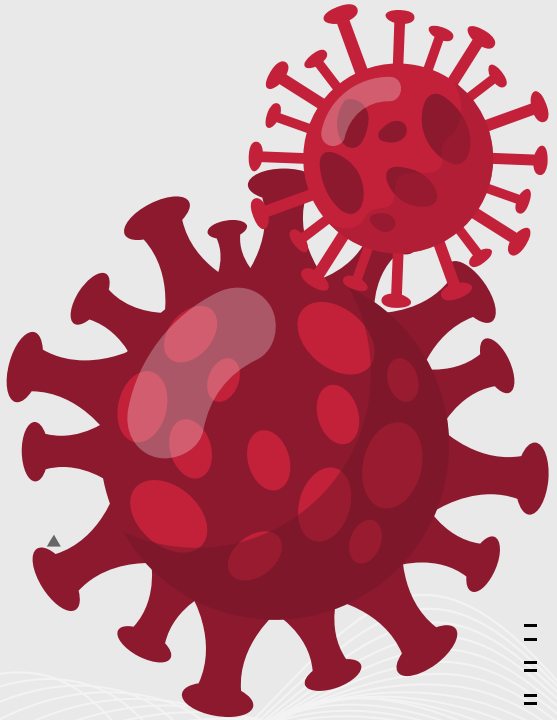
COVID-19 MULTI-OMIC ANALYSIS



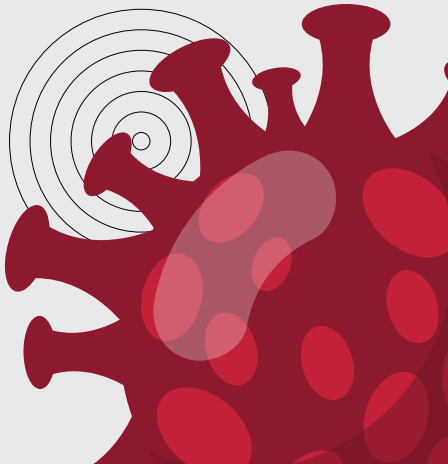
LARGE-SCALE MULTI-OMIC ANALYSIS OF COVID-19 SEVERITY

MOTIVATION

CLINICAL PRESENTATION OF COVID-19 INFECTION IS VERY DIVERSED

ITS SEVERITY DEPENDS ON VIRAL FACTORS, BUT ALSO THE HOST

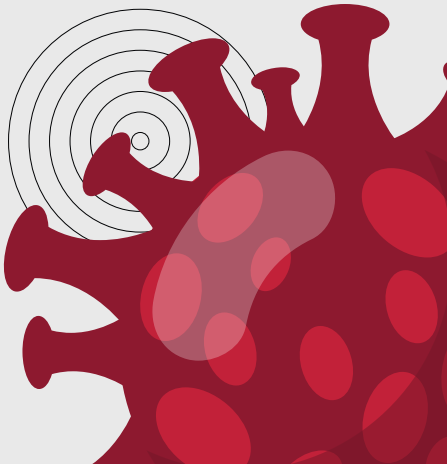
EVIDENCE INDICATES A POSSIBILE DIFFERENCIATION BETWEEN A POSITIVE AND A NEGATIVE DIAGNOSIS BASED ON GENETIC, PATOLOGIC AND CLINICAL MARKERS



GOALS

TO UNDERSTAND HOST RESPONSES IN A MOLECULAR LEVEL THAT MIGHT EXPLAIN DIFFERENT CLINICAL PRESENTATIONS

TO CREATE A MACHINE LEARNING MODEL THAT ALLOWS TO PREDICT THE SEVERITY OF THE COVID-19 INFECTION BASED ON OMIC, PHISIOLOGICAL AND CLINICAL DATA

