3%$_{a}$	63.3%$_{a}$	F(2,850)=1.027;p=.358; η^2=0.003;1-β =0.23
Hospital stay and outcome measures					
Inpatient hospital days†		8.5 (4.9)$_{a}$	8.6 (6.4)$_{a}$	8.3 (5.1)$_{a}$	F(2,850)=0.363;p=.695; η^2=0.001;1-β =0.109
Discharge reason†	*Recovered*	31 (53.4%)	246 (82.0%)	443 (89.5%)	F(2,850)=26.054;p<.0001; η^2=0.062;1-β=1
	Deceased	27 (46.6%)$_{a}$	54 (18.0%)$_{b}$	52 (10.5%)$_{b}$	
Intensive care unit admission†	*No*	52 (89.7%)	277 (92.3%)	458 (92.5%)	F(2,850)=1.12;p=.327; η^2=0.003;1-β=0.248
	Yes	6 (10.3%)$_{a}$	23 (7.7%)$_{a}$	37 (7.5%)$_{a}$	
Days until intensive care unit admission†		0.2 (0.4)$_{a}$	3.4 (6.3)$_{a}$	2.3 (4.3)$_{a}$	F(2,850)=1.393;p=.256; η^2=0.042;1-β=0.289
Days in intensive care unit†		0.2 (0.4)$_{a}$	4.8 (6.5)$_{a,b}$	7.6 (6.9)$_{b}$	F(2,850)=3.747;p=.029; η^2=0.106;1-β=0.665
Mechanical ventilation need†		35 (60.3%)$_{a}$	177 (59.0%)$_{a}$	277 (56.0%)$_{a}$	F(2,850)=0.163;p=.850; η^2<.0001;1-β=0.075
Vitals and laboratory tests					
First heart ratio measurement in the emergency department		98.4 (25.0)$_{a,b}$	100.1 (26.2)$_{a}$	93.5 (24.4)$_{b}$	F(2,850)=8.45;p<.0001;η^2=0.021;1-β=0.965
First oxygen saturation measurement in the emergency department		84.2 (12.3)$_{a}$	90.1 (7.6)$_{b}$	94.2 (3.6)$_{c}$	F(2,850)=81.732;p<.0001;η^2=0.171;1-β=1
Last heart ratio measurement in the emergency department		99.0 (25.1)$_{a,b}$	100.1 (26.)$_{a}$	93.6 (24.7)$_{b}$	F(2,850)=8.104;p<.0001; η^2=0.02;1-β=0.958
Last oxygen saturation measurement in the emergency department		84.2 (12.2)$_{a}$	90.0 (7.52)$_{b}$	94.2 (3.6)$_{c}$	F(2,850)=82.554;p<.0001; η^2=0.172;1-β=1
Red cell distribution width		13.6 (1.9)$_{a}$	12.9 (1.84)$_{b}$	13.0 (1.9)$_{a,b}$	F(2,850)=3.28;p=.038; η^2=0.008;1-β=0.623
Basophils		0.03 (0.03)$_{a}$	0.02 (0.02)$_{a,b}$	0.02 (0.0)$_{b}$	F(2,850)=5.545;p=.004; η^2=0.014;1-β=0.854
Mean corpuscular hemoglobin concentration		33.9 (1.5)$_{a}$	34.0 (1.17)$_{a}$	33.6 (1.2)$_{b}$	F(2,850)=8.602;p<.0001; η^2=0.021;1-β=0.968
Creatinine		1.3 (1.4)$_{a}$	1.0 (0.47)$_{b}$	1.0 (0.5)$_{b}$	F(2,850)=9.591;p<.0001; η^2=0.024;1-β=0.981
Eosinophils		0.02 (0.04) $_{a}$	0.02 (0.04)$_{a}$	0.04 (0.1)$_{b}$	F(2,850)=6.518;p=.002; η^2=0.016;1-β=0.908
% of Eosinophils		0.20 (0.5)$_{a}$	0.3 (0.60)$_{a}$	0.6 (1.2)$_{b}$	F(2,850)=10.000;p<.0001; η^2=0.025;1-β=0.985
Aspartate transaminase		80.3 (48.0)$_{a}$	55.8 (33.4)$_{b}$	32.8 (18.7)$_{c}$	F(2,850)=109.193;p<.0001; η^2=0.216;1-β=1
Alanine transaminase		57.2 (69.1)$_{a}$	50.7 (48.1)$_{a}$	29.5 (23.8)$_{b}$	F(2,850)=32.686;p<.0001; η^2=0.076;1-β=1
Potasium		4.6 (0.8)$_{a}$	4.2 (0.6)$_{b}$	4.2 (0.5)$_{b}$	F(2,850)=16.957;p<.0001; η^2=0.041;1-β=1
Lactate dehydrogenase		1339.72 (240.56)$_{a}$	742.5 (122.0)$_{b}$	447.7 (91.5)$_{c}$	F(2,850)=1666.635;p<.0001; η^2=0.808;1-β=1
Leucocytes		9.9 (4.8)$_{a}$	8.5 (4.2)$_{a}$	6.9 (5.2)$_{b}$	F(2,850)=13.055;p<.0001; η^2=0.032;1-β=0.997
Lymphocytes		1.0 (0.5)$_{a,b}$	1.0 (0.6)$_{a}$	1.3 (2.1)$_{b}$	F(2,850)=3.692;p=.025; η^2=0.009;1-β=0.679
% of Lymphocites		12.6 (7.8)$_{a}$	14.0 (7.7)$_{a}$	20.0 (9.8)$_{b}$	F(2,850)=46.962;p<.0001; η^2=0.106;1-β=1
% of Monocytes		5.1 (2.9)$_{a}$	6.6 (3.9)$_{a}$	8.7 (4.8)$_{b}$	F(2,850)=29.321;p<.0001; η^2=0.069;1-β=1
Sodium		136.2 (7.1)$_{a,b}$	136.2 (4.4)$_{a}$	137.2 (4.6)$_{b}$	F(2,850)=4.016;p=.018; η^2=.01;1-β=0.718
Neutrophils		8.4 (4.7)$_{a}$	6.9 (4.0)$_{b}$	4.9 (2.7)$_{c}$	F(2,850)=45.584;p<.0001; η^2=0.103;1-β=1
% of Neutrophils		81.8 (10.2)$_{a}$	78.8 (9.9)$_{a}$	70.4 (11.9)$_{b}$	F(2,850)=62.070;p<.0001; η^2=0.135;1-β=1

