 doi:10.20944/preprints202209.0291.v1

Article

Multidimensional Machine Learning on 2173 COVID-19 Patients in Vietnam

Tue Trong Nguyen PhD.[4,5*], Tu Cam Ho MSc.[1,2*], Huong Thi Thu Bui MD. PhD. [6*], Ulrike Protzer MD.PhD. [2,3] and Van Thanh Ta MD.PhD. [1,4,5]

[1]: Center of Gene and Protein Research, Hanoi Medical University, Hanoi, Vietnam
[2]: Institute of Virology, School of Medicine, Technical University of Munich, Munich, Germany
[3]: Institute of Virology, School of Medicine, Technical University of Munich, Helmholtz Zentrum München, Munich, Germany; German Center for Infection Research (DZIF), Munich and Hamburg Partner sites, Germany
[4]: Medical Laboratory Department, Hanoi Medical University, Hanoi, Vietnam
[5]: Clinical Laboratory Department, Hanoi Medical University Hospital, Hanoi, Vietnam
[6]: Thai Nguyen University of Medicine and Pharmacy
* *Correspondence*: trongtue@hmu.edu.vn

Abstract:The purpose of the study was to determine (a) the overall preclinical character; (b) the cumulative cutoff values and the risk ratio, and (c) the factors associated with severity by a unidimensional and multidimensional analysis on 2173 Sars-Cov2 patients. **METHODS:** The machine learning study population consisted of 2173 patients (1587 mild and non symptoms patients, 377 moderate patients, 209 severe patients). The status of the patients was recorded from September 2021 to March 2022. **RESULTS**: The Covid19 Severity directly links with a significant correlation to Age, Score index of the chest X-ray, percentage and quantity of neutrophils, Albumin, C reactive protein, and ratio of Lymphocytes. Their important cut off values (from regression analysis) respectively are: 77.56 years old (the mild-moderate group), 5.53 (the mild-moderate group) and 10.51 (the moderate-severe group), 84.80% (the mild-moderate group) and 87.74%(the moderate-severe group), 11.77G/L (the moderate-severe group), 29.73g/L (the moderate-severe group), 7.46mg/dL (the mild-moderate group), 6.32% (the moderate-severe group). Their significant ($p<0.0001$) R score correlation with the severity of Covid19, are: 0.44, 0.52 and 0.52, 0.33 and 0.44, 0.42, -0.43, 0.40, -0.41. Their significant risk ratio ($p<0.00001$) from the meta-analysis, respectively are: 4.19 [3.58-4.95], 3.29 [2.76-3.92] and 3.03 [2.4023;3.8314], 3.18 [2.73-3.70] and 3.32 [2.6480;4.1529], 3.15 [2.6153;3.8025], 3.4[2.91-3.97], 0.46 [0.3650;0.5752] ($p<0.00001$), 0.34 [0.2743;0.4210]. The pair ALT – Leucocytes and Transferrin – Anion Chloride get the most important correlation shift. ALT – Leucocytes show the important negative link ($R=-1$, $p<0.00001$) in the mild group to the significant positive correlation in the moderate group ($R=1$, $p<0.00001$). Transferrin–anion Chloride has an important positive association ($R=1$, $p<0.00001$) in the mild group with a significant negative correlation in the moderate group ($R=-0.59$, $p<0.00001$). The network map and HCA show thain the mild-moderate group, the closest neighbors with the Covid19 severity are ferritins, Age. Then there is C-reactive protein, SI of X-ray, Albumin, and Lactate dehydrogenase, which are the next close neighbors of these three factors. In the moderate-severe group, the closest neighbors with the Covid19 severity are Ferritin, Fibrinogen, Albumin, the quantity of Lymphocytes, SI of X-ray, white blood cells count, Lactate dehydrogenase, and quantity of neutrophils. **CONCLUSIONS:** Complete multidimensional study in 2173 Covid19 patients in Vietnam shows the whole picture of all the preclinical factors, which may become the clinical reference marker for surveillance and diagnostic management.

Keywords: COVID-19; Multidimensional Analysis; HCA; Hierarchical cluster analysis; regression analysis; mild; moderate; severe; Age; Score index of the chest X-ray; percentage and quantity of neutrophils; Albumin; C reactive protein; ratio of Lymphocytes

 doi:10.20944/preprints202209.0291.v1

1. Introduction

The COVID-19 pandemic has become one of the most severe health crises in human history, spreading rapidly across the globe from January 2020 to the present. The outbreak has been crossing over 200 countries in the world. Vietnam has the first sixteen typical cases confirmed positive, updated to Feb 28th, 2020. After thoroughly applying medical prevention and active control, Vietnam can take control of the outbreak of COVID-19 as a recent WHO assessment. Vietnam has been reported as an influential country in preventing and controlling the outbreak of COVID-19.[1] We performed a multi-dimension study of 2173 hospitalized Covid-19 patients from September 2021 to March 2022 at the Hospital of Hanoi Medical University in Vietnam. All of them have the PCR Ct of SARS-CoV < 30 cycles. We take all the medical records at our study's first diagnosis of Covid19. All information in the digital medical records was collected. Machine Learning (ML) provides methods, techniques, and tools to help solve diagnostic and prognostic problems in various medical domains. ML is used to analyze the importance of clinical parameters and their combinations for prognosis, e.g., prediction of disease progression, extraction of medical knowledge for outcome research, therapy planning and support, and overall patient management. ML is also essential for data analysis, such as detecting regularities in the data by appropriately dealing with imperfect data, interpreting continuous data used in the Intensive Care Unit, and intelligent alarming, resulting in effective and efficient monitoring. Medical diagnostic reasoning is a critical application area of intelligent systems.[2] ML techniques are based on algorithms – sets of mathematical procedures which describe the relationships between variables. Our study is a combination of unidimensional and multidimensional analysis, using algorithms including regularized General Linear Model regression (GLMs), classing data following Support Vector Machines (SVMs) or the hierarchical analysis (HCA), and Artificial Networks Contribution.[3]

2. Patients and Methods

2.1. Study Population

Following the Living guidance for clinical management of COVID-19 from WHO (23 November 2021) and the latest guidance on diagnosis and treatment of COVID-19, the Ministry of Health (article number 250/QD-BYT), the asymptomatic group F0 was added to the classification of COVID-19 disease severity, in addition to the four groups of mild, moderate, severe and critical already present. In our study, there are 1098 mild patients, 386 moderate patients, and 203 severe patients.

The mild level is Covid19 patients who have non-specific clinical symptoms such as fever, dry cough, sore throat, nasal congestion, fatigue, headache, muscle pain, loss of taste, smell, diarrhea, etc. Breathing rate < 20 breaths/minute, SpO_2 > 96% when breathing air. Besides, the patient is awake and self-service; Chest X-ray is normal or present with minimal damage.

The moderate level is when assessing the overall condition. The patient has non-specific clinical symptoms such as mild severity. Regarding respiration, the patient showed signs of pneumonia, shortness of breath, rapid breathing 20-25 times/minute, lung crackles, and no signs of severe respiratory failure, SpO_2 94-96% when breathing room air. F0 may have difficulty breathing on exertion (walking indoors, climbing stairs). The patient's pulse may be fast or slow, dry skin, tachycardia, normal blood pressure, and alert consciousness. In addition, chest X-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

ray and chest computed tomography detected lesions, lesions less than 50%. Ultrasound shows B wave, arterial blood gas: $PaO2/FiO_2 > 300$.

The severity case has respiratory signs of pneumonia accompanied by any of the following: respiratory rate > 25 breaths/min; severe shortness of breath, contraction of accessory respiratory muscles; $SpO_2 < 94\%$ when breathing room air. Regarding circulation, the patient's heart rate can be fast or slow, and normal blood pressure or increased. Neurologically, the patient may be restless or lethargic and tired. Chest X-ray and computed tomography: there are lesions, and lesions are more than 50%. Ultrasound shows many B waves and arterial blood gas: PaO_2/FiO_2 about 200-300.

2.2. Statistic and multidimensional analysis

The data collection, storage tools, and analysis of this study were conducted in R.4.1.0 environment. The groups were compared using the independent sample t-test, Pearson's chi-square test, and Fisher's exact test to clarify the differences. The cutoff values, risk ratio (RR), odds ratio (OR), and 95% confidence interval (CI) were calculated by the regression analysis and the meta-analytical methods. The significant factor will be selected for the HCA and Network study in R.4.1.0 environment, which is performed based on the correlation test results. K-means was optimized following 20 methods. (Figure 1)

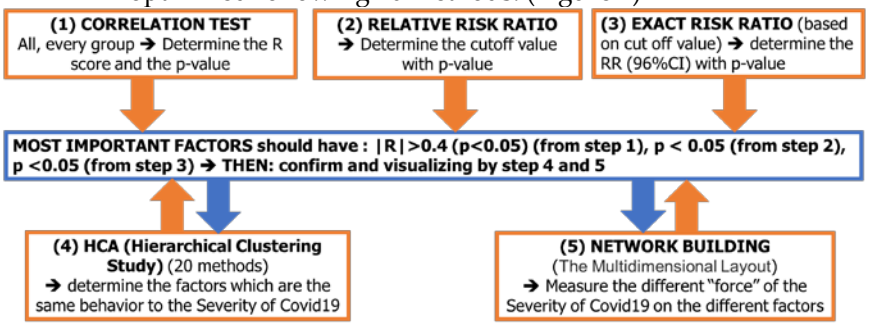


Figure 1: Multidimensional machine learning approach

3. Results

3.1. Overview of cohort

2173 COVID-19 patients in our study, there are 1587 mild and non symptom patients, 377 moderate patients, and 209 severe patients. Taking all collected information, we found significantly different in several factors between the severity groups.

All information about the patient's medical history and cause of hospitalization was collected and analyzed. Medical history was recorded and classified according to ICD 10 list of the Ministry of Health of Vietnam and WHO. (http://icd.kcb.vn/)

The ANOVA test showed a statistically significant difference between the distribution of medical history and the level of Covid19. No correlation was found between the level of Covid 19 and the patient's medical history.

Table 1. Patient's medical history from 2173 Covid19 patients, 3 groups of patients: Severe, Moderate and Mild.

Diagnostic Report	Number of patients	Number patients in the mild and non symptom group	Number patients in the moderate group	Number patients in the severe group
Diabetes	315	155	107	53
Lung related disease	195	64	57	74

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

Shock	72	8	1	63
Cardio related disease	559	278	176	105
Hepatic related disease	93	58	25	10
Brain infarction related disease	87	35	36	16
Neural related disease	38	11	18	9
Renal related disease	168	82	42	44
Cancer	102	61	19	22
Bone cartilage related disease	56	29	20	7
Digestive systems related disease	107	145	18	7
Immune related disease	30	23	4	3

ICD-10-CM (http://icd.kcb.vn/)	**Number of patients**
W18 (Slipping, tripping, stumbling and falls)	1
W01 (Slipping, tripping, stumbling and falls)	1
Z80_Z89 (Family history of malignant neoplasm of digestive organs)	1
Z70_Z76 (persons encountering health services for examination and investigation)	1
Z40_Z54 (Persons encountering health services for specific procedures and health care)	9
V52 (Accidents)	1
T66_T78 (Other and unspecified effects of external causes)	1
T15_T19 (Effects of foreign body entering through natural orifice)	1
T00_T07 (Injuries involving multiple body regions)	1
U04 (Severe acute respiratory syndrome, unspecified)	1
U07.1 (COVID-19, virus identified)	927
U07.2 (COVID-19, virus not identified)	4
Z00_Z13 (Persons encountering health services for examination and investigation)	1
S80_S89 (Injuries to the knee and lower leg)	2
S70_S79 (Injuries to the hip and thigh)	1
S50_S59 (Injuries to the elbow and forearm)	2
S40_S49 (Injuries to the shoulder and upper arm)	1
S30_S39 (Injuries to the abdomen, lower back, lumbar spine and pelvis)	3
S20_S29 (Injuries to the thorax)	1
S00_S09 (Injuries to the head)	1
R57.2 (Septic shock)	1
R57.1 (Hypovolaemic shock)	1
Q60_Q64 (Congenital malformations of the urinary system)	1
Q38_Q45 (Other congenital malformations of the digestive system)	1
Q30_Q34 (Congenital malformations of the respiratory system)	1
N80_N89 (Noninflammatory disorders of female genital tract)	1
N40_N51 (Diseases of male genital organs)	1
N30_N39 (Other diseases of urinary system)	1
N17_N19 (Renal failure)	1
M80_M94 (Osteopathies and chondropathies)	3
M60_M79 (Soft tissue disorders)	2
M40_M54 (Dorsopathies)	11

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

M00_M25 (Arthropathies)	4
K90_K93 (Other diseases of the digestive system)	1
K80_K87 (Disorders of gallbladder, biliary tract and pancreas)	2
K70_K77 (Diseases of liver)	1
K65_K67 (Diseases of peritoneum)	1
K55_K64 (Other diseases of intestines)	8
K50_K52 (Noninfective enteritis and colitis)	1
K40_K46 (Hernia)	1
K35_K38 (Diseases of appendix)	1
K20_K31 (Diseases of oesophagus, stomach and duodenum)	1
I10_I15 (Hypertensive diseases)	7
J90_J94 (Other diseases of pleura)	1
J80_J84 (Other respiratory diseases principally affecting the interstitium)	1
J09_J18 (Influenza and pneumonia)	1
G90_G99 (Other disorders of the nervous system)	1
G80_G83 (Cerebral palsy and other paralytic syndromes)	1
F40_F48 (Neurotic, stress-related and somatoform disorders)	1
F00_F09 (Organic, including symptomatic, mental disorders)	1
E70_E90 (Metabolic disorders)	1
E20_E35 (Disorders of other endocrine glands)	1
E10_E14 (Diabetes mellitus)	4
D65_D69 (Coagulation defects, purpura and other hemorrhagic conditions)	2
D37_D48 (Neoplasms of uncertain or unknown behavior)	1
D10_D36 (Benign neoplasms)	1
D00_D09 (In situ neoplasms)	1
C00_C97 (Neoplasms)	4
B35_B49 (Mycoses)	1
B25_B34 (Other viral diseases)	11
B15_B19 (Viral hepatitis)	1
A30_A49 (Other bacterial diseases)	1

Classification of Covid19 patients is based on the guidelines of the Ministry of Health, in which the SPO_2 index is used as the main criterion. A sudden decrease in SPO_2 is believed to be the hallmark of a worsening patient. The parameters of the Covid19 level were given by our colleagues, the doctors at the Field Hospital - Hanoi Medical University Hospital to have a suitable treatment regimen for each patient.