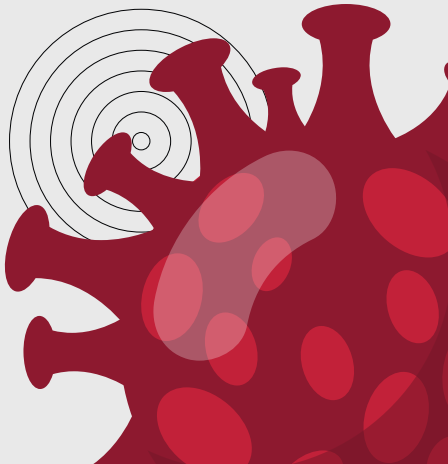


DATASET

GEO ACCESS CODE: GSE157103

CLINICAL DATA FROM 102 COVID-19 POSITIVE PATIENTS AND 26 COVID-19 NEGATIVE PATIENTS WITH RESPIRATORY ISSUES

LABORATORY AND GENOMIC ANALYSIS FROM PACIENTS'S BLOOD SAMPLES



RNA SEQ DATA

#symbol	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	...	NC17	NC18	NC19	NC20	NC21	NC22	NC23	NC24	NC25	NC26
A1BG	0.49	0.29	0.26	0.45	0.17	0.21	0.49	0.12	0.51	0.10	...	0.37	0.33	0.25	0.20	0.40	0.30	0.70	0.75	2.80	0.22
A1CF	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.01	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
A2M	0.21	0.14	0.03	0.09	0.00	0.08	0.23	0.08	0.88	0.13	...	0.07	0.06	0.11	0.01	0.04	0.02	0.02	0.27	0.04	0.28
A2ML1	0.04	0.00	0.02	0.07	0.05	0.04	0.03	0.01	0.02	0.01	...	0.01	0.00	0.00	0.02	0.02	0.02	0.01	0.00	0.00	0.00
A3GALT2	0.07	0.00	0.00	0.00	0.07	0.00	0.07	0.00	0.79	0.15	...	0.18	0.00	0.04	0.00	0.00	0.00	0.00	0.06	0.00	0.00
A4GALT	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.03	0.07	0.00	0.00	0.00	0.00	0.00	0.00
A4GNT	0.03	0.05	0.07	0.00	0.00	0.03	0.00	0.00	0.00	0.12	...	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00
AAAS	18.92	18.68	13.85	22.11	8.45	19.60	28.59	10.50	22.78	15.47	...	16.31	10.98	11.28	13.57	24.83	17.06	20.31	27.25	21.64	5.54
AACS	4.07	3.00	1.83	4.22	1.17	3.15	4.24	2.10	4.86	2.90	...	2.47	1.60	1.30	2.00	2.74	1.96	1.54	3.42	1.43	0.35
AADAC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	...	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

10 rows × 126 columns

19472, 126

METADATA

```
Index(['title', 'geo_accession', 'status', 'submission_date',
       'last_update_date', 'type', 'channel_count', 'source_name_ch1',
       'organism_ch1', 'taxid_ch1', 'characteristics_ch1.0.disease state',
       'characteristics_ch1.1.age (years)', 'characteristics_ch1.2.Sex',
       'characteristics_ch1.3.icu', 'characteristics_ch1.4.apacheii',
       'characteristics_ch1.5.charlson score',
       'characteristics_ch1.6.mechanical ventilation',
       'characteristics_ch1.7.ventilator-free days',
       'characteristics_ch1.8.dm',
       'characteristics_ch1.9.hospital-free days post 45 day followup (days)',
       'characteristics_ch1.10.ferritin (ng/ml)',
       'characteristics_ch1.11.crp (mg/l)',
       'characteristics_ch1.12.ddimer (mg/l_feu)',
       'characteristics_ch1.13.procalcitonin (ng/ml)',
       'characteristics_ch1.14.lactate (mmol/l)',
       'characteristics_ch1.15.fibrinogen', 'characteristics_ch1.16.sofa',
       'characteristics_ch1.17.associated controlled-vocabulary terms',
       'characteristics_ch1.18.cell type',
       'characteristics_ch1.19.tissue of origin',
       'characteristics_ch1.20.sample type', 'treatment_protocol_ch1',
       'growth_protocol_ch1', 'molecule_ch1', 'extract_protocol_ch1',
       'description', 'data_processing', 'platform_id', 'contact_name',
       'contact_email', 'contact_department', 'contact_institute',
       'contact_address', 'contact_city', 'contact_state',
       'contact_zip/postal_code', 'contact_country', 'instrument_model',
       'library_selection', 'library_source', 'library_strategy', 'relation',
       'supplementary_file_1', 'series_id', 'data_row_count',
       'characteristics_ch1.14.fibrinogen', 'characteristics_ch1.15.sofa',
       'characteristics_ch1.16.associated controlled-vocabulary terms',
       'characteristics_ch1.17.cell type',
       'characteristics_ch1.18.t                                    ',
       'characteristics_ch1.19.s                                    ',
      dtype='object')
```