C-reactive protein		206.1 (131.7)$_a$	152.1 (110.0)$_b$	64.2 (63.7)$_c$	F(2,850)=12.930;p<.0001; η^2=0.233;1-β=1
Platelet count		229.0 (92.2)$_{a,b}$	236.3 (96.6)$_a$	210.3 (87.2)$_b$	F(2,850)=7.541;p=.001; η^2=0.019;1-β=0.944
Blood urea nitrogen		58.9 (56.6)$_a$	41.8 (29.0)$_b$	40.5 (29.7)$_b$	F(2,850)=7.579;p=.001; η^2=0.019;1-β=0.945
Diseases and Surgical procedures					
Previous history of disorders of purine and pyrimidine metabolism		6.9%$_b$	1.3%$_a$	5.1%$_b$	F(2,850)=4.179;p=.016; η^2=0.01;1-β =0.736
Previous history of epilepsy and recurrent seizures		5.2%$_a$	1.30%$_b$	0.4%$_b$	F(2,850)=5.660;p=.004; η^2=.014;1-β =.862
Previous history of emphysema		5.2%$_a$	0.70%$_b$	0.4%$_b$	F(2,850)=6.663;p=.001; η^2=0.017;1-β=0.914
Previous history of thoracic, thoracolumbar, and lumbosacral intervertebral disc disorders		0.0%$_{a,b}$	3.0%$_a$	0.6%$_b$	F(2,850)=4.385;p=.013; η^2=0.011;1-β=0.758
Previous history of surgical procedures		0.0% $_{a,b}$	1.3% $_a$	0.0% $_b$	F(2,850)=3.753;p=.024; η^2=0.009;1-β =0.686
Surgical operations during the current hospitalization		5.2%$_a$	0.7%$_b$	0.8%$_b$	F(2,850)=4.880;p=.008; η^2=0.012;1-β =0.804