66 subclinical indicators are separated for better analysis. There is a significant dissimilarity between the age of mild patients with moderate and severe patients ($p<0.00001$), the same result appears in AST, ALT, Hemoglobin, Mean Corpuscular hemoglobin, Quantity of Monocytes, and Covid19 vaccine-boosted times. Several factors have essential differences when we compare every two groups of the severity (Mild vs. Moderate, Moderate vs. Severe, Mild vs. Severe) as Urea, pH, SI of X-ray, Fibrinogen, percentage, and quantity of Monocytes, the number of lymphocytes, percentage of basophils, RBC, WBC, the ratio of lymphocytes, quantity and percentage of neutrophils, anion Cl, pCO_2, SPO_2 D-dimer, Glucose, Albumin, Transferrin, Ferritin, CRP, Pro-BNP, Troponin-T, Creatinine, LDH. Some other factors show the significant difference between the non-severe group (mild and moderate) with the severe group, for instant: the number of basophils,

 doi:10.20944/preprints202209.0291.v1

Beta-adrenergic blockers, the ratio of prothrombin, time and ratio of aPTT, Leucocytes, and Specific gravity. Total bilirubin shows only a significant difference between the mild and moderate groups. (Table 2)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

		MILD (N=1587 patients)							MODERATE(N=377 patients)							SEVERE(N=209 patients)							p value of ANOVA		
		Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Mild vs Moderate	Moderate vs Severe	Mild vs Severe
1	SPO_2 (%)	7	97	98	97.2	98	100	608	80	95	96	95.77	98	100	64	35	87	94	89.84	97	100	16	7.2E-07	<2E-16	<2E-16
2	Age (years old)	1	27	43	44.72	62	108	10	8	59	72	68.48	83	99	3	17	62	73	70.20	83	103	2	2.85E-69	2.79E-01	1.62E-50
3	Gender(1= Male, 2 = Female)	1	1	2	1.52	2	2	11	1	1	2	1.51	2	2	3	1	1	1	1.45	2	2	2	5.54E-01	1.96E-01	5.16E-02
4	Severity of Covid19 (1= Mild, 2=Moderate, 3=Severe)	1	1	1	1	1	1	1	2	2	2	2	2	2	2	3	3	3	3	3	3	3	NA	NA	NA
5	Covid19 vaccine (0 = no. 1=1 boost, 2= 2 boosts, 3= 3 boosts)	0	0	2	1.58	3	3	971	0	0	0	0.90	2	3	278	0	0	0	0.85	2	3	155	2.63E-05	8.03E-01	2.80E-04
6	Urobilinogen (µmol/L)	3	3	3	3	3	3	1074	3	3	3	3	3	3	363	1	3	3	2.90	3	3	161	NA	NA	7.13E-01
7	Unrinary Creatinine (umol/L)	676	3761.50	6860	5963	9061.50	9456	1089	NA	NA	NA	NaN	NA	NA	377	7818	7818	7818	7818	7818	7818	208	NA	NA	NA
8	Urea (mmol/L)	1.30	3.80	4.60	5.31	5.80	34.70	316	1.60	4.50	5.70	6.94	7.88	39.60	23	2	6.50	9.40	11.34	13.95	45.80	2	5.38E-13	1.02E-18	4.95E-63
9	Urinary Glucose (mmol/L)	3	5.88	15.50	26.69	56	56	1085	5.50	5.50	6	21.04	35	56	365	3	5.50	14	20.48	28	56	153	6E-01	9.30E-01	4.17E-01
10	Cetone (mmol/L)	1.50	1.50	1.50	6	8.25	15	1090	0.50	1.25	1.50	3.38	2.38	15	369	0.50	0.50	1.50	3.13	1.50	15	179	5.12E-01	8.94E-01	3.25E-01
11	Leucocytes (LEU/uL)	15	100	125	242.69	500	500	1080	15	15	85	201.39	500	500	359	1	25	70	101.1	100	500	161	6.07E-01	3.33E-02	6.44E-03
12	pH	5.50	7	7.41	7.14	7.44	8	1001	5	7.39	7.42	7.34	7.45	8	134	5	7.09	7.35	6.91	7.43	7.50	15	1.65E-04	1.18E-11	2E-02
13	Specific Gravity	1005	1012	1015	1016.62	1020	1033	1048	1007	1013	1019	1019.65	1026	1042	340	1006	1016	1021	1021	1026	1047	98	7.92E-02	2.88E-01	8.66E-04
14	SI of Xray	1	2	3	3.55	4	15	705	1	4	7	6.98	9	15	74	2	10	12	11.08	13	16	24	1.50E-49	1.02E-34	2.51E-134
15	Ejection Fraction	32	45.75	57	53.70	60.75	69	1083	17	41.50	62	55.79	67	78	358	30	44	53	57.29	75	85	188	7.21E-01	7.75E-01	5.55E-01
16	Lactate (mmol/L)	1	1.40	1.90	2.39	2.40	17.80	1028	0.80	1.60	2	2.41	2.60	20	140	0.60	1.60	2.30	2.75	3	20	18	9.47E-01	8.44E-02	2.66E-01
17	Fibrinogen (g/L)	0.74	2.76	3.26	3.35	3.81	8.45	391	0.87	3.12	3.88	4.08	4.72	9.63	63	0.64	3.32	4.34	4.46	5.67	9.46	21	1.67E-20	8.69E-03	2.33E-29

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

18	Percentage of Monocytes (%)	0	5.40	7.40	7.77	9.60	60	252	0	3.20	5.30	6.02	7.90	21.40	9	0	NULL	NULL	0	NULL	NULL	0	3.53E-13	2.20E-16	9.46E-46
19	Quantity of Monocytes (G/L)	0.02	0.33	0.45	0.50	0.61	1.66	265	0	0.23	0.36	0.46	0.61	3.06	11	0	NULL	NULL	0	NULL	NULL	0	4.01E-02	2.63E-01	4.64E-04
20	Mean Corpuscular volume (fL)	63.40	87.90	91.10	90.43	94.40	119.7	253	59.10	88.48	91.70	90.78	95.20	126.60	9	65.80	89.50	92.70	91.93	96.43	121.90	3	4.37E-01	1.04E-01	NA
21	Mean Corpuscular hemoglobin (pg/cell)	18	28	30	38.95	31	344	253	18	29	30	32.63	31	329	9	0	NULL	NULL	0	NULL	NULL	0	3.11E-02	4.84E-01	3.14E-02
22	Mean Corpuscular hemoglobin concentration (g/dL)	29	317	325	323.42	334	360	253	30	317.8	324	323.23	332	362	9	0	NULL	NULL	0	NULL	NULL	0	8.93E-01	3.31E-01	3.15E-01

Table 2: 60 subclinical factors from 2173 Covid19 patients, 3 groups of patients: Severe, Moderate and Mild. (cont)

		MILD (N=1587 patients)							MODERATE(N=377 patients)							SEVERE(N=209 patients)							p value of ANOVA		
		Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Mild vs Moderate	Moderate vs Severe	Mild vs Severe
23	Quantity of Lymphocytes (G/L)	0.09	0.89	1.25	1.36	1.71	7.43	266	0.09	0.58	0.89	1	1.29	3.80	10	0.11	0.32	0.56	0.70	0.82	3.49	1	4.01E-19	1.68E-09	5.59E-36
24	Percentage of Eosinophils (%)	0.20	0.50	0.70	0.88	1.08	1.90	1089	0	0.08	0.10	1.03	1.05	3.90	373	0	0	0	0	0	0	208	8.88E-01	6.65E-01	3.61E-01
25	Quantity of Basophils (G/L)	0	0.02	0.04	0.05	0.06	1.03	258	0	0.02	0.03	0.05	0.05	0.80	9	0	0.02	0.04	0.07	0.07	2.60	1	5.49E-01	2.51E-02	3.71E-03
26	Percentage of Basophils (%)	0	0.40	0.60	0.75	0.90	10.90	267	0	0.30	0.40	0.60	0.70	6.90	11	0	0.20	0.30	0.48	0.60	3.30	1	5.13E-04	1.16E-02	1.79E-07
27	Red blood cell count	2.19	4.32	4.71	4.70	5.12	6.92	268	2.25	4.07	4.53	4.51	5	6.73	11	1.44	3.84	4.40	4.35	4.90	6.79	1	1.73E-05	1.23E-02	2.97E-10
28	White blood cell count (G/L)	0.34	4.92	6.37	6.77	7.93	29.51	263	0.15	5.46	7.38	9.03	9.72	306	12	2.14	8.28	12.50	13.22	16.91	40.43	4	1.68E-04	6.69E-04	3.55E-78
29	Hemoglobin (T/L)	100	129	139	139.21	150	191	286	100	121.8	133	133.93	146	190	25	0	NULL	NULL	0	NULL	NULL	0	5.95E-07	1.12E-01	5.99E-09
30	Ratio of Lymphocytes (%)	1	14	21	22.32	28.60	77.90	268	1	7.68	12.05	14.81	19.53	66	9	0.60	2.55	5.40	6.66	8.70	34	2	1.10E-26	6.99E-25	1.34E-71

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

31	Quantity of Neutrophils (G/L)	0	2.80	4.10	4.63	5.60	28.70	264	0.20	3.80	5.55	6.62	8.23	55.50	13	1.80	7.20	11.25	11.79	15.53	38.10	1	2.99E-18	3.52E-24	2.59E-103
32	Percentage of Neutrophils (%)	0.80	57.10	66.30	65.34	75.60	97.20	256	2.60	67.53	79.55	75.99	86.10	95.90	11	47.70	84.58	89.50	87.97	93.23	98.20	1	4.47E-31	3.65E-29	1.50E-86
33	Hematocrit (L/L)	0.19	0.39	0.43	0.62	0.46	48	255	0.17	0.37	0.41	0.41	0.45	0.63	9	0.14	0.36	0.41	0.40	0.45	0.59	1	1.52E-01	7.01E-02	2.60E-01
34	Beta adrenergic blockers (mmol/L)	33.10	42.10	43.85	43.53	45.63	51.50	1029	26.90	42.40	44.70	44.48	46.53	56.10	153	3	39.15	42.70	41.83	44.95	56.80	2	6.71E-02	7.07E-08	3.53E-02
35	anion HCO3 (mmol/L)	6.20	20.05	22.10	21.37	23.10	30.10	1026	13.70	20.90	22.95	22.95	24.90	36.10	139	6.10	18.68	21.50	21.18	23.63	42.20	9	6.49E-04	9.15E-06	7.76E-01
36	anion Cl (mmol/L)	74	97	99	98.34	101	107	263	73	93	96	95.84	99	126	8	76	95	99	99.45	104	132	1	1.41E-16	3.77E-09	3.69E-03
37	ion K (mmol/L)	2.50	3.60	3.80	3.86	4.10	6.40	264	2.20	3.60	3.90	3.94	4.30	6.10	10	2.30	3.70	4	4.58	4.40	100	9	3.97E-03	7.40E-02	2.58E-03
38	ion Na (mmol/L)	114	134	137	135.80	138	145	264	110	130	133	133.30	137	165	9	106	132	135	135.7	138	169	6	2.67E-17	3.05E-05	8.41E-01
39	pCO2 (mmHg)	16.10	30.75	34.10	33.64	37.55	47	1026	23.10	31.95	35.30	36.46	39.10	79.60	138	15.80	30.88	36.90	40.71	45.43	122.70	5	4.05E-03	9.56E-05	1.47E-04
40	FiO_2 (Fraction of inspired oxygen)	0.21	0.21	0.21	0.21	0.21	0.21	1026	0	0.21	0.21	0.21	0.21	0.21	138	0.21	0.21	0.21	0.22	0.21	1	4	5.80E-01	1.03E-01	4.44E-01
41	INR (International normalized ratio)	0.89	1.17	1.35	1.45	1.60	4.36	967	0.97	1.21	1.36	1.50	1.59	4.85	260	0.93	1.27	1.42	1.56	1.64	7.37	29	4.24E-01	3.89E-01	9.15E-02
42	Time of Prothrombin (s)	11.90	14.90	16.60	17.48	19	41.80	966	12.80	15.30	16.8	18	19.05	44.90	259	12.50	15.80	17.30	18.44	19.33	62.10	29	3.41E-01	4.69E-01	8.39E-02
43	Ratio of Prothrombin (%)	17	52	65	65.22	78.50	120	966	14.90	51	62.5	66.79	74	633	259	12	51	60	60.63	70	112	29	7.64E-01	1.63E-01	2.52E-02
44	D-dimer (ng/mL)	100	194	310	777.29	639.50	27100	303	20	329.75	585	1223.35	1157.50	26700	23	100	903.50	1550	4149	3960	36400	10	5.57E-04	2.05E-13	3.17E-35