126, 61

				ource_name_ch1	organism_ch1
GSM4753021	COVID_01_39y			Leukocytes from whole blood	Homo sapiens
GSM4753022	COVID_02_63y			Leukocytes from whole blood	Homo sapiens
GSM4753023	COVID_03_33y			Leukocytes from whole blood	Homo sapiens

RNA SEQ DATA DESCRIPTION

#symbol	count	mean	std	min	25%	50%	75%	max
A1BG	126.0	0.362381	0.301175	0.05	0.2100	0.310	0.4200	2.80
A1CF	126.0	0.002222	0.004714	0.00	0.0000	0.000	0.0000	0.02
A2M	126.0	0.123651	0.136109	0.00	0.0400	0.090	0.1500	0.88
A2ML1	126.0	0.014127	0.019111	0.00	0.0000	0.010	0.0200	0.10
A3GALT2	126.0	0.129206	0.231398	0.00	0.0000	0.050	0.1675	1.75
...	...	...	...	...	...	...	...	...
ZYG11A	126.0	0.155794	0.075877	0.02	0.1000	0.140	0.2000	0.37
ZYG11B	126.0	5.195238	2.542537	0.65	3.2600	4.815	7.1250	12.42
ZYX	126.0	375.867063	147.352997	115.92	276.8475	354.515	455.0200	806.75
ZZEF1	126.0	9.408016	4.535389	0.74	6.2800	8.465	12.4700	19.77
ZZZ3	126.0	3.075476	1.997453	0.27	1.7825	2.555	3.8250	11.01

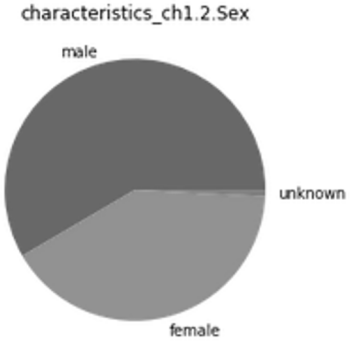
19472 rows × 8 columns

METADATA DESCRIPTION

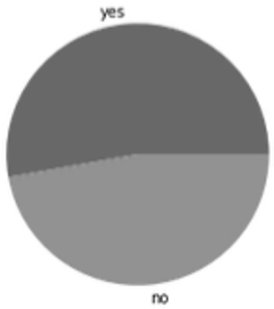
	title	geo_accession	status	submission_date	last_update_date	type	channel_count	source_name_ch1	organism_ch1	taxid_
count	126	126	126	126	126	126	126	126	126	
unique	126	126	1	1	1	1	1	1	1	
top	COVID_83_85y_female_NonICU	GSM4753107	Public on Aug 29 2020	Aug 28 2020	Aug 29 2020	SRA	1	Leukocytes from whole blood	Homo sapiens	9
freq	1	1	126	126	126	126	126	126	126	

SAMPLE COHORT

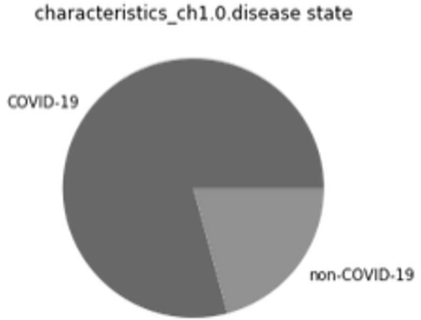
Counter({'male': 73, 'female': 51, 'unknown': 1})