Mean (median) ± standard deviation and frequency (%) are reported.
*Values in the same row not sharing subscript letters showed significant difference after Bonferroni post-hoc correction.
η^2: Effect size; 1- β: Observed power.

For a clearer characterization of the clusters, the Figure 1 shows a radar chart with the variables (hospital stay, outcome measures and laboratory tests), which showed statistically significant differences among the clusters and a medium or high effect size ($\eta^2 > 0.06$).[36].

An online cluster assignment tool, according to the results reported here, can be found at: https://g-nec.car.upm-csic.es/COVID19-severity-group-assessment/

To test the robustness of the identified clusters, we performed a validation analysis using the initially excluded patients without missing values in the variables that statistically differed among the three clusters (Table 3). Specifically, it was based on six variables (first and last oxygen saturation measurement in the emergency department, AST, LDH, neutrophils and CRP). For this purpose, we selected 349 patients initially excluded who were assigned to one of the three previously identified clusters by the minimum Manhattan distance to the average values of the six mentioned variables of those clusters. Table 4 shows the differences in demographics, hospital stay and outcome measures in the three clusters. Indeed, the clusters initially obtained were consistent with the cluster assigned in the validation analysis in terms of age, sex, hospital stay, and outcome measures. Specifically, the cluster 1 was the one with the oldest and highest number of men, as well as with the highest mortality rate and number of patients who required mechanical ventilation. By contrast, the cluster 3 was the one with the youngest and lowest number of men, and with the lowest mortality rate and number of patients who required mechanical ventilation.

Table 4. Demographics, hospital stay and prognosis of the patients (N= 349) selected for the validation analysis in the three clusters.

		Cluster 1 (N=18)	Cluster 2 (N=112)	Cluster 3 (N=219)	Statistics
Demographics					
Age		72.8 (14.2)$_{a,b}$	71.3 (14.3)$_{a}$	64.2 (15.8)$_{b}$	F(2,346)=9.414; p<.0001;η^2=0.052;1-β=0.979
Sex (men)		77.8%$_{a}$	60.7%$_{a}$	56.2%$_{a}$	F(2,346)=1.738;p=.177; η^2=0.010;1-β =0.364
Hospital stay and outcome measures					
Inpatient hospital days		9.1 (6.4)$_{a}$	9.3 (5.9)$_{a}$	8.0 (5.3)$_{a}$	F(2,346)=2.320;p=.100; η^2=0.013;1-β =0.469
Discharge reason	*Recovered*	8 (44.4%)	80 (71.4%)	200 (91.3%)	F(2,346)=22.025;p<.0001; η^2=0.113;1-β=1
	Deceased	10 (55.6%)$_{a}$	32 (28.6%)$_{b}$	19 (8.7%)$_{c}$	
Intensive care unit admission	*No*	16 (88.9%)	101 (90.2%)	213 (97.3%)	F(2,346)=4.268;p=.015; η^2=0.024;1-β=0.743
	Yes	2 (11.1%)$_{a,b}$	11 (9.8%)$_{a}$	6 (2.7%)$_{b}$	
Days until intensive care unit admission		6.5 (7.8)$_{a}$	4.1 (3.9)$_{a}$	6.3 (13.7)$_{a}$	F(2,16)=.170;p=.845; η^2=0.021;1-β=0.072
Days in intensive care unit		4.5 (0.7)$_{a}$	3.8 (4.5)$_{a}$	3.2 (4.6)$_{a}$	F(2,16)=.082;p=.922; η^2=0.010;1-β=0.060
Mechanical ventilation need		12 (66.7%)$_{a}$	54 (48.2%)$_{a}$	96 (43.8%)$_{a}$	F(2,346)=1.854;p=.158; η^2=0.011;1-β=0.385