Table 2: 60 subclinical factors from 2173 Covid19 patients, 3 groups of patients: Severe, Moderate and Mild. (cont)

		MILD (N=1587 patients)							MODERATE(N=377 patients)							SEVERE(N=209 patients)							p value of ANOVA		
		Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Min.	1stQu.	Median	Mean	3rdQu	Max.	no information	Mild vs Moderate	Moderate vs Severe	Mild vs Severe
45	Time of aPTT (s)	24.30	29.85	32.20	34.06	35.20	100	999	21.80	28.83	32.1	34.90	37.53	189.70	283	20	29.75	34.10	40.15	42.25	123.30	30	6.88E-01	2.51E-02	3.78E-03
46	Ratio of aPTT (Disease/control)	0.90	1.10	1.10	1.17	1.20	1.90	999	0.80	1.03	1.10	1.25	1.30	6.70	283	0.80	1.08	1.20	1.32	1.50	4.70	33	2.65E-01	2.85E-01	4.04E-03

 doi:10.20944/preprints202209.0291.v1

47	Total Protein	NA	NA	NA	NA	NA	NA	1093	54.80	54.80	54.8	54.80	54.80	54.80	376	48.50	48.50	48.50	48.50	48.50	48.50	208	NA	NaN	NA
48	Protein of pleural fluid (g/L)	11.40	11.50	23.60	22.18	23.60	40.80	1088	21.50	26.28	31.1	31.05	35.83	40.60	375	2	19.70	27.50	25.53	32.78	40.70	201	4.30E-01	5.84E-01	6.36E-01
49	Protein of cerebrospinal fluid	0.33	0.41	0.49	0.52	0.61	0.73	1090	0.24	0.99	1.73	1.73	2.48	3.23	374	0.81	0.81	0.81	0.81	0.81	0.81	208	2.35E-01	6.46E-01	3.34E-01
50	Glucose of cerebrospinal fluid	3.72	3.72	3.72	3.72	3.72	3.72	1092	2.41	4.34	5.16	4.86	5.69	6.72	373	6.17	6.17	6.17	6.17	6.17	6.17	208	6.10E-01	5.62E-01	NaN
51	Glucose (mmol/L)	2.90	5.30	6.10	6.89	7.40	29.80	340	3.50	5.95	7.30	9.34	10.30	46.70	42	0.60	6.90	10.15	12.74	15.03	89.50	45	4.10E-21	1.52E-06	3.10E-39
52	Albumin (g/L)	18	31.58	36.50	35.92	40.90	49.30	937	18.10	29.20	32.9	32.72	35.70	45.70	180	7.20	24.10	27.80	27.68	31	41.10	32	3.85E-07	1.18E-18	1.41E-28
53	Transferrin (mg/dL)	230	255.50	281	312.67	354	427	1090	66	141	155	159.87	181.5	265	362	64	89	112	118.6	141	239	178	7.93E-04	3.58E-03	5.87E-08
54	Pro Calcitonin (ng/mL)	0.05	0.24	2.75	16.83	10.72	100	1085	0.03	0.10	0.58	41.66	8.20	418	361	0.06	0.50	1.63	29.03	7	616	114	5.33E-01	6.39E-01	7.28E-01
55	Ferritin (ng/mL)	7.42	144.20	338.17	466.19	626.21	2000	304	9.87	427.7	919.7	992.66	1676	2000	39	97.80	1088	1676	1422	1720.75	2000	25	2.22E-52	1.33E-14	2.98E-111
56	C reactive protein (mg/dL)	0.07	0.30	0.70	1.96	1.94	21.50	333	0.10	1.58	4.20	6.47	8.90	41.10	37	0.10	3.60	8.70	10.66	16.10	38.60	56	6.47E-44	1.43E-08	4.28E-83
57	Pro b-type natriuretic peptide (pro-BNP, pg/ml)	5	35.01	93.20	385.55	428.85	4983	977	5	102.1	348.9	1330.09	913.7	35000	195	12.70	413	926	3836	2891	35000	64	1.03E-02	9.49E-05	9.28E-07
58	Troponin-T (ng/L)	3	7	10.70	21.45	18	423.7	956	3.70	10.35	15.9	36.88	28.13	760.70	165	4.30	17.25	37.05	146.4	90.25	3253	43	3.32E-02	3.14E-05	9.97E-05
59	Creatinine (μmol/L)	19	56	69	76.88	87	954	252	18	58	74	84.11	93	852	13	24	62	85	107.8	124	608	12	2.83E-02	1.76E-04	8.26E-13
60	Phosphor (mmol/L)	0.70	0.90	1	1.17	1.28	2.50	1079	0.40	0.80	1	1.03	1.20	2.30	332	0.30	0.80	1	1.07	1.23	3.40	77	2.27E-01	6.64E-01	4.36E-01
61	Creatin kinase (U/L 37°C)	48	105	184	1217.23	792	9764	1080	32	84	102	345.93	197	2800	362	20	102.25	249	716.7	739.25	5294	129	2.31E-01	1.95E-01	2.27E-01
62	Lactate Dehydrogenase (U/L 37°C)	93	177	202	232.12	244	5548	348	144	229	311	352.40	409	3146	36	145	386.50	540.5	579.2	665.25	3518	29	9.48E-17	9.80E-19	1.35E-57
63	Total Bilirubin (μmol/L)	5.10	7.10	8.40	26.27	9.70	136.8	1086	4	5	7.30	43.90	9.20	194	372	4	8.50	15.50	20.61	23.40	61.20	188	6.54E-01	2.31E-01	6.51E-01
64	Direct bilirubin (μmol/L)	3.20	3.40	5.30	31.38	14.40	130.6	1088	2.70	2.85	3	62.80	92.85	182.70	374	2.40	4.20	6.80	11.28	11.45	50.30	186	5.88E-01	1.56E-02	1.10E-01
65	AST (U/L 37°C)	14	26	44	108.25	130.50	1570	1014	20	101.75	128	243.91	257.50	3109	323	16	104	126	232	196	2376	122	1.39E-02	8.62E-01	8.59E-03
66	ALT (U/L 37°C)	11	19	43	81.92	132	441	1008	15	100	125	193.51	213	1815	332	19	102	124	335.4	190	4049	150	8.24E-04	1.98E-01	1.03E-03

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

1	ABO Rh (1 = O/+, 2 = A/+, 3=B/+, 4=AB/+)	N(1)=4	N(2)=7	N(3)=2	N(4)=15			1065	N(1)=7	N(2)=5	N(3)=1	N(4)=9			355	N(1)=20	N(2)=18	N(3)=4	N(4)=23			144	2.09E-01	7.91E-01	5.81E-02

Table 2: 60 subclinical factors from 2173 Covid19 patients, 3 groups of patients: Severe, Moderate and Mild.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

3.2. Correlation between the severity of Covid-19 and more than 60 preclinical factors

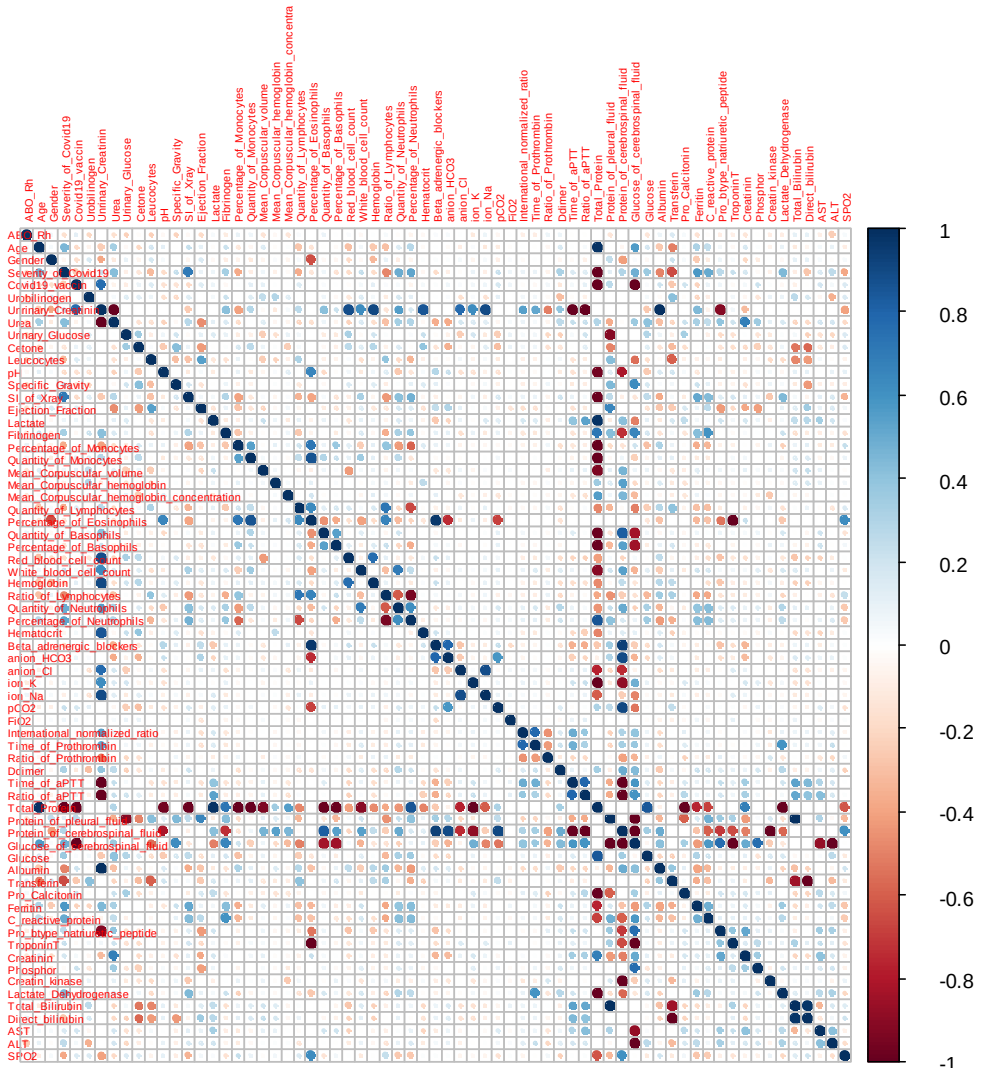
The correlation R measures the strength of the linear relationship between two quantitative variables. Pearson R formula is:

$$\boldsymbol{R} = \frac{\mathbf{1}}{\boldsymbol{n-1}} \sum \left(\frac{\boldsymbol{x_i - \bar{x}}}{\boldsymbol{S_x}}\right)\left(\frac{\boldsymbol{y_i - \bar{y}}}{\boldsymbol{S_y}}\right) \qquad 4$$

Equation 1: The point biserial correlation coefficient. R: Pearson correlation coefficient, x and y: two vectors of length i and j.

R is between -1 and 1. R > 0 reveals a positive association. R < 0 indicates a negative association. Values of R near 0 indicate a fragile linear relationship. The strength of the linear affinity increases as R moves away from 0 toward -1 or 1. The extreme values R = -1 and R = 1 occur only in the case of a perfect linear relationship.[4] Taking the database from the electronic medical records of 2173 Covid-19 patients, we calculate the R score, which shows the direction of correlation between the factors. (Figure 1, Supplementary table 1 – 2: excel file)

Figure 2: Correlation matrix between biomarkers depicted as a heat-map. Heat map represents the color-coded



correlation factors between more than 66 preclinical factors. The color value of the cells is proportional to the strength of the associations, ranging from red (negative correlations) to blue (positive correlations). The strength of the correlation is indicated in the color scale (at the right of the panel). Pair-wise Pearson correlation coefficients are shown in Supplementary Table 1-2.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

Figure 2 shows the first general view of some strong associations in both two directions, which shows a strong association in two directions (negative and positive). We focus on the association of the severity of Covid19 to all other factors. It strongly and positively correlates with Age, Urea, Score index of the chest X-ray, quantity and ratio of neutrophils, the quantity of protein and glucose in cerebrospinal fluid, the quantity of glucose in the blood, Ferritin, C reactive protein, and lactate dehydrogenase. Their R score respectively are: 0.45, 0.43, 0.71, 0.50, 0.52, 0.29, 0.49, 0.37, 0.58, 0.50, 0.44 ($p<0.0001$). The severity strongly and negatively correlates with quantity of Albumin, Transferrin, SPO_2 with R score are: -0.50, -0.65, -0.39 ($p<0.00001$) (Figure 2 and Supplementary Table 1-2)

Separating 2173 patients in 3 groups: the mild and no symptoms (1093 and 494 patients), the moderate (377 patients), the severe (209 patients); taking all collected information, we found the significant different of several factors and the critical shift of associated direction between the severity groups. (Figure 3, Table 2-3).