Counter({'yes': 66, 'no': 59})



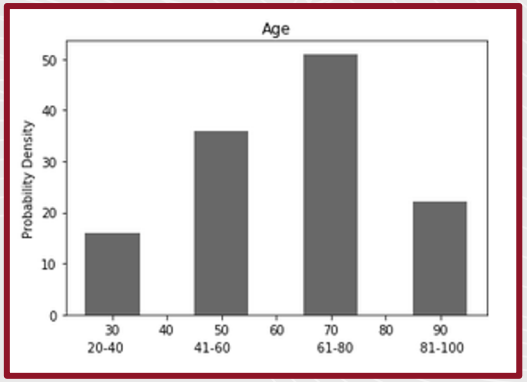
Counter({'COVID-19': 99, 'non-COVID-19': 26})

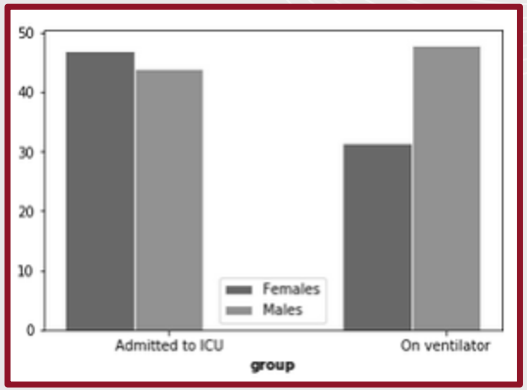


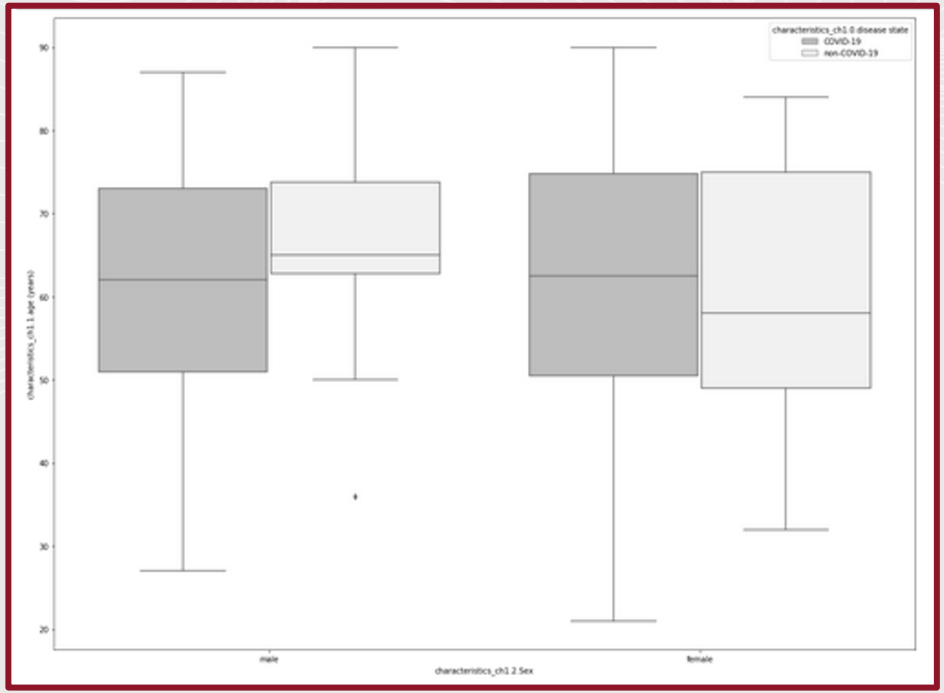
SAMPLE COHORT

```
characteristics_ch1.2.Sex  characteristics_ch1.0.disease state
female                     COVID-19                              38
                           non-COVID-19                          13
male                       COVID-19                              61
                           non-COVID-19                          12
unknown                    non-COVID-19                           1
dtype: int64

characteristics_ch1.0.disease state  characteristics_ch1.3.icu
COVID-19                             no                        49
                                     yes                       50
non-COVID-19                         no                        10
                                     yes                       16
dtype: int64

characteristics_ch1.3.icu  characteristics_ch1.6.mechanical ventilation
no                         no                                          54
                           yes                                          5
yes                        no                                          20
                           yes                                         46
dtype: int64
```