Mean (median) ± standard deviation and frequency (%) are reported.
*Values in the same row not sharing subscript letters presented significant differences after Bonferroni post-hoc correction.
η^2: Effect size; 1- β: Observed power.

4. DISCUSSION

With the application of unsupervised machine learning approach, we could identify and segregate patients with COVID-19 into subgroups depending upon the severity of disease, simply using standard laboratory tests performed during the first medical assessment in the emergency department. We found that inflammatory (CRP), hematologic (number of neutrophils and percentage of monocytes and lymphocytes), and serum biochemical abnormalities (AST, ALT, and LDH), mainly indicating liver dysfunction, detected upon admission to the hospital, can predict the severity of the disease. From a sum of 850 variables, collected in the emergency department, only four standard laboratory tests, i.e. serum AST, LDH, CRP, and the number of neutrophils, were enough to segregate these patients into three separate clusters. Of these, the levels of LDH showed the biggest effect size, practically allowing us to differentiate the three clusters linearly. Further, the percentage of monocytes and lymphocytes as well as ALT distinguished the cluster 3 (less severe) from the other two. The cluster 1 was characterized by the higher mortality rate, early ICU admission and high values of AST, ALT, LDH, number of neutrophils and CRP, and low values of monocytes (%) and lymphocytes (%) (See Figure 1). The cluster 2 included a group of patients with a moderate mortality rate, late ICU admission and medium values of the previous laboratory tests (see Figure 1). Finally, the cluster 3 was characterized by the lower mortality rate, medium time to ICU admission and lower values of AST, ALT, LDH, number of neutrophils and CRP, and high values of monocytes (%) and lymphocytes (%) (See Figure 1).

Our results have several clinical implications. First, age and sex were not considered while building the clusters. Therefore, our unsupervised machine

learning approach, based exclusively on the performance of simple laboratory tests at a primordial stage, would permit to establish a strategy for rationing of healthcare resources, and to settle a triage protocol, which would support medical decisions in a transparent and ethical way. Second, since the analyzed data are standard laboratory tests, this method would have a special value for underdeveloped and developing areas with lack of medical resources and affordability issues. Finally, we could tailor treatment to each severity group accordingly at a primordial stage (i.e., in the emergency department). For example, more aggressive therapies could be considered in patients classified in the cluster 1 (the most severe), and not in those ones in the cluster 3 (the least severe).

Initially, the novel severe acute respiratory syndrome coronavirus (SARS-CoV-2) was primarily considered as a respiratory pathogen. However, with time it has behaved as a virus with the potential to cause multi-system involvement. [37, 38] Specifically, hepatic injury related to COVID-19 is only beginning to unravel. Elevated liver injury indicators, particularly AST, are strongly associated with a higher mortality risk in COVID-19 patients.[39] Of note, high serum levels of LDH predicts higher in-hospital mortality in severe and critically ill COVID-19 patients.[40] Significant increased CRP levels in early stages of COVID-19 disease are correlated with the severity of disease and the degree of internal tissue pathologies.[41] Further, a significant increase in the number of neutrophils with decrease in the number of lymphocytes, monocytes and eosinophils may indicate clinical worsening and increased risk of a poor outcome among COVID-19 patients.[42] Taken together, the presence of elevated biomarkers of inflammation and that of liver injury in serum, as well as

the number of neutrophils at admission are heralding a multiple organ failure in COVID-19 patients that could lead to death. Our laboratory findings are in agreement with other previous studies worldwide.[43-45]