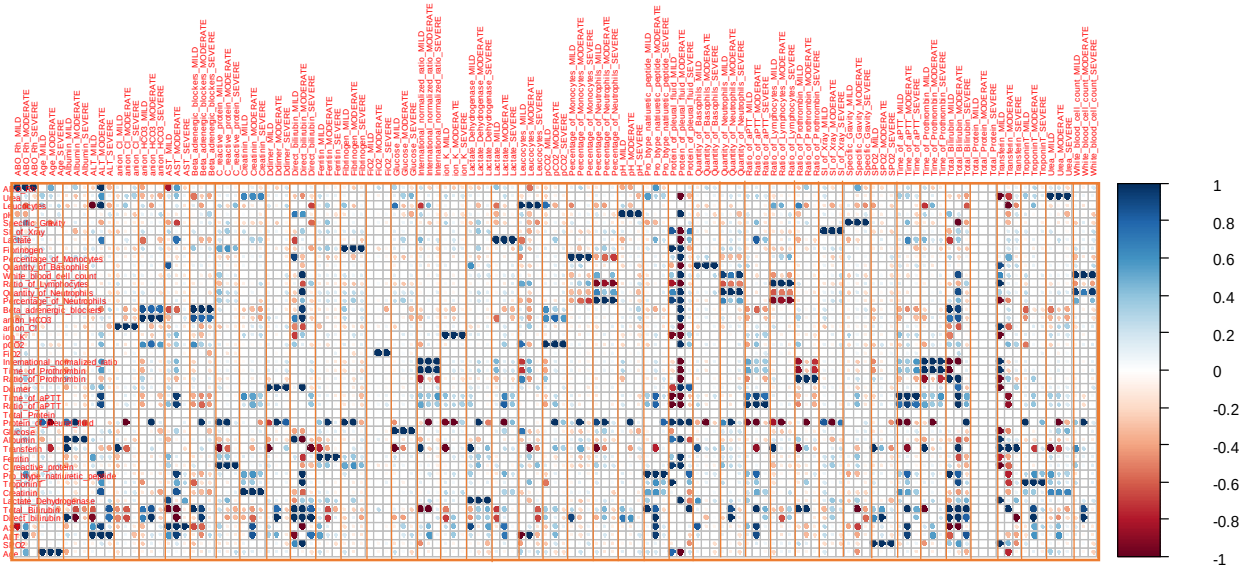


Figure 3: Correlation matrix between biomarkers depicted as a heat-map. Heat map represents the color-coded correlation factors between all the subclinical indexes. The color value of the cells is proportional to the strength of the associations, ranging from red (negative correlations) to blue (positive correlations). The strength of the correlation is indicated in the color scale (at the right of the panel).

We selected common forty-two factors for three groups to run the second correlation test; we found different correlative indications. Table 3 shows all the pairs which have not only a significant link but also the inverse correlation following the severity of Covid19. In which their absolute R score values are from 0.20 to 1 (Figure 3, Table 3, and Supplementary table 3-8).

1st factor	2nd factor	R scores MILD	P values MILD	R scores MODERATE	P values MODERATE	R scores SEVERE	P values SEVERE	Inverse correlation
Urea	anion Cl	-0.20	1.41E-08	0.29	4.27E-08	0.32	1.99E-06	- in mild to + in moderate/ severe
Leucocytes	ALT	-1.00	0.00E+00	1.00	0.00E+00	-0.34	2.11E-01	- in mild to + in moderate
Ratio of Lymphocytes	Creatinine	-0.08	1.76E-02	-0.02	6.44E-01	0.24	6.12E-04	- in mild to + in severe
Beta adrenergic blockers	ALT	-0.11	7.32E-01	-0.47	6.19E-03	0.28	2.87E-02	- in moderate to + in severe
anion Cl	Urea	-0.20	1.41E-08	0.29	4.27E-08	0.32	1.99E-06	- in mild to + in moderate/ severe
anion Cl	D-dimer	-0.10	7.85E-03	0.07	2.16E-01	0.21	2.89E-03	- in mild to + in severe

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

anion Cl	Transferrin	1.00	0.00E+00	-0.59	2.01E-02	0.05	7.87E-01	+ in mild to - in moderate
D-dimer	Protein of pleural fluid	-0.70	1.86E-01	1.00	0.00E+00	0.82	1.23E-02	- in mild to + in moderate/ severe
Protein of pleural fluid	Lactate	0.00	0.00E+00	-1.00	0.00E+00	0.43	3.35E-01	- in moderate to + in severe
Protein of pleural fluid	Fibrinogen	1.00	4.18E-02	1.00	0.00E+00	-0.37	4.17E-01	+ in mild to - in severe
Albumin	Direct bilirubin	1.00	0.00E+00	-1.00	0.00E+00	-0.30	2.34E-01	+ in mild to - in moderate/ severe
Direct bilirubin	ALT	-0.94	2.25E-01	0.00	0.00E+00	0.76	4.87E-02	- in mild to + in severe
ALT	Total Bilirubin	-0.61	3.87E-01	0.00	0.00E+00	0.85	3.03E-02	- in mild to + in severe
SPO2	Transferrin	1.00	0.00E+00	0.46	1.00E-01	-0.46	8.90E-03	+ in mild/moderate to - in severe
Age	Glucose	0.22	7.70E-10	0.04	5.20E-01	-0.20	9.27E-03	+ in mild to - in severe

Table 3: inverse correlation between the groups: Severe, Moderate and Mild. The color value of the cells is proportional to the strength of the associations, ranging from red (negative correlations) to blue (positive correlations).

	SEVERE vs MODERATE	SEVERE vs MILD	MODERATE vs MILD
Albumin	5.08E-02	8.09E-02	9.91E-05
ALT	2.70E-01	4.18E-02	6.23E-02
anion Cl	5.66E-03	3.93E-03	6.86E-02
anion HCO3	1.10E-03	9.02E-05	6.89E-05
Beta-adrenergic blockers	1.21E-02	5.04E-04	4.56E-04
C reactive protein	1.25E-01	1.80E-03	4.41E-04
Creatinine	9.23E-04	3.24E-04	7.76E-03
D-dimer	2.60E-04	6.71E-01	6.58E-01
Ejection Fraction	9.73E-02	2.78E-02	8.45E-02
Ferritin	3.06E-03	9.82E-05	1.45E-02
Fibrinogen	6.12E-01	1.15E-01	3.23E-03
FiO2	6.08E-01	1.15E-01	3.23E-03
Glucose	6.08E-01	1.15E-01	6.50E-01
ion K	8.40E-02	5.38E-02	3.32E-06
Lactate	1.59E-01	8.53E-02	2.43E-01
Lactate Dehydrogenase	4.04E-01	3.86E-03	3.39E-02
Percentage of Neutrophils	4.04E-01	3.86E-03	2.38E-07
pH	3.29E-03	9.94E-03	3.55E-05
Pro b-type natriuretic peptide	3.14E-04	3.75E-01	6.61E-01
Protein of pleural fluid	9.89E-01	9.65E-01	5.40E-01
Quantity of Basophils	2.13E-04	6.20E-02	3.70E-01
Quantity of Neutrophils	7.62E-03	1.81E-01	3.25E-04
Ratio of Lymphocytes	1.52E-01	2.00E-01	8.66E-08
SI of X-ray	5.34E-01	3.64E-05	7.07E-01
White blood cell count	3.68E-03	3.84E-01	1.84E-02
SPO2	1.62E-03	2.62E-02	5.13E-09

Table 4: P-value from the pairwise t-test, compare the correlation (R score) of each factor with other (27 factors in the table) between three groups of patients: Severe, Moderate, and Mild.

We have taken all the R score in three groups for the pairwise t-test. We found again these significant differences ($p<0.05$) of several factors. That confirms the important shift of correlation between the group of the severity of Covid19. (Table 4)

3.3. Cut-off value and exact Risk ratio shows the predicted values affect on the severity of COVID-19 patients

We fit a logistic regression model to predict the subclinical values, which link with 50:50 probabilities that the status of mild/moderate patients could be severe. (Table 5)

	C L H	E S T L	S T E D	N - V A	P - R	N P	N P	N S	N S	N C	N C	NORMAL VALUE

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

												Female	Male	units
SPO_2	87.80	-2.00E-01	2.16E-02	-9.24E+00	2.45E-20	1112	74	144	49	968	25	95-100	95-100	%
UREA	15.87	2.07E-01	1.80E-02	1.15E+01	1.33E-30	82	1463	36	171	46	1292	2.76 - 8.07	2.76 - 8.07	mmol/L
LEUCOCYTES	273.69	-3.59E-03	1.35E-03	-2.67E+00	7.63E-03	16	80	5	43	11	37	<10	<10	LEU/uL
PH	6.52	-8.80E-01	1.52E-01	-5.78E+00	7.37E-09	606	84	150	44	456	40	4.8 - 7.4	4.8 - 7.4	
SPECIFIC GRAVITY	1013.98	5.40E-02	1.95E-02	2.77E+00	5.61E-03	178	54	90	21	88	33	1.005 - 1.030	1.005 - 1.030	
SI OF XRAY	11.03	4.46E-01	3.17E-02	1.41E+01	5.29E-45	211	815	103	82	108	733			
LACTATE	8.33	7.93E-02	4.56E-02	1.74E+00	8.17E-02	10	639	6	185	4	454	0.6 - 1.4	0.6 - 1.4	mmol/L
FIBRINOGEN	7.75	4.47E-01	5.62E-02	7.96E+00	1.76E-15	23	1357	6	182	17	1175	2.00 - 4.00	2.00 - 4.00	g/L
PERCENTAGE OF MONOCYTES	1.21	-4.53E-01	3.74E-02	-1.21E+01	9.77E-34	1594	39	200	8	1394	31	0.00 - 0.80	0.00 - 0.80	%
QUANTITY OF BASOPHILS	0.76	2.47E+00	1.06E+00	2.32E+00	2.01E-02	3	1623	1	207	2	1416	0.00 - 0.10	0.00 - 0.10	G/L
WHITE BLOOD CELL COUNT	21.18	1.45E-01	1.52E-02	9.59E+00	8.83E-22	48	1565	28	177	20	1388	4.00 - 10.00	4.00 - 10.00	G/L
RATIO OF LYMPHOCYTES	3.99	-2.39E-01	1.82E-02	-1.31E+01	2.65E-39	1450	165	124	83	1326	82	20 - 45	20 - 45	%
QUANTITY OF NEUTROPHILS	14.27	2.59E-01	1.89E-02	1.37E+01	9.45E-43	113	1503	66	142	47	1361	1.8 -7.5	1.8 -7.5	G/L
PERCENTAGE OF NEUTROPHILS	90.67	1.80E-01	1.32E-02	1.36E+01	2.46E-42	186	1440	87	121	99	1319	45 - 75	45 - 75	%
BETA ADRENERGIC BLOCKERS	40.22	-1.12E-01	2.19E-02	-5.10E+00	3.33E-07	528	121	143	64	385	57	45 - 52	45 - 52	mmol/l
ANION HCO3	17.04	-8.69E-02	2.35E-02	-3.69E+00	2.24E-04	601	62	168	32	433	30	21.00 - 29.50	21.00 - 29.50	mmol/L
ANION CL	127.08	6.12E-02	1.36E-02	4.49E+00	7.09E-06	3	1620	2	206	1	1414	98.0 - 107.0	98.0 - 107.0	mmol/L
ION K	6.12	8.43E-01	1.51E-01	5.60E+00	2.14E-08	6	1599	4	196	2	1403	3.40 - 4.50	3.40 - 4.50	mmol/L
pCO_2	47.10	4.39E-02	9.68E-03	4.54E+00	5.65E-06	73	597	46	158	27	439	35.00 - 45.00	35.00 - 45.00	mmHg
FIO_2	0.21	1.34E+03	6.05E+04	2.22E-02	9.82E-01	2	667	2	203	0	464			
INTERNATIONAL NORMALIZED RATIO	2.56	2.85E-01	1.85E-01	1.54E+00	1.23E-01	18	494	6	174	12	320	0.8 - 1.2	0.8 - 1.2	
TIME OF PROTHROMBIN	27.87	3.14E-02	2.12E-02	1.48E+00	1.38E-01	20	494	7	173	13	321			
RATIO OF PROTHROMBIN	30.98	-9.77E-03	5.54E-03	-1.76E+00	7.77E-02	492	22	173	7	319	15	70.00 - 140.0	70.00 - 140.0	%
DDIMER	8676.77	2.51E-04	3.15E-05	7.99E+00	1.37E-15	51	1498	26	173	25	1325	<500.00	<500.00	ng/mL
TIME OF APTT	38.63	2.51E-02	8.35E-03	3.00E+00	2.68E-03	98	350	50	129	48	221			
RATIO OF APTT	1.36	6.03E-01	2.88E-01	2.09E+00	3.64E-02	100	345	51	125	49	220	0.8-1.2	0.8-1.2	times
TOTAL PROTEIN	51.65	-7.48E+00	1.78E+04	-4.19E-04	1.00E+00	1	2	0	1	1	1	64 - 83	64 - 83	g/L
PROTEIN OF PLEURAL FLUID	4.32	6.42E-03	4.61E-02	1.39E-01	8.89E-01	15	1	7	1	8	0	0.15 - 0.45	0.15 - 0.45	g/L
GLUCOSE	23.60	1.33E-01	1.50E-02	8.85E+00	8.44E-19	41	1400	15	149	26	1251	4.56 - 6.38	4.11 - 5.89	mmol/L
ALBUMIN	27.29	-1.92E-01	2.03E-02	-9.46E+00	3.07E-21	501	159	98	79	403	80	35 - 52	35 - 52	g/L
TRANSFERIN	166.85	-2.36E-02	8.07E-03	-2.92E+00	3.47E-03	12	49	3	28	9	21	200 - 360	200 - 360	mg/dL
FERRITIN	1887.07	2.09E-03	1.51E-04	1.39E+01	1.24E-43	110	1396	38	146	72	1250	10 - 291	30 - 400	ng/ mL
C REACTIVE PROTEIN	20.69	1.35E-01	1.22E-02	1.11E+01	1.29E-28	58	1385	24	129	34	1256	<0.50	<0.50	mg/dL
PRO BTYPE NATRIURETIC PEPTIDE	6569.28	1.50E-04	3.65E-05	4.11E+00	3.92E-05	44	521	25	120	19	401	<125	<125	pg/mL
TROPONIN-T	167.28	6.58E-03	1.39E-03	4.73E+00	2.27E-06	44	608	27	139	17	469	<14.0	<14.0	ng/L
CREATININ	382.15	6.16E-03	1.27E-03	4.84E+00	1.31E-06	15	1602	3	194	12	1408	44 - 80	62 - 106	umol/L
LACTATE DEHYDROGENASE	659.69	6.02E-03	4.89E-04	1.23E+01	9.12E-35	85	1374	48	132	37	1242	135 - 214	135 - 225	U/L
TOTAL BILIRUBIN	95.27	-8.09E-03	9.33E-03	-8.67E-01	3.86E-01	2	40	0	21	2	19	<15	<24	µmol/l
DIRECT BILIRUBIN	72.14	-2.08E-02	1.35E-02	-1.54E+00	1.23E-01	2	37	0	23	2	14	≤3.4	≤3.4	µmol/l
AST	886.48	6.11E-04	4.41E-04	1.38E+00	1.67E-01	6	258	3	84	3	174	≤32	≤40	U/I
ALT	588.41	1.91E-03	7.82E-04	2.44E+00	1.46E-02	15	216	7	52	8	164	≤33	≤41	U/I

Table 5: Relative risk ratio. Predicted values linked with a 50:50 probability that Covid-19 patients change from mild/moderate to severe status.