SAMPLE COHORT







STATISTICAL ANALYSIS

FISHER'S EXACT TEST

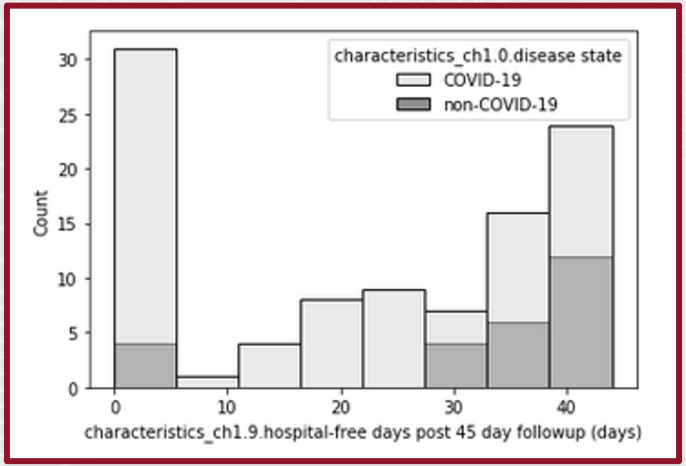
(0.6802721088435374, 0.5025623499870514)

```
check Age distr; COVID-19/non-COVID-19 groups
shapiro
COVID-19 (0.969968318939209, 0.02303556352853775)
non-COVID-19 (0.9655020833015442, 0.5111427307128906)
MannwhitneyuResult(statistic=1169.0, pvalue=0.2372767989887578)
```

```
check Age distr; female COVID-19/female non-COVID-19; male COVID-19/male non-COVID-19; female COVID-19/male COVI
D-19; female non-COVID-19/male non-COVID-19
shapiro
COVID-19 + Female (0.9534679651260376, 0.11588154733181)
non-COVID-19 + Female (0.9619491696357727, 0.7834566831588745)
COVID-19 + male (0.9731571674346924, 0.19940562546253204)
non-COVID-19 + male (0.9297250509262085, 0.3772066831588745)

ttest
COVID19 female/COVID19 male: Ttest_indResult(statistic=-0.5167418333477375, pvalue=0.606512568083234)
nonCOVID19 female/nonCOVID19 male: Ttest_indResult(statistic=-1.1834691341842989, pvalue=0.2487115499521349 6)
COVID19 female/nonCOVID19 female: Ttest_indResult(statistic=0.09977742688908672, pvalue=0.9209281394772523)
COVID19 male/nonCOVID19 male: Ttest_indResult(statistic=-1.1234425724835997, pvalue=0.26503392638932777)
```

SEVERITY SCORES

HOSPITAL-FREE DAYS
POST 45 DAY (HFD-45)



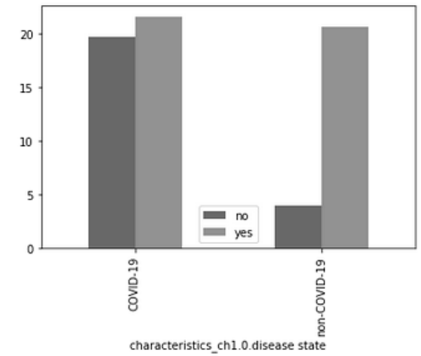

```
Teste à normalidade dos dados:
COVID-19 (0.8425031900405884, 6.32884189499805e-09)
non-COVID-19 (0.6970590353012085, 4.834085302718449e-06)
Rejeita-se H0, uma vez que as amostras não seguem distribuição nor
mal.

Teste U de Mann-Whitney
MannwhitneyuResult(statistic=791.5, pvalue=0.0009940081528354407)
Rejeita-se H0, os dois grupos não têm a mesma mediana.
```