Although one previous multi-center study, based on the analyses of demographics, comorbidities, vital signs, and laboratory test results upon admission, on the prediction of disease course in COVID-19, has been undertaken,[18] there remains much to learn about applying machine learning techniques in this novel infectious disease. Comparison with that study is difficult, as they have used different variables and techniques. The accuracy of the model could be influenced by several factors including methods. Feature extraction methods, feature selection or classification tools, number of subjects, and demographics are also important considerations. Besides, most COVID-19 diagnostics and prognostic models that have evolved to date have a high risk of generating bias leading to inequality,[46] mainly due to the high influence of demographic variables (specially age and sex) in those models and to the non-blinded nature of the supervised machine-learning approach between predictors and outcome measures. In fact, our results confirmed that age and sex had a similar and little discriminant value to separate the three clusters (table 3). Nevertheless, the results obtained in our study are in line with most previous works based on supervised machine-learning techniques in COVID-19.[18, 46]

The study should be interpreted within the context of several limitations. First, the patients in the current study may represent a selected group of COVID-19 patients (i.e., patients with a more severe disease since all of them were admitted to the hospital), and hence it is questionable to what extent our results could be generalized to the entire COVID-19 population. The reason for

this was that the extreme circumstances in our hospitals at the peak of this pandemic permitted only to hospitalize the most severe cases. Notwithstanding, our aim was to detect severity subgroups among COVID-19 patients upon admission to the hospital. Second, we only kept the records (patients), laboratory tests, and clinical variables from 853 patients from the dataset due to the high number of missing values in the remaining 1,457 patients. Despite this, the results have been robust.

In closing, to the authors' knowledge, the work presented in this paper is the first attempt to use unsupervised machine learning to identify severity subgroups among COVID-19 patients upon admission. A few affordable, simple, and standard laboratory tests, which are expected to be available in any emergency department, have shown promising discriminative power for characterization of severity subgroups among COVID-19 patients. We have also provided an online severity cluster assignment tool for COVID-19 patients who are admitted to the emergency department. This could permit to classify the patients according to severity subgroups and hence initiate earlier interventions.