Several factors have a significant cutoff50 ($p<0.05$) which is the boundary between the group of mild/moderate and severe. Taking these values to the Meta Risk Ratio (RR) analysis, using several meta-analytical methods, we found the RR=1.493 [1.42-1.57] ($z=16.12$, $p<0.0001$ (Table 6)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

Cut off value from:	RR	95%-CI	%W (common)	Cut off value from:	RR	95%-CI	%W(common)
Ratio of Lymphocytes	0.17	[0.1355; 0.2132]	12.9	Ratio of aPTT	1.4076	[1.1098; 1.7853]	4.9
SPO_2	0.1956	[0.1565; 0.2444]	8	AST	1.5357	[0.6769; 3.4840]	0.3
Direct bilirubin	0.3191	[0.0264; 3.8550]	0.3	Creatinin	1.6515	[0.5951; 4.5832]	0.3
Total Bilirubin	0.3767	[0.0310; 4.5715]	0.3	ALT	1.9385	[1.0739; 3.4990]	0.6
Albumin	0.3937	[0.3107; 0.4988]	10.4	Fibrinogen	1.9451	[0.9649; 3.9211]	0.5
Transferrin	0.4375	[0.1594; 1.2006]	1	Lactate	2.0724	[1.2316; 3.4874]	0.5
pH	0.4725	[0.3692; 0.6047]	6.7	pCO2	2.381	[1.9092; 2.9693]	3
Protein of pleural fluid	0.4839	[0.2894; 0.8091]	0.2	Pro b-type natriuretic peptide	2.4669	[1.8245; 3.3354]	1.6
Beta adrenergic blockers	0.512	[0.4114; 0.6373]	9	Quantity of Basophils	2.6135	[0.5248; 13.0144]	0.1
anion HCO3	0.5416	[0.4122; 0.7116]	5	TroponinT	2.6841	[2.0363; 3.5380]	1.6
Total Protein	0.5556	[0.0466; 6.6286]	0.1	FiO2	3.2801	[2.9251; 3.6782]	0.2
Leucocytes	0.5814	[0.2734; 1.2366]	1.2	Ferritin	3.3031	[2.4481; 4.4567]	1.9
Percentage of Monocytes	0.6117	[0.3254; 1.1499]	1.4	Glucose	3.4376	[2.2348; 5.2877]	0.7
International normalized ratio	0.9464	[0.4871; 1.8387]	1.1	Urea	3.7561	[2.8322; 4.9813]	1.6
Time of Prothrombin	0.9994	[0.5435; 1.8379]	1.2	Ddimer	4.4144	[3.2591; 5.9792]	1
Ratio of Prothrombin	1.1051	[0.5925; 2.0612]	1.2	C reactive protein	4.4427	[3.1382; 6.2894]	0.9
Specific Gravity	1.3002	[0.9030; 1.8720]	2.8	SI of Xray	4.8518	[3.7883; 6.2138]	2.9
Time of aPTT	1.3843	[1.0916; 1.7555]	4.9	White blood cell count	5.1577	[3.9121; 6.8000]	0.9
				anion Cl	5.2427	[2.3317; 11.7882]	0.1
				ion K	5.4388	[3.0428; 9.7216]	0.1
				Percentage of Neutrophils	5.5665	[4.4259; 7.0012]	2.4
				Lactate Dehydrogenase	5.8781	[4.5904; 7.5270]	1.3
				Quantity of Neutrophils	6.1821	[4.9578; 7.7087]	1.7

Table 6: risk ratio on 41 preclinical factors. The Meta analytical method, such as: the Mantel-Haenszel method, the restricted maximum-likelihood estimator for tau^2, the Q-profile method for confidence interval of tau^2 and tau, the continuity correction of 0.5 in studies with zero cell frequencies, we found the RR=1.39 [1.33-1.46] (z=13.73, p<0.0001), tau^2=0.9682 [0.6084;1.5641], tau=0.9840 [0.7800;1.2506], I^2=98.0% [97.7%;98.3%], H=7.05[6.58;7.56]. Test of heterogeneity: Q=2040.66, d.f. = 41, p-value <0.0001. The Outcome + : the severe group, the Outcome -: the mild/moderate group.

We see a very high RR from 2.47 to 6.18 of these factors: Pro-BNP (protein b type natriuretic peptide), the number of basophils, Troponin T, FiO2, Ferritin, Glucose, Urea, D-dimer, CRP, SI of X-ray, WBC, anion Cl, ion K, ratio and quantity of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

Neutrophils, lactate dehydrogenase. Other factors such as the Ratio of Lymphocytes, direct and total bilirubin, Albumin, transferrin, pH, Protein of pleural fluid, SPO_2 show very low RR (0.17 to 0.48), which fits with their inverse correlation with the Covid-19 severity.

MILD vs MODERATE RR=2.78 [2.66;2.90] (z=45.29, p<0.000001)	**Percentage of Neutrophils**	**Protein of pleural fluid**	**C reactive protein (CRP)**	**Ferritin**	**Age**	**SI of X-ray**	**Protein of cerebrospinal fluid**
cutoff50 mild to moderate	84.8	38.98	7.46	1194.7	77.56	5.53	0.92
Estimate	6.05E-02	7.36E-02	2.02E-01	1.70E-03	5.24E-02	4.26E-01	1.88E+00
Std-Error	5.55E-03	8.14E-02	1.84E-02	1.29E-04	3.47E-03	3.63E-02	1.81E+00
Z-value	1.09E+01	9.05E-01	1.10E+01	1.32E+01	1.51E+01	1.17E+01	1.04E+00
P	1.13E-27	3.66E-01	4.40E-28	1.25E-39	1.92E-51	9.09E-32	2.97E-01
N patients in upper of cutoff50	171	2	148	204	238	247	2
N patients in lower of cutoff50	1201	13	1141	1082	1610	490	5
N moderate patients upper of cutoff50	114	1	104	137	143	189	2
N moderate patients lower of cutoff50	252	1	236	201	231	114	1
N other patient in upper of cutoff50	57	1	44	67	95	58	0
N other patient in lower of cutoff50	949	12	905	881	1379	376	4
RR	**3.18**	**6.5**	**3.4**	**3.62**	**4.19**	**3.29**	**3.67**
95%-CI	[2.73-3.7]	[0.63-67.35]	[2.91-3.97]	[3.09-4.23]	[3.58-4.9]	[2.77-3.92]	[0.94-14.35]
%W(common)	5.1	0	4.4	5.2	4.8	6.2	0.1
test statistic	5.14E+01	1.88E-01	2.07E+02	1.92E-29	2.66E+02	2.21E+01	2.72E-01
df	1	1	1	1	1	1	1
p value 1s	3.74E-13	3.33E-01	3.85E-47	5.00E-01	4.52E-60	1.27E-06	3.01E-01
p value 2s	7.48E-13	6.65E-01	7.70E-47	1.00E+00	9.03E-60	2.54E-06	6.02E-01
R score with Severity of Covid19	0.33	0.36	0.4	0.43	0.44	0.52	0.57
P(pearson) with Severity of Covid19	0.00E+00	4.30E-01	0.00E+00	0.00E+00	0.00E+00	0.00E+00	2.35E-01

MODERATE vs SEVERE RR=1.28 [1.22;1.33] (z=11.23, p<0.0001)	**Quantity of Neutrophils**	**Ratio of Lymphocytes**	**Transferrin**	**Albumin**	**Percentage of Neutrophils**	**Direct bilirubin**	**SI of X-ray**
cutoff50 moderate to severe	11.77	6.32	170.73	29.73	87.74	108.97	10.51
Estimate	1.79E-01	-1.71E-01	-2.16E-02	-1.87E-01	1.31E-01	-2.45E-02	3.53E-01
Std-Error	2.04E-02	1.88E-02	8.48E-03	2.44E-02	1.38E-02	1.65E-02	3.49E-02
Z-value	8.76E+00	-9.07E+00	-2.55E+00	-7.68E+00	9.50E+00	-1.49	10.12
P	1.97E-18	1.18E-19	1.08E-02	1.65E-14	2.01E-21	1.37E-01	4.50E-24
N patients in upper of cutoff50	126	389	7	209	200	1	174
N patients in lower of cutoff50	446	186	39	165	374	25	314
N severe patients upper of cutoff50	98	86	3	65	133	0	116
N severe patients lower of cutoff50	110	121	28	112	75	2	69
N other patient in upper of cutoff50	28	303	4	144	67	1	58
N other patient in lower of cutoff50	336	65	11	53	299	23	245
RR	**3.15**	**0.34**	**0.60**	**0.46**	**3.32**	**3.40**	**3.03**
95%-CI	[2.6153;3.8025]	[0.2743;0.4210]	[0.2482;1.4359]	[0.3650;0.5752]	[2.6480;4.1529]	[0.2652;43.5901]	[2.4023;3.8314]
%W(common)	3.3	11	0.6	8.4	3.5	0	3.3
test statistic	1.17E+02	9.89E+01	1.14E+00	4.86E+01	1.20E+02	1.08E-30	9.31E+01
df	1	1	1	1	1	1	1
p value 1s	1.12E-27	1.35E-23	1.43E-01	1.60E-12	3.74E-28	5.00E-01	2.47E-22
p value 2s	2.24E-27	2.69E-23	2.86E-01	3.19E-12	7.49E-28	1.00E+00	4.93E-22
R score with Severity of Covid19	0.42	-0.41	-0.42	-0.43	0.44	-0.47	0.52
P(pearson) with Severity of Covid19	0.00E+00	0.00E+00	3.58E-03	0.00E+00	0.00E+00	1.56E-02	0.00E+00

Table 7: Most important factor in 2 groups: Mild-Moderate, Moderate-Severe. The Meta analytical method, such as: the Mantel-Haenszel method, the restricted maximum-likelihood estimator for tau^2, the Q-profile

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

method for confidence interval of tau^2 and tau, the continuity correction of 0.5 in studies with zero cell frequencies, we found in the **Mild – Moderate group for 47 factors (upper part of table)**: RR=2.78 [2.66;2.90] (z=45.29, p<0.000001), tau^2=0.2635 [0.2043;0.9629], tau=0.5133 [0.4520;0.9813], I^2=91.6% [89.7%;93.1%], H=3.45 [3.11;63.82]. Test of heterogeneity: Q=546.33, d.f. = 46, p-value <0.0001. The Outcome + : moderate, the Outcome -: mild**. In the Moderate – Severe group for 37 factors (lower part of table)**: RR=1.28 [1.22;1.33] (z=11.23, p<0.00001), tau^2=0.5394 [0.3239;0.9054], tau=0.7344 [0.5691;0.9515], I^2=98.0% [97.7%;98.3%], H=7.15 [6.65;7.70]. Test of heterogeneity: Q=1842.55, d.f. = 36, p-value <0.05. The Outcome + : severe, the Outcome -: moderate.

See more in detail when we have taken every two groups together for the relative risk ratio imputation and then the exact risk ratio. The Meta Risk Ratio (RR) analysis show the pooled RR in the mild-moderate group is RR=2.78 [2.66;2.90] (z=45.29, p<0.000001) and in the moderate-severe group is RR=1.28 [1.22;1.33] (z=11.23, p<0.0001) (Table 5). We selected the R score (|R|>0.3) with the severity of covid19 to calculate their risk ratio values. The percentage of neutrophils, CRP, Ferritin, Age, and SI of X-ray show the not only their strong and significant correlation with the severity of Covid19 (mild and moderate), with R scores respectively: 0.33, 0.4, 0.43, 0.44, 0.52 (p<0.0001), their RR for the cutoff50 value were also significantly higher than the pooled RR, they respectively are: 3.18, 3.4, 3.62, 4.19, 3.29 (p<0.00001).