SEVERITY SCORES

SCORES CHARLSON, APACHEII E SOFA

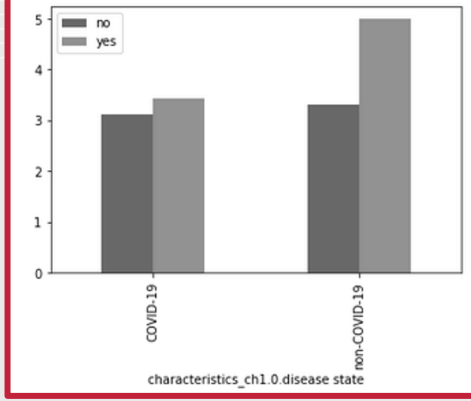
characteristics_ch1.4.apacheii

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	19.714286	21.580
non-COVID-19	4.000000	20.625



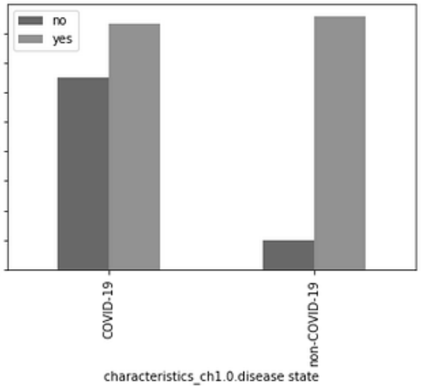
characteristics_ch1.5.charlson score

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	3.12	3.44
non-COVID-19	3.30	5.00



characteristics_ch1.16.sofa

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	6.5	8.3000
non-COVID-19	1.0	8.5625



SEVERITY SCORES

STATISTICAL ANALYSIS

```
Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.4.apacheii (0.9830883741378784, 0.687343418598175)
non-COVID-19 / characteristics_ch1.4.apacheii (0.92143315076828, 0.177878737449646)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.5.charlson score (0.8693323731422424, 5.393660831032321e-05)
non-COVID-19 / characteristics_ch1.5.charlson score (0.9138711094856262, 0.13436259329319)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.16.sofa (0.9536594152450562, 0.04831695556640625)
non-COVID-19 / characteristics_ch1.16.sofa (0.9044427871704102, 0.09470410645008087)
```

```
Teste U de Mann-Whitney
MannwhitneyuResult(statistic=227.0, pvalue=0.004494960710277233) characteristics_ch1.5.charlson score
MannwhitneyuResult(statistic=365.0, pvalue=0.3022069285035096) characteristics_ch1.16.sofa
ttest
Ttest_indResult(statistic=-0.22080066654865027, pvalue=0.8259505410977093) characteristics_ch1.4.apacheii
```

BIOMARKERS

```
characteristics_ch1.10.ferritin (ng/ml) 833.5181818181818

 characteristics_ch1.0.disease state
COVID-19         932.755319
non-COVID-19     250.500000
Name: characteristics_ch1.10.ferritin (ng/ml), dtype: float64


characteristics_ch1.11.crp (mg/l) 131.17570093457948

 characteristics_ch1.0.disease state
COVID-19         140.538043
non-COVID-19      73.753333
Name: characteristics_ch1.11.crp (mg/l), dtype: float64


characteristics_ch1.12.ddimer (mg/l_feu) 10.765999999999996

 characteristics_ch1.0.disease state
COVID-19         11.722000
non-COVID-19      5.348667
Name: characteristics_ch1.12.ddimer (mg/l_feu), dtype: float64
```

```
characteristics_ch1.13.procalcitonin (ng/ml) 3.0777450980392174

 characteristics_ch1.0.disease state
COVID-19         3.242989
non-COVID-19     2.119333
Name: characteristics_ch1.13.procalcitonin (ng/ml), dtype: float64


characteristics_ch1.14.lactate (mmol/l) 1.462470588235294

 characteristics_ch1.0.disease state
COVID-19         1.240476
non-COVID-19     2.098182
Name: characteristics_ch1.14.lactate (mmol/l), dtype: float64


characteristics_ch1.15.fibrinogen 516.516129032258

 characteristics_ch1.0.disease state
COVID-19         543.848101
non-COVID-19     362.285714
Name: characteristics_ch1.15.fibrinogen, dtype: float64
```