REFERENCES

1. Joebges S, Biller-Andorno N. Ethics guidelines on COVID-19 triage-an emerging international consensus. Critical care (London, England). 2020 May 6;24(1):201. PMID: 32375855. doi: 10.1186/s13054-020-02927-1.
2. Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. The New England journal of medicine. 2019 Apr 4;380(14):1347-58. PMID: 30943338. doi: 10.1056/NEJMra1814259.
3. Bohr A, Memarzadeh K. The rise of artificial intelligence in healthcare applications. Artificial Intelligence in Healthcare. 2020:25-60. PMID: PMC7325854. doi: 10.1016/B978-0-12-818438-7.00002-2.
4. Waring J, Lindvall C, Umeton R. Automated machine learning: Review of the state-of-the-art and opportunities for healthcare. Artificial intelligence in medicine. 2020 Apr;104:101822. PMID: 32499001. doi: 10.1016/j.artmed.2020.101822.
5. Serrano JI, Romero JP, Castillo MDD, Rocon E, Louis ED, Benito-León J. A data mining approach using cortical thickness for diagnosis and characterization of essential tremor. Sci Rep. 2017 May 19;7(1):2190. PMID: 28526878. doi: 10.1038/s41598-017-02122-3.
6. Benito-León J, Louis ED, Mato-Abad V, Sánchez-Ferro A, Romero JP, Matarazzo M, et al. A data mining approach for classification of orthostatic and essential tremor based on MRI-derived brain volume and cortical thickness. Ann Clin Transl Neurol. 2019 Dec;6(12):2531-43. PMID: 31769622. doi: 10.1002/acn3.50947.
7. Mato-Abad V, Labiano-Fontcuberta A, Rodríguez-Yáñez S, García-Vázquez R, Munteanu CR, Andrade-Garda J, et al. Classification of radiologically isolated syndrome and clinically isolated syndrome with machine-learning techniques. European journal of neurology. 2019 Jul;26(7):1000-5. PMID: 30714276. doi: 10.1111/ene.13923.
8. Michalski RS, Bratko I, Bratko A. Machine Learning and Data Mining; Methods and Applications: John Wiley & Sons, Inc.; 1998. ISBN: 0471971995.
9. Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. Journal of critical care. 2018 Jun;45:156-62. PMID: 29486341. doi: 10.1016/j.jcrc.2018.02.010.
10. Alsuliman T, Humaidan D, Sliman L. Machine learning and artificial intelligence in the service of medicine: Necessity or potentiality? Current research in translational medicine. 2020 Feb 3. PMID: 32029403. doi: 10.1016/j.retram.2020.01.002.
11. Lopez C, Tucker S, Salameh T, Tucker C. An unsupervised machine learning method for discovering patient clusters based on genetic signatures. Journal of Biomedical Informatics. 2018 2018/09/01/;85:30-9. doi: https://doi.org/10.1016/j.jbi.2018.07.004.
12. Shang Y, Liu T, Wei Y, Li J, Shao L, Liu M, et al. Scoring systems for predicting mortality for severe patients with COVID-19. EClinicalMedicine. 2020 Jul;24:100426. PMID: 32766541. doi: 10.1016/j.eclinm.2020.100426.
13. Ioannidis JPA, Axfors C, Contopoulos-Ioannidis DG. Population-level COVID-19 mortality risk for non-elderly individuals overall and for non-elderly individuals without underlying diseases in pandemic epicenters. Environmental research. 2020 Sep;188:109890. PMID: 32846654. doi: 10.1016/j.envres.2020.109890.
14. Fu J, Huang PP, Zhang S, Yao QD, Han R, Liu HF, et al. The value of serum