Looking at the moderate-severe group, the percentage and the number of neutrophils, SI of X-ray also show a significant positive correlation with the Covid19 severity with their important RR, respectively are: 3.32, 3.15, 3.03 (p<0.00001), which are higher than the multi-variant RR. Albumin and Ratio of Lymphocytes appear as a powerfully negative correlated factors with the Covid19 severity with their RR are 0.46 and 0.34 (p<0.00001), significantly lower than pooled RR. (Table 7)

3.4. HCA (Hierarchical Clustering Study)

Based on the correlation score (R) above, the clustering imputation shows the difference in branching between the factors. (Figure 4) The agglomerative hierarchical clustering algorithms in this program module build a cluster hierarchy commonly displayed as a tree diagram of the dendrogram. They start with each object in a separate cluster. At each step, the two most similar clusters joined into a single new cluster. Once fused, objects are never separated. Within each cluster, the value for this measure is displayed from smallest to largest. Clustering is partitioning a set of objects into groups (clusters) where objects within a group are more similar than objects in different groups. Most clustering algorithms depend on assumptions to define the subgroups present in a data set. Consequently, the resulting clustering scheme requires some evaluation regarding its validity. [5]

Based on the dendrogram, with seeing clearly, the neighboring factors around the Severity of Covid19 are nine factors: D-dimer, Fibrinogen, C reactive protein, lactate dehydrogenase, white blood cell count, the number of Neutrophils, Glucose, SI of X-ray and Ferritin. Going deeper into every group of patients, the neighbors of those nine factors above changed from 3 groups. In the severe group, the glucose and the lactate dehydrogenase split into the other groups. They show their behaviour to the severity is changed. We used more than 20 classifier clustering algorithms to determine and predict the number of clusters. (Table 8)

ALL GROUPS			MILD			MODERATE			SEVERE		
Methods	**Number clusters**	**Value Index**	**Methods**	**Number clusters**	**Value Index**	**Methods**	**Number clusters**	**Value Index**	**Methods**	**Number clusters**	**Value Index**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

kl	3	2.90	*kl*	59	97.69	*kl*	5	5.31	*kl*	2	3.16
ch	41	28.44	*ch*	59	3153.89	*ch*	63	3268.76	*ch*	2	5.80
hartigan	41	Inf	*hartigan*	59	Inf	*hartigan*	63	Inf	*hartigan*	60	Inf
cindex	41	0.00	*cindex*	59	0.00	*cindex*	63	0.00	*cindex*	2	0.45
db	41	0.11	*db*	59	0.01	*db*	63	0.01	*db*	60	0.37
silhouette	41	0.98	*silhouette*	59	1.00	*silhouette*	63	1.00	*silhouette*	60	0.97
duda	5	0.97	*duda*	2	0.92	*duda*	5	0.89	*duda*	2	0.95
pseudot2	5	0.03	*pseudot2*	2	4.70	*pseudot2*	5	2.14	*pseudot2*	2	2.28
ratkowsky	3	0.30	*ratkowsky*	4	0.23	*beale*	18	-1.90	*beale*	6	-2.88
ball	3	29.45	*ball*	3	85.77	*ratkowsky*	4	0.26	*ratkowsky*	13	0.15
ptbiserial	2	0.58	*ptbiserial*	2	0.55	*ball*	3	28.26	*ball*	3	39.26
gap	2	-0.31	*gap*	2	-0.18	*ptbiserial*	5	0.59	*ptbiserial*	11	0.49
mcclain	2	0.03	*mcclain*	2	0.05	*gap*	2	0.14	*gap*	2	0.31
gamma	38	1.00	*gamma*	47	1.00	*mcclain*	2	0.89	*mcclain*	2	0.49
gplus	38	0.00	*gplus*	47	0.00	*gamma*	48	1.00	*gamma*	59	1.00
tau	4	143.44	*tau*	9	248.99	*gplus*	48	0.00	*gplus*	59	0.00
dunn	39	1.43	*dunn*	59	9.89	*tau*	5	279.64	*tau*	4	273.61
sdindex	32	1.86	*sdindex*	44	1.03	*dunn*	63	2.12	*dunn*	60	1.02
sdbw	41	0.00	*sdbw*	59	0.00	*sdindex*	49	2.36	*sdindex*	60	0.88
						sdbw	63	0.00	*sdbw*	60	0.01

Table 8: result of clustering imputation. Abbreviations of clustering methods: *CH* (Calinski and Harabasz 1974), *CCC* (Sarle 1983), Pseudot2 (Duda and Hart 1973), *KL* (Krzanowski and Lai 1988), *Gamma* (Baker and Hubert 1975), *Gap* (Tibshirani et al. 2001), *Silhouette* (Rousseeuw 1987), *Hartigan* (Hartigan 1975), *Cindex* (Hubert and Levin 1976), *DB* (Davies and Bouldin 1979), *Ratkowsky* (Ratkowsky and Lance 1978), *Scott* (Scott and Symons 1971), *Marriot* (Marriot 1971), *Ball* (Ball and Hall 1965), *Trcovw* (Milligan and Cooper 1985), *Tracew* (Milligan and Cooper 1985), *Friedman* (Friedman and Rubin 1967), *Rubin* (Friedman and Rubin 1967), *Dunn* (Dunn 1974).

The evaluation procedure has to tackle challenging problems such as the quality of clusters, the degree to which a clustering scheme fits a specific data set, and the optimal number of clusters in partitioning. These methods proposed the different optimal numbers of clusters. We chose the result proposed by the *tau* and the *ratskowsky* method, which propose 4 clusters as optimal (Table 8). The cluster plot in figure 5 describes the different distributions of the preclinical factors following the result above. We see the Covid-19 severity locates in the same group with age, D-dimer, Glucose in blood, C reactive protein (CRP), lactate dehydrogenase (LDH), ferritin, SI of X-ray, quantity/percentage of neutrophils, white blood cells count, Fibrinogen, Specific gravity and Protein of pleural fluid. Looking deeper in every group of severity, we see the group of factors above may contain pH, D-dimer, troponin T (in the mild group); transferrin, albumin, FiO2, anion HCO3, BB, pCo2, glucose (in the moderate group); or split into two different clusters in the severe group. That confirms the inversion of correlating direction of certain factors via their distance, such as Albumin - direct bilirubin– a ratio of monocytes (%), pCO2 – pH, the SI of chest X-ray – Protein in pleural fluid, ALT – Leucocytes, Albumin/Fibrinogen/Lactate Dehydrogenase - Protein in the pleural fluid, D-dimer – Protein in the pleural fluid, the ratio INR – quantity of leucocytes, Protein n the pleural fluid – lactate again. (Figure 5)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

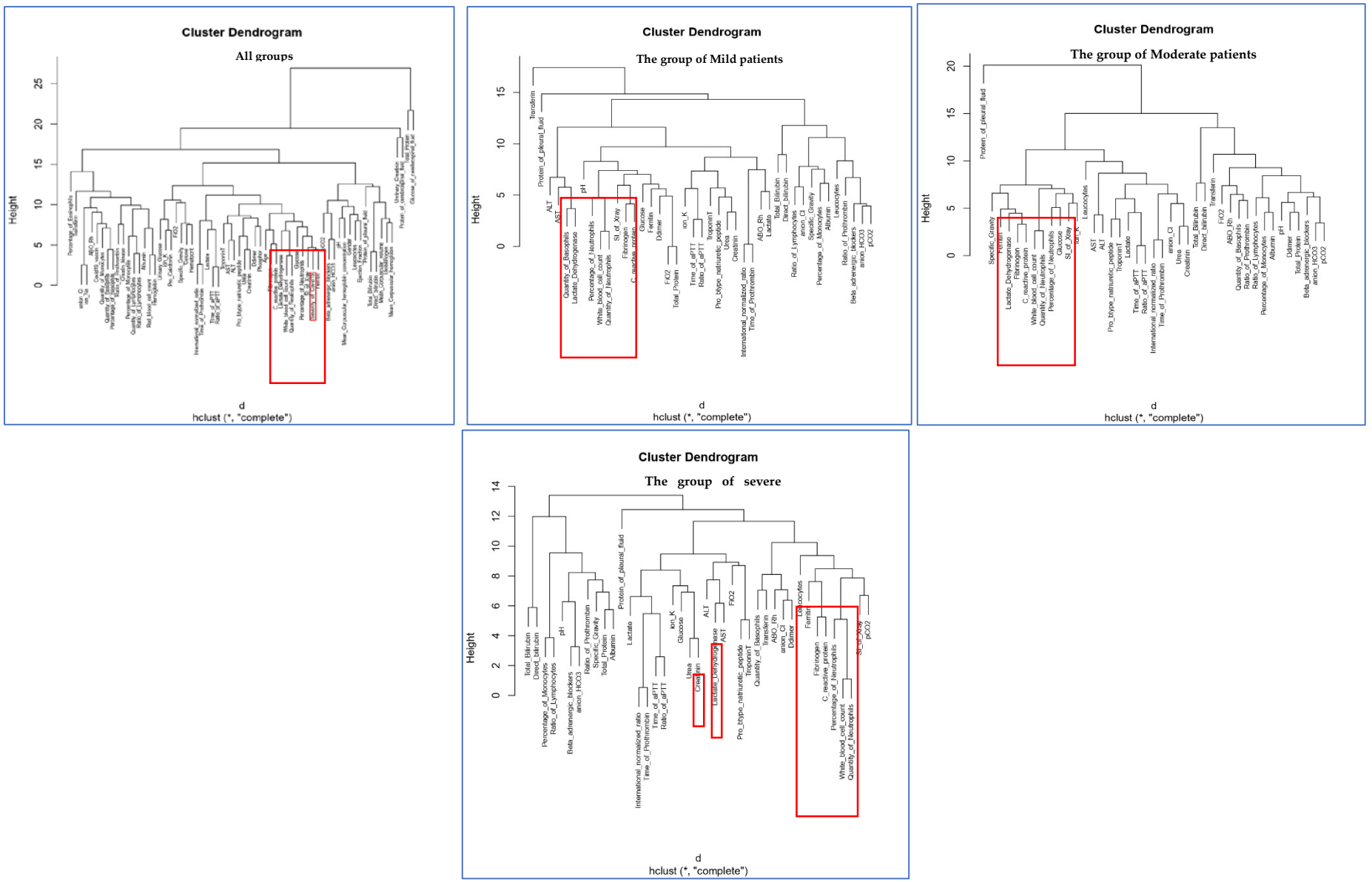


Figure 4: Cluster Dendrogram. The height axis displays the Euclidean distance between observations or clusters. The horizontal bars indicate when two clusters/observations merged.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

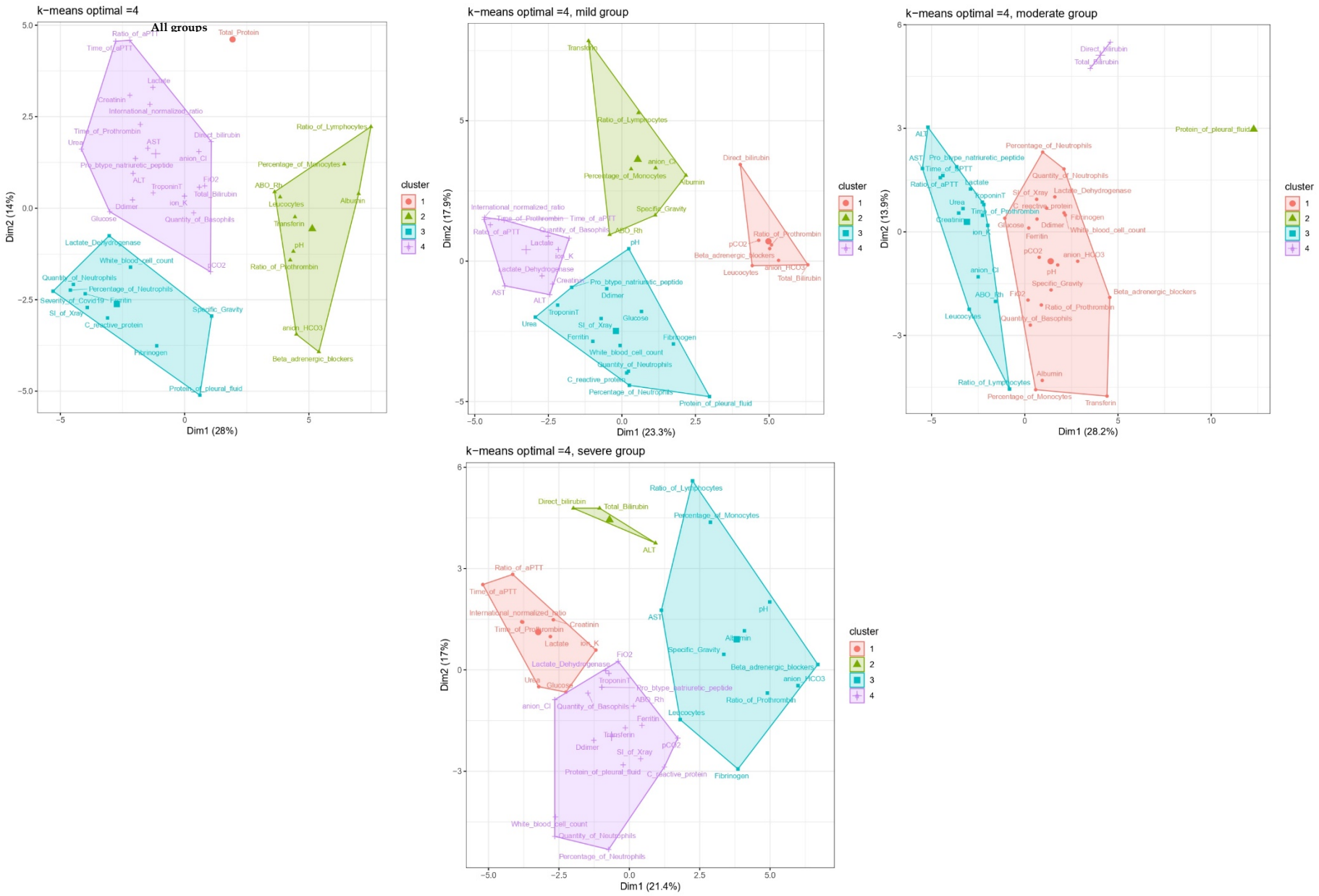


Figure 5: Cluster plot. (upper),All group, (lower, left to right) the group of Mild patients, the group of Moderate patients, the group of severe patients.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

3.5. Network building

Network analysis is a visualization approach that can represent various types of data. Several models can be used to build the map of nodes with their force of influence on each other. One of them is the layout called Multidimensional scaling. Multidimensional scaling (MDS) is a means of visualizing the level of similarity of individual cases of a dataset.