BIOMARKERS

STATISTICAL ANALYSIS

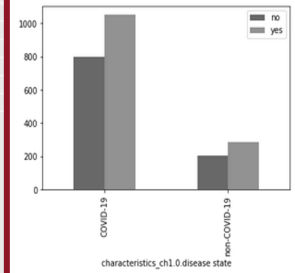
```
Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.10.ferritin (ng/ml) (0.6724420189857483, 3.5811387492576396e-13)
non-COVID-19 / characteristics_ch1.10.ferritin (ng/ml) (0.8367587924003601, 0.008751677349209785)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.11.crp (mg/l) (0.9439736604690552, 0.0006211342988535762)
non-COVID-19 / characteristics_ch1.11.crp (mg/l) (0.861149787902832, 0.025073396041989326)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.12.ddimer (mg/l_feu) (0.5426771640777588, 7.535703154147241e-15)
non-COVID-19 / characteristics_ch1.12.ddimer (mg/l_feu) (0.6030751466751099, 2.7178120944881812e-05)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.13.procalcitonin (ng/ml) (0.30038517713546753, 2.0670906431600546e-18)
non-COVID-19 / characteristics_ch1.13.procalcitonin (ng/ml) (0.4998037815093994, 3.5043137813772773e-06)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.14.lactate (mmol/l) (0.8686490058898926, 7.334798738156678e-06)
non-COVID-19 / characteristics_ch1.14.lactate (mmol/l) (0.6681157350540161, 7.897288924141321e-06)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.15.fibrinogen (0.9711021184921265, 0.06983786821365356)
non-COVID-19 / characteristics_ch1.15.fibrinogen (0.7982048988342285, 0.004742539953440428)
```

```
Teste U de Mann-Whitney
MannwhitneyuResult(statistic=304.0, pvalue=7.410975126923796e-05) characteristics_ch1.10.ferritin (ng/ml)
MannwhitneyuResult(statistic=418.0, pvalue=0.007421319689347778S) characteristics_ch1.11.crp (mg/l)
MannwhitneyuResult(statistic=518.5, pvalue=0.12632120997310708) characteristics_ch1.12.ddimer (mg/l_feu)
MannwhitneyuResult(statistic=576.5, pvalue=0.237717239036226) characteristics_ch1.13.procalcitonin (ng/ml)
MannwhitneyuResult(statistic=526.0, pvalue=0.047377184070242O9) characteristics_ch1.14.lactate (mmol/l)
MannwhitneyuResult(statistic=256.5, pvalue=0.0007349754311390S5) characteristics_ch1.15.fibrinogen
```

BIOMARKERS

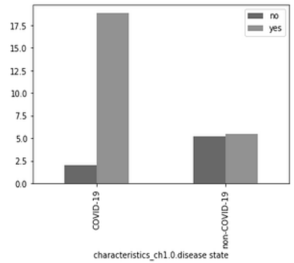
characteristics_ch1.10.ferritin (ng/ml)

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	796.954545	1052.260000
non-COVID-19	205.285714	285.666667



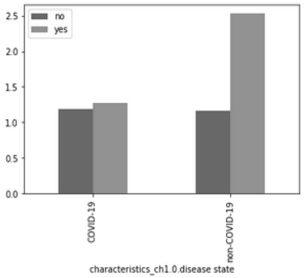
characteristics_ch1.12.ddimer (mg/l_feu)

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	1.979167	18.880000
non-COVID-19	5.212500	5.504286



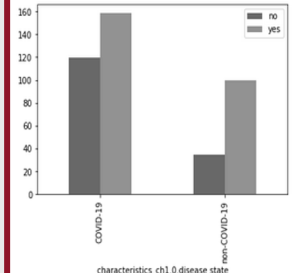
characteristics_ch1.14.lactate (mmol/l)

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	1.186087	1.27175
non-COVID-19	1.164286	2.53400