amyloid A for predicting the severity and recovery of COVID-19. Experimental and therapeutic medicine. 2020 Oct;20(4):3571-7. PMID: 32855710. doi: 10.3892/etm.2020.9114.
15. Hou H, Zhang B, Huang H, Luo Y, Wu S, Tang G, et al. Using IL-2R/lymphocytes for predicting the clinical progression of patients with COVID-19. Clinical and experimental immunology. 2020 Jul;201(1):76-84. PMID: 32365221. doi: 10.1111/cei.13450.
16. Zeng Z, Yu H, Chen H, Qi W, Chen L, Chen G, et al. Longitudinal changes of inflammatory parameters and their correlation with disease severity and outcomes in patients with COVID-19 from Wuhan, China. Critical care (London, England). 2020 Aug 27;24(1):525. PMID: 32854750. doi: 10.1186/s13054-020-03255-0.
17. Zhang C, Qin L, Li K, Wang Q, Zhao Y, Xu B, et al. A Novel Scoring System for Prediction of Disease Severity in COVID-19. Frontiers in cellular and infection microbiology. 2020;10:318. PMID: 32582575. doi: 10.3389/fcimb.2020.00318.
18. Bertsimas D, Lukin G, Mingardi L, Nohadani O, Orfanoudaki A, Stellato B, et al. COVID-19 mortality risk assessment: An international multi-center study. PloS one. 2020;15(12):e0243262. PMID: 33296405. doi: 10.1371/journal.pone.0243262.
19. Ye W, Lu W, Tang Y, Chen G, Li X, Ji C, et al. Identification of COVID-19 Clinical Phenotypes by Principal Component Analysis-Based Cluster Analysis. Front Med (Lausanne). 2020;7:570614. PMID: 33282887. doi: 10.3389/fmed.2020.570614.
20. England NHS. COVID-19 Decision support tool. 2020 [cited 2021 March 22]; Available from: http://prod-upp-image-read.ft.com/765d3430-7a57-11ea-af44-daa3def9ae03. .
21. guideline N. COVID-19 rapid guideline: critical care in adults. 2021 [cited 2021 March, 22]; Available from: https://www.nice.org.uk/guidance/ng159COVID-19 rapid guideline: critical care in adults.
22. Ahadi S, Zhou W, Schüssler-Fiorenza Rose SM, Sailani MR, Contrepois K, Avina M, et al. Personal aging markers and ageotypes revealed by deep longitudinal profiling. Nat Med. 2020 Jan;26(1):83-90. PMID: 31932806. doi: 10.1038/s41591-019-0719-5.
23. Marengoni A, Zucchelli A, Vetrano DL, Armellini A, Botteri E, Nicosia F, et al. Beyond Chronological Age: Frailty and Multimorbidity Predict In-Hospital Mortality in Patients With Coronavirus Disease 2019. J Gerontol A Biol Sci Med Sci. 2021 Feb 25;76(3):e38-e45. PMID: 33216846. doi: 10.1093/gerona/glaa291.
24. Gómez-Moreno C, Hernández-Ruiz V, Hernández-Gilsoul T, Avila-Funes JA. CLINICAL DECISION MAKING IN OLDER ADULTS WITH COVID-19 IN DEVELOPING COUNTRIES: LOOKING BEYOND CHRONOLOGICAL AGE. Rev Invest Clin. 2020;72(3):127-34. PMID: 32584321. doi: 10.24875/ric.20000131.
25. Ehni HJ, Wahl HW. Six Propositions against Ageism in the COVID-19 Pandemic. J Aging Soc Policy. 2020 Jul-Oct;32(4-5):515-25. PMID: 32491963. doi: 10.1080/08959420.2020.1770032.
26. Perrotta F, Corbi G, Mazzeo G, Boccia M, Aronne L, D'Agnano V, et al. COVID-19 and the elderly: insights into pathogenesis and clinical decision-making. Aging Clin Exp Res. 2020 Aug;32(8):1599-608. PMID: 32557332. doi: 10.1007/s40520-020-01631-y.
27. Juul S, Nielsen EE, Feinberg J, Siddiqui F, Jørgensen CK, Barot E, et al. Interventions for treatment of COVID-19: Second edition of a living systematic review with meta-analyses and trial sequential analyses (The LIVING Project). PloS one. 2021;16(3):e0248132. PMID: 33705495. doi: 10.1371/journal.pone.0248132.
28. Rahman A, Jahan Y. Defining a ‘Risk Group’ and Ageism in the Era of COVID-