MDS is used to decode the pairwise 'distances' among a set of n objects or individuals into a layout of n points mapped into a conceptual Cartesian distance. MDS refers to a set of corresponding ordination techniques used in information visualization to display the information in a matrix distance. It is a structure of non-linear dimensionality reduction. They are given a distance matrix with the distances between each pair of objects in a set and a chosen N number of dimensions. An MDS algorithm sets each object into an N-dimensional area to preserve the between-object distances as well as possible. With N = 1, 2, and 3, the resulting points can be visualized on a scatter plot. The main theoretical contributions to MDS were made by James O. Ramsay of McGill University, founder of functional data analysis.[6]

Taking all the above results and using the MDS layout, we continue to look more at the influence of the Covid19 severity on the other factors and how their link was disturbed by other neighbors. Based on the R scores, the data to be analyzed is a collection of M objects (in our study, it will be the number of preclinical factors in each imputation) on which a distance function defines the entries of the dissimilarity matrix:

$$D := \begin{pmatrix} d_{1,1} & \cdots & d_{1,M} \\ \vdots & & \vdots \\ d_{M,1} & \cdots & d_{M,M} \end{pmatrix} \quad 6$$

Equation 2: The distance function from the Multidimensional layout. Where $d_{i,j:=}$ distance between $i-$th and $j-$ th objects.

The distance between each factor in the network shows how well they connect and which other factors could disturb their link. (Supplementary table 9)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 20 September 2022 doi:10.20944/preprints202209.0291.v1

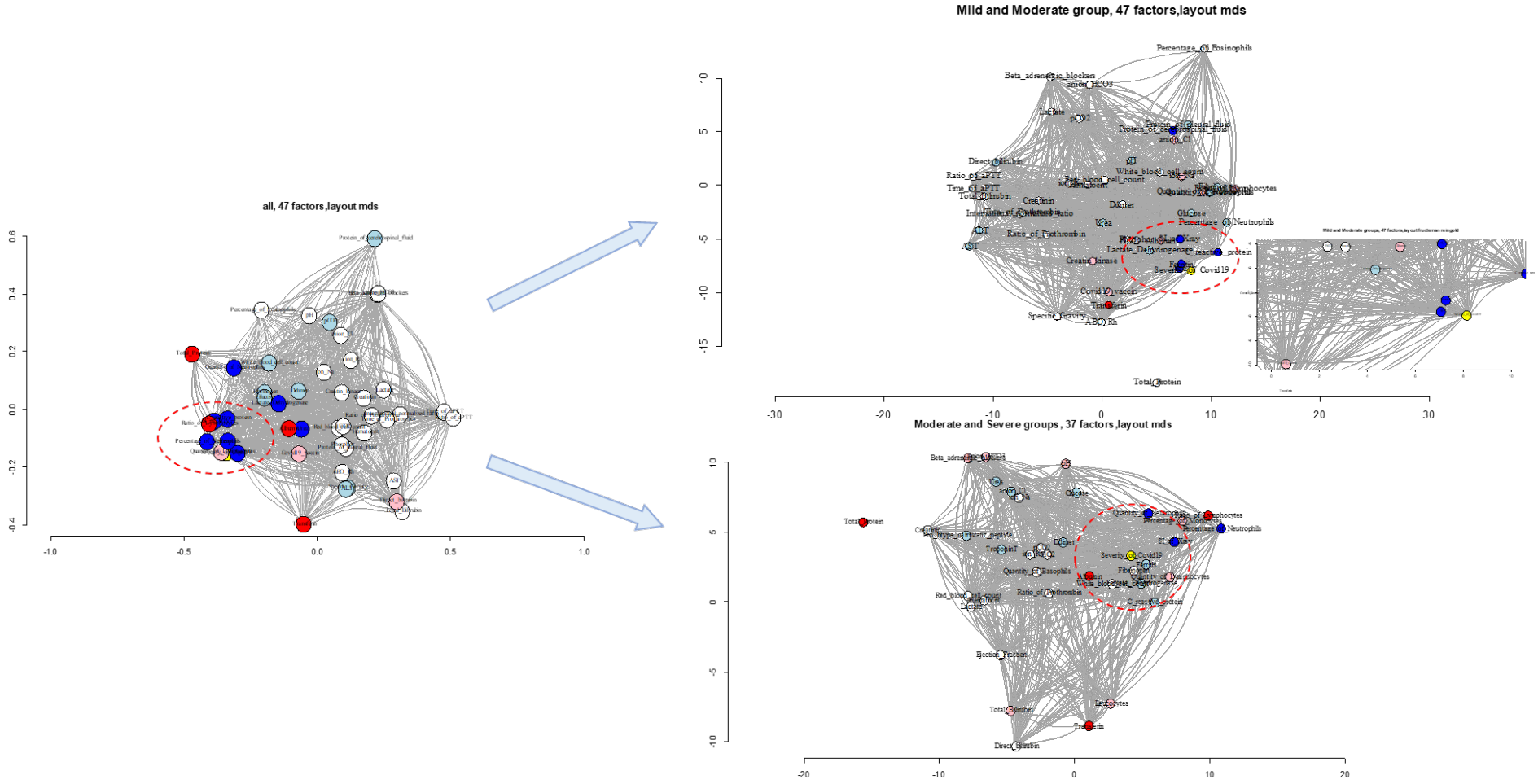


Figure 6: Network plot for Mild-Moderate group and Moderate-Severe group, 2 dimensions. Node in yellow color: Severity of Covid19. Node in red color: factors has strong negative correlation with the Covid19 severity (-0.4 > R > -1). Node in blue color: factors has strong positive correlation with the Covid19 severity (0.4 < R < 1). Node in pink color: factors has moderate negative correlation with the Covid19 severity (-0.20 > R > -0.40). Node in color light-blue: factors has moderate positive correlation with the Covid19 severity (0.20 < R < 0.40). Node in grey color: factors has weak correlation with the Covid19 severity (-0.20 < R < 0.20).

 doi:10.20944/preprints202209.0291.v1

4. Discussion

4.1. The Covid19 Severity directly links to Age, Score index of the chest X-ray, percentage of neutrophils, Albumin, C reactive protein, and ratio of Lymphocytes.

Age, Score index of X-ray, Percentage of Neutrophils, and C reactive protein (CRP) show a strong positive link with the severity of Covid19. Moreover, their exact risk ratio (RR) based on the cutoff50s significantly differs from the pooled RR.

77.56 years old is the significant cutoff50 with a risk ratio is 4.19 [3.58-4.95] ($p<0.00001$) in the mild-moderate group. This risk ratio is higher than the pooled RR (RR=2.78, $p<0.00001$) by about 1.41, and the probability of the mild patients becoming moderate increases three times. In this group, the R score between Age and severity is 0.44 ($p<0.0001$) (Table 6)

There are many views on the age-related difference in the severity of COVID-19. The explanation for the marked age gradient is likely multifactorial. The proposed mechanisms that relate specifically to the pathogenesis of SARS-CoV-2 seem more likely to be critical than those that would also apply to other viral infections for which a similar age gradient is not seen. Differences in innate, adaptive, and heterologous immunity and differences in the endothelial and clotting function are the most potential mechanisms to explain the observed age gradient in COVID-19. Following exposure to SARS-CoV-2, immunologic factors in children are essential in preventing or controlling the virus after infection. Age-related differences in endothelial and clotting function are more important in putting the elderly at risk of complications of COVID-19 that lead to higher mortality.[7]

Neutrophils are a typical type of white blood cell produced by the bone marrow and stimulate phagocytosis to attack bacteria. In normal status, neutrophils have a value of around 2-8 G/L, equivalent to 43-76% of the total leukocytes. Neutrophil totals are used in clinical practice to inhibit the body's ability to fight infections, particularly bacterial infections. Patients had neutropenia when the absolute blood count was below 2000/µl and an increased risk of disease when the count was below 1000/µl. The neutrophil decrease may be due to decreased production or increased peripheral destruction. The percentage of Neutrophils also appears in all of the groups. The cutoff50 values in the mild-moderate group are 84.80% ($p<0.00001$) with RR=3.18 [2.73-3.70] ($p<0.00001$), higher than the multiple variant RR about 0.40, the probability for the mild patients become moderate increase two times. In the moderate-severe group, the cutoff50 value is 87.74% ($p<0.00001$) with RR=3.32 [2.6480;4.1529] ($p<0.00001$), higher than the pooled RR about 2.04, the probability for the moderate patients become more severe increase 2.3 times. The cutoff50 value of the number of neutrophils in this group is 11.77G/L ($p<0.00001$) with RR=3.15 [2.6153;3.8025] ($p<0.00001$), higher than the pooled RR about 1.25, the probability for the moderate patients become more severe increase 2.15 times.

Substantial evidence has accumulated since the beginning of the COVID-19 pandemic that neutrophils play an essential role in the pathophysiology, particularly in those with severe disease courses. [10] The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) engages the inflammasome in monocytes and macrophages, leading to the cytokine storm. Neutrophils, the most abundant leukocytes, release neutrophil extracellular traps (NETs), which have been implicated in the pathogenesis of COVID-19. The recent study of Aymonnier *et al*. shows that activation of the NLRP3 inflammasome is essential for NET release in sterile inflammation. Their clinical study shows that in COVID-19 infection, ASC speck is formed in neutrophils, and the percentage of neutrophils showing a speck is similar to that seen in mononuclear cells. Given their vast presence in the blood and lung compared with other types of leukocytes, we can conclude that neutrophils could be the primary producers of specks. These findings support the possibility that neutrophils and monocytes contribute to the cytokine storm through the inflammasome, and in vivo, ASC speck detection could indicate innate immune system activation. [11] Bohnacker *et al.* showed in a recent study that the monocyte-derived macrophages (MDM) cause the inflammatory response to SARS-CoV-2 (severe acute respiratory

 doi:10.20944/preprints202209.0291.v1

syndrome coronavirus 2), and they are a significant source of eicosanoids in airway inflammation. RNA-sequencing (RNAseq) analysis determined 163 differentially expressed genes (DEGs) in MDM differentiated from monocytes of seropositive individuals 3–5 months post-infection compared to MDM from seronegative subjects. Post-COVID-19 MDM showed higher expression of pro-inflammatory chemokines (CCL2, CCL8, CCL7), driving neutrophil recruitment, including in COVID-19. [12]

The scoring index of X-ray (Brixia) appears in all of the groups. The cutoff50 values in the mild-moderate group are 5.53 (p<0.00001) with RR=3.29 [2.76-3.92] (p<0.00001), higher than the multiple variant RR about 0.51, the probability for the mild patients become moderate increase 4.5 times. In the moderate-severe group, the cutoff50 value is 10.51 (p<0.00001) with RR=3.03 [2.4023;3.8314] (p<0.00001), higher than the pooled data RR=1.28 [1.22;1.33] with z=11.23, p<0.0001 about 1.75, that means the probability for the moderate patients become more severe increase two times.

Integration of imaging and clinical parameters could improve the stratification of COVID-19 patients on emergency department (ED) admission. Borghesi et al. (2020) showed that according to receiver operating characteristic curve analyses, the optimal cut-off values for Brixia score and patient age were 8 points and 71 years, respectively.[8] Similar to Hoang et al. (2022) showed in the univariate analysis, age, vaccination status, previous disease, NEWS2, a saturation of peripheral oxygen (Sp02), the Brixia and TSS scores were significant predictors of mortality (p-value < 0.05). In multivariate analysis, there were statistically significant differences in mortality between age, $Sp0_2$, and Brixia score, and patients with previous diseases were independent predictors of mortality and hospitalization. [9]

C-reactive protein is produced by the liver, binding to the polysaccharide C of pneumococcus. CRP is one of the proteins released into the bloodstream in response to inflammation, which is considered an early marker of infection and inflammation. The C-reactive protein test is a quantitative blood test for CRP, which measures the overall level of inflammation in the body. Normal CRP levels are deficient. At the onset of inflammation, CRP in the blood rises rapidly within 6-8 hours. As the inflammation or tissue damage resolves, CRP levels drop. Thus, although CRP is a nonspecific indicator of inflammation, it is a valuable marker for monitoring disease severity. C reactive protein (CRP) has the cutoff50 values, which is 7.46mg/dL (p<0.00001) in the mild-moderate group with RR=3.4[2.91-3.97] (p<0.00001), higher than the multiple variant RR about 0.62, the probability for the mild patients become moderate increase 2.4 times. In the recent study of Faucheux et al. on two cohorts are COVID-19 hematological patients from France and Brazil during the pre-vaccination period, two profiles of patients were identified. One is young patients, with few comorbidities and low C-reactive protein (CRP), D-dimers, lactate dehydrogenase (LDH) and creatinine levels. And the other is the older patients, with several comorbidities and high levels of the four biology markers. The profiles were strongly associated with survival (p < 0.0001), even after adjusting for age (p = 0.0002).[13] The plasma CRP level is positively correlated to the severity of COVID-19 on CT performance, and a higher level of CRP showed a longer inpatient duration. For the first time, plasma CRP level is demonstrated to assist in discerning patients with moderate to severe COVID-19 pneumonia from those with mild conditions. This suggests CRP testing may be useful as an earlier indicator for severe illness and help physicians to stratify patients for intense care unit transfer.[14]

Albumin is an important protein, accounting for a large part of 58-74% of the total protein in the body. Produced in the liver, quantified at about 10.5g per day with the main function of preventing water from going out of blood vessels, maintaining colloidal osmotic pressure at a stable level, and also a bridge to bind and transport Fatty acids, steroid hormones, vitamins, bilirubin, and drugs go to every organ in the body. Albumin and ratio of lymphocytes show a strong negative link with the severity of Covid19. Their exact risk ratio (RR) based on the cutoff50s shows a large difference with the pooled RR. Albumin has the cutoff50 values, which is 29.73g/L (p<0.00001) in the moderate-severe group

 doi:10.20944/preprints202209.0291.v1

with RR=0.46 [0.3650;0.5752] (p<0.00001), lower than the multiple variant RR about 0.82, the probability for the moderate patients become severe decrease 0.54 times.

Hypo-albuminemia status has been associated with critically ill patients and mortality across numerous clinical settings. The pathophysiology behind hypo-albuminemia in the disease state is thought to be secondary to increased capillary permeability, decreased protein synthesis, the half-life of serum albumin, serum albumin total mass, increased volume of distribution, and increased expression of vascular endothelial growth factor. [15]

The lymphocyte index reflects the number of lymphocytes present in the body. This is essential in helping doctors understand the patient's condition and detect abnormalities early so they can diagnose and treat them promptly. In Normal people, the number of white blood cells usually has a very high range with an average number of 4 - 10 G/L with a percentage in the blood of 17-48%. The ratio of lymphocytes has the cutoff50 values, which is 6.32% (p<0.00001) in the moderate-severe group with RR=0.34 [0.2743;0.4210] (p<0.00001), lower than the multiple variant RR about 0.94, the probability for the moderate patients become severe decrease 0.66 times.

Yamasaki et al. show in the study in 2020 that Lymphocyte counts approximately six days after onset were significantly lower in the severe pneumonia group compared to both the non-severe pneumonia group and the non-pneumonia group (p = 0.0159, 0.0016, respectively). The severe pneumonia group had a low mean lymphocyte count at 659 cells/mm3 (SD 318.9). Patients in the severe pneumonia group were significantly older than those in the non-severe pneumonia group (p = 0.0079) but not significantly different from those in the non-pneumonia group. [16]

4.2. The significant inversion of correlation between the group of severity shows the important remark

Look more in detail the results, the important factors which are mentioned above show the week and medium link with other factor beside the severity (table 1,2,6). The most important shift is come from two pairs: Transferrin – anion Cl and ALT – Leucocytes.

The pair ALT – Leucocytes show the important negative link (R=-1, p<0.00001) in the mild group to the significant positive correlation in the moderate group (R=1, p<0.00001). The mean values of ALT and Leucocytes in each group (mild, moderate), respectively are: 81.92[19.00-132.00]U/I, 193.51[100.00-213.00]U/I, 335.37[102.00-109.00]U/I, 242.96[100.00-500.00]LEU/µL, 201.39[15.00-500.00] LEU/µL, 101.17[25.00-100.00] LEU/µL. We see the significant increase of ALT with p<0.001. ALT could take the stronger influence on the inversion than Leucocytes. The complete blood count (CBC) is the most common examination used to monitor overall health in clinical practice. Whether there is a relationship between CBC indexes and alanine transaminase (ALT) and aspartate aminotransferase (AST) has been unclear. Aminotransferase includes AST and ALT. They are markers of hepatocellular injury. They also correlate with obesity with a normal reference range higher in those with higher body mass index. [17] Leukocytes can be induced to express tissue factors and release pro-inflammatory and pro-coagulant molecules (granular enzymes, cytokines, and damage-associated molecular patterns). These intermediaries can influence all aspects of thrombus formation, including platelet activation, adhesion, and activation of the intrinsic and extrinsic coagulation pathways. [18]

The pair Transferrin – anion Chloride shows an important positive association (R=1, p<0.00001) in the mild group to a significant negative correlation in the moderate group (R=-0.59, p<0.00001). The mean values of Transferrin and anion Chloride in each group (mild, moderate), respectively are: 312.67 [255.50-354.00] mg/dL, 159.87 [141.00-181.50] mg/dL, 98.4 [97.00-101.00] mmol/L, 95.84 [93.00-99.00] mmol/L. The major function of human transferrin is to deliver iron from the bloodstream to actively dividing cells. Upon iron release, the protein changes its conformation from 'closed' to 'open.' The studies in vitro indicate that iron release from the transferrin is very complex and involves many factors, including pH, the chelator used, an anion effect, temperature, receptor binding, and intra-lobe interactions. The nature of the dual effect of chloride: the anion effect on

iron release is closely related to the strength of anion binding to the apo-protein. The negative effect seems to originate from competition between chloride and the chelator for an anion-binding site(s) near the metal center.[19] The least transferrin, the more anion chloride is released in plasma.

4.3. The important factors have the separated link with the severity of Covid19

The network map and HCA show the closest neighbors with the Covid19 severity in the network map of the mild-moderate group. The closest to the Covid19 severity are still ferritin, Age; then there is C-reactive protein, SI of X-ray, Albumin, and Lactate dehydrogenase, which are the next close neighbors of these three factors. That fit with the correlation result, in which these factors show a weak link with other subclinical indices. (Table 2) and exact correlation with the severity of covid19 (table 6).

We found different factors close to the severity in the moderate-severe group: Ferritin, Fibrinogen, Albumin, the quantity of Lymphocytes, SI of X-ray, white blood cells count, Lactate dehydrogenase, and quantity of neutrophils. The reasons we found in the whole group map these factors: Age, SI of X-ray, the number of Lymphocytes, Percentage of Neutrophils, Ferritin, and C-reactive protein. The distances between Age/Ferritin/quantity of Lymphocytes – Severity of Covid19 are shorter than the distance between Age/Ferritin/quantity of Lymphocytes with others. That shows the direct link between their four factors. The four others: SI of X-ray, the number of Lymphocytes, Percentage of Neutrophils, and C-reactive protein, also have a closer link than the severity.

5. Conclusion

Following visible successes on a wide range of predictive tasks, machine learning techniques are attracting substantial interest from medical researchers and clinicians. 3 Classifying data is a common task in machine learning. Suppose some given data points belong to one of two classes, and the goal is to decide which class a new data point will be in. Link to our Covid19, several missing points in every factor need to predict. Where is the importance of the complete multidimensional study in 2173 Covid19 patients in Vietnam? The answer will give the whole picture of all the preclinical factors. We could confirm they are not only dependent on the severity of patients. Otherwise, this multidimensional analysis study shows a possible link between several factors to give a clinical reference marker for surveillance and diagnostic management and enhance the prevention of severe outcomes.

Acknowledgments: We thank the participation of patients, healthcare workers, and technicians who participated and assisted us in completing this research.

Authors' contribution: TTN, Sample collection, data compilation and analysis, manuscript writing and review. CTH, study design, data compilation and analysis, manuscript writing and review. HTTB, data compilation and analysis and review. UP, Supervisor. VTT, Project Leader. All authors read and approved the final manuscript.

Conflicts of interest: All authors declare that they have no conflicts of interest.

Funding: The project: "Database study on SARS-Cov2 genome in Vietnam" (number: DTDLCN.102/21, MOST VIETNAM) was recognized and accepted by the Institution Review Board for Ethics in Biomedical Research of Hanoi Medical University's institutional ethics council (Number 546/HMU-IRD), and all participants' rights were protected. Ethics approval and consent to participate

The project: "Database study on SARS-Cov2 genome in Vietnam" (number: DTDLCN.102/21, MOST VIETNAM) was recognized and accepted by the Institution Review Board for Ethics in Biomedical Research of Hanoi Medical University's institutional ethics council (Number 546/HMU-IRD), and all participants' rights were protected.

Reference:

1. Nguyen NH, Van Nguyen T, Nguyen AQ, Van Nguyen P, Nguyen TNM. The first cohort of the COVID-19 patients in Vietnam and the national response to the pandemic. *Int J Med Sci*. 2020;17(16):2449-2453. doi:10.7150/ijms.45449

2. Magoulas GD, Prentza A. Machine Learning in Medical Applications. In: Paliouras G, Karkaletsis V, Spyropoulos CD, eds. *Machine Learning and Its Applications: Advanced Lectures*. Springer Berlin Heidelberg; 2001:300-307. doi:10.1007/3-540-44673-7_19

3. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Medical Research Methodology*. 2019;19(1):64. doi:10.1186/s12874-019-0681-4

4. Daren SS, Josh T. *The Basic Practice of Statistics*.; 2018.

5. Charrad M, Ghazzali N, Boiteau V, Niknafs A. NbClust: An R Package for Determining the Relevant Number of Clusters in a Data Set. *J Stat Soft*. 2014;61(6):1-36. doi:10.18637/jss.v061.i06

6. Cox, T.F., Cox, M.A.A. *Multidimensional Scaling*. Chapman and Hall.

7. Zimmermann P, Curtis N. Why Does the Severity of COVID-19 Differ With Age?: Understanding the Mechanisms Underlying the Age Gradient in Outcome Following SARS-CoV-2 Infection. *Pediatr Infect Dis J*. 2022;41(2):e36-e45. doi:10.1097/INF.0000000000003413

8. Monaco CG, Zaottini F, Schiaffino S, et al. Chest x-ray severity score in COVID-19 patients on emergency department admission: a two-centre study. *Eur Radiol Exp*. 2020;4(1):68. doi:10.1186/s41747-020-00195-w

9. Hoang SV, Nguyen KM, Huynh TM, Huynh KLA, Nguyen PH, Tran HPN. Chest X-ray Severity Score as a Putative Predictor of Clinical Outcome in Hospitalized Patients: An Experience From a Vietnamese COVID-19 Field Hospital. *Cureus*. 2022;14(3):e23323. doi:10.7759/cureus.23323

10. Reusch N, De Domenico E, Bonaguro L, et al. Neutrophils in COVID-19. *Front Immunol*. 2021;12:652470. doi:10.3389/fimmu.2021.652470

11. Aymonnier K, Ng J, Fredenburgh LE, et al. Inflammasome activation in neutrophils of patients with severe COVID-19. *Blood Adv*. 2022;6(7):2001-2013. doi:10.1182/bloodadvances.2021005949

12. Bohnacker S, Hartung F, Henkel F, et al. Mild COVID-19 imprints a long-term inflammatory eicosanoid- and chemokine memory in monocyte-derived macrophages. *Mucosal Immunology*. 2022;15(3):515-524. doi:10.1038/s41385-021-00482-8

13. Faucheux L, Bassolli de Oliveira Alves L, Chevret S, Rocha V. Comparison of characteristics and laboratory tests of COVID-19 hematological patients from France and Brazil during the pre-vaccination period: identification of prognostic profiles for survival. *Hematology, Transfusion and Cell Therapy*. doi:10.1016/j.htct.2022.05.003

14. Chen W, Zheng KI, Liu S, Yan Z, Xu C, Qiao Z. Plasma CRP level is positively associated with the severity of COVID-19. *Annals of Clinical Microbiology and Antimicrobials*. 2020;19(1):18. doi:10.1186/s12941-020-00362-2

15. Aziz M, Fatima R, Lee-Smith W, Assaly R. The association of low serum albumin level with severe COVID-19: a systematic review and meta-analysis. *Critical Care*. 2020;24(1):255. doi:10.1186/s13054-020-02995-3

16. Yamasaki Y, Ooka S, Tsuchida T, et al. The peripheral lymphocyte count as a predictor of severe COVID-19 and the effect of treatment with ciclesonide. *Virus Research*. 2020;290:198089. doi:10.1016/j.virusres.2020.198089

17. Yu J, Pan Q, Yang J, et al. Correlations of Complete Blood Count with Alanine and Aspartate Transaminase in Chinese Subjects and Prediction Based on Back-Propagation Artificial Neural Network (BP-ANN). *Med Sci Monit*. 2017;23:3001-3009. doi:10.12659/msm.901202

18. Swystun LL, Liaw PC. The role of leukocytes in thrombosis. *Blood*. 2016;128(6):753-762. doi:10.1182/blood-2016-05-718114

 doi:10.20944/preprints202209.0291.v1

19. He QY, Mason AB, Nguyen V, MacGillivray RT, Woodworth RC. The chloride effect is related to anion binding in determining the rate of iron release from the human transferrin N-lobe. *Biochem J*. 2000;350 Pt 3(Pt 3):909-915.