19. Journal of Loss and Trauma. 2020 2020/11/16;25(8):631-4. doi: 10.1080/15325024.2020.1757993.
29. Cirillo D, Catuara-Solarz S, Morey C, Guney E, Subirats L, Mellino S, et al. Sex and gender differences and biases in artificial intelligence for biomedicine and healthcare. npj Digital Medicine. 2020 2020/06/01;3(1):81. doi: 10.1038/s41746-020-0288-5.
30. Chang ES, Kannoth S, Levy S, Wang SY, Lee JE, Levy BR. Global reach of ageism on older persons' health: A systematic review. PloS one. 2020;15(1):e0220857. PMID: 31940338. doi: 10.1371/journal.pone.0220857.
31. HM Hospitales (2020) Dataset “Covid Data Saves Lives”. [cited 2020 May 14]; Available from: https://www.hmhospitales.com/coronavirus/covid-data-save-lives/english-version.
32. Pelleg D, Moore AW. X-means: Extending k-means with efficient estimation of the number of clusters. In ICML2000. p. 727--34.
33. Cha S, editor. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions. 2007.
34. Shirkhorshidi AS, Aghabozorgi S, Wah TY. A Comparison Study on Similarity and Dissimilarity Measures in Clustering Continuous Data. PloS one. 2015;10(12):e0144059. PMID: 26658987. doi: 10.1371/journal.pone.0144059.
35. Halkidi M, Batistakis Y, Vazirgiannis M. On Clustering Validation Techniques. Journal of Intelligent Information Systems. 2001 2001/12/01;17(2):107-45. doi: 10.1023/A:1012801612483.
36. Vacha-Haase T, Thompson B. How to Estimate and Interpret Various Effect Sizes. Journal of Counseling Psychology. 2004;51(4):473-81. doi: 10.1037/0022-0167.51.4.473.
37. Yang X, Yu Y, Xu J, Shu H, Xia J, Liu H, et al. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. The Lancet Respiratory medicine. 2020 May;8(5):475-81. PMID: 32105632. doi: 10.1016/s2213-2600(20)30079-5.
38. Roy D, Ghosh R, Dubey S, Dubey MJ, Benito-León J, Kanti Ray B. Neurological and Neuropsychiatric Impacts of COVID-19 Pandemic. Can J Neurol Sci. 2020 Aug 5:1-66. PMID: 32753076. doi: 10.1017/cjn.2020.173.
39. Lei F, Liu YM, Zhou F, Qin JJ, Zhang P, Zhu L, et al. Longitudinal Association Between Markers of Liver Injury and Mortality in COVID-19 in China. Hepatology (Baltimore, Md). 2020 Aug;72(2):389-98. PMID: 32359177. doi: 10.1002/hep.31301.
40. Dong X, Sun L, Li Y. Prognostic value of lactate dehydrogenase for in-hospital mortality in severe and critically ill patients with COVID-19. International journal of medical sciences. 2020;17(14):2225-31. PMID: 32922185. doi: 10.7150/ijms.47604.
41. Tan C, Huang Y, Shi F, Tan K, Ma Q, Chen Y, et al. C-reactive protein correlates with computed tomographic findings and predicts severe COVID-19 early. Journal of medical virology. 2020 Jul;92(7):856-62. PMID: 32281668. doi: 10.1002/jmv.25871.
42. Henry BM, de Oliveira MHS, Benoit S, Plebani M, Lippi G. Hematologic, biochemical and immune biomarker abnormalities associated with severe illness and mortality in coronavirus disease 2019 (COVID-19): a meta-analysis. Clinical chemistry and laboratory medicine. 2020 Jun 25;58(7):1021-8. PMID: 32286245. doi: 10.1515/cclm-2020-0369.
43. Poggiali E, Zaino D, Immovilli P, Rovero L, Losi G, Dacrema A, et al. Lactate dehydrogenase and C-reactive protein as predictors of respiratory failure in CoVID-19 patients. Clinica chimica acta; international journal of clinical chemistry. 2020 2020/10//;509:135-8. PMID: 32531257. doi: 10.1016/j.cca.2020.06.012.

44. Pourbagheri-Sigaroodi A, Bashash D, Fateh F, Abolghasemi H. Laboratory findings in COVID-19 diagnosis and prognosis. Clinica Chimica Acta. 2020 2020/11/01/;510:475-82. doi: https://doi.org/10.1016/j.cca.2020.08.019.
45. Soraya GV, Ulhaq ZS. Crucial laboratory parameters in COVID-19 diagnosis and prognosis: An updated meta-analysis. Med Clin (Engl Ed). 2020 Aug 28;155(4):143-51. PMID: 32864456. doi: 10.1016/j.medcle.2020.05.004.
46. Wynants L, Van Calster B, Collins GS, Riley RD, Heinze G, Schuit E, et al. Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. BMJ. 2020;369:m1328. doi: 10.1136/bmj.m1328.

Figure 1. Hospital stay, outcome measures and laboratory tests that showed statistically significant differences among clusters with a medium or high effect size ($\eta 2 > 0.06$). Note that some variables are scaled (transformation between brackets) for the sake of graph legibility.

Supplementary Files

Figures

Hospital stay, outcome measures and laboratory tests that showed statistically significant differences among clusters with a medium or high effect size (?2 > 0.06). Note that some variables are scaled (transformation between brackets) for the sake of graph legibility.