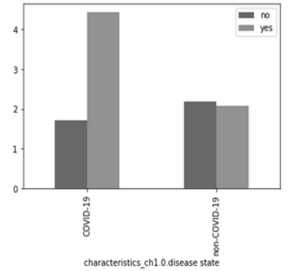
characteristics_ch1.11.crp (mg/l)

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	119.818605	158.720408
non-COVID-19	34.733333	99.766667



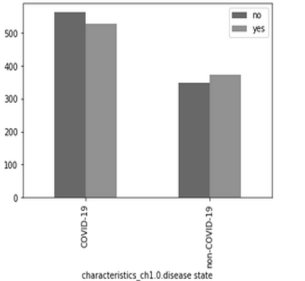
characteristics_ch1.13.procalcitonin (ng/ml)

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	1.707105	4.434082
non-COVID-19	2.191667	2.071111



characteristics_ch1.15.fibrinogen

characteristics_ch1.3.icu	no	yes
characteristics_ch1.0.disease state		
COVID-19	563.558824	528.955556
non-COVID-19	348.000000	373.000000



BIOMARKERS

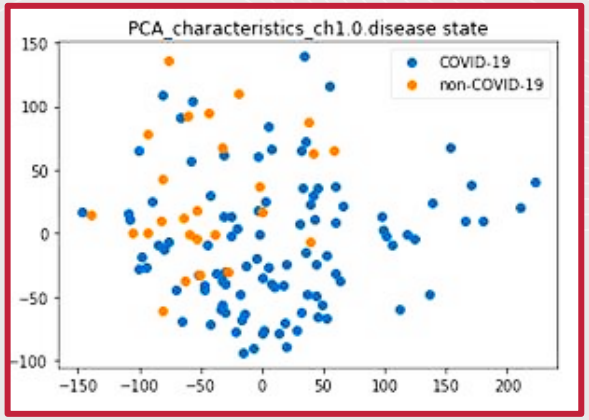
STATISTICAL ANALYSIS

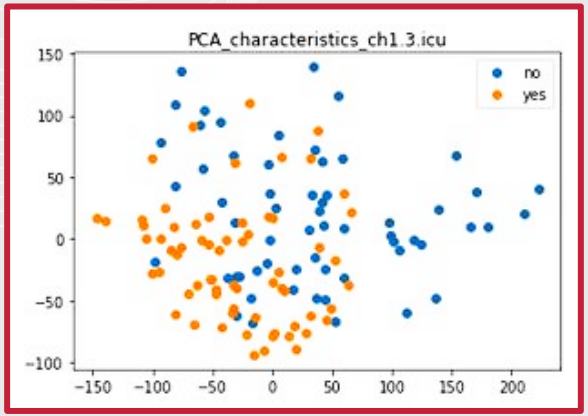
```
Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.10.ferritin (ng/ml) (0.7296851873397827, 2.939570187265872e-08)
non-COVID-19 / characteristics_ch1.10.ferritin (ng/ml) (0.8502141833305359, 0.0749269500374794)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.11.crp (mg/l) (0.9528987407684326, 0.048432618379592896)
non-COVID-19 / characteristics_ch1.11.crp (mg/l) (0.9312303066253662, 0.49314913153648376)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.12.ddimer (mg/l_feu) (0.6747757196426392, 3.824088157244887e-09)
non-COVID-19 / characteristics_ch1.12.ddimer (mg/l_feu) (0.8394935131072998, 0.09828334301710129)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.13.procalcitonin (ng/ml) (0.33338654041290283, 1.3885003173272348e-13)
non-COVID-19 / characteristics_ch1.13.procalcitonin (ng/ml) (0.4710007309913635, 2.9137365800124826e-06)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.14.lactate (mmol/l) (0.9343225359916687, 0.022340837866067886)
non-COVID-19 / characteristics_ch1.14.lactate (mmol/l) (0.7340185642242432, 0.0005886571598239243)

Teste à normalidade dos dados:
COVID-19 / characteristics_ch1.15.fibrinogen (0.9738714694976807, 0.39675429463386536)
non-COVID-19 / characteristics_ch1.15.fibrinogen (0.8005968332290649, 0.029049644246697426)
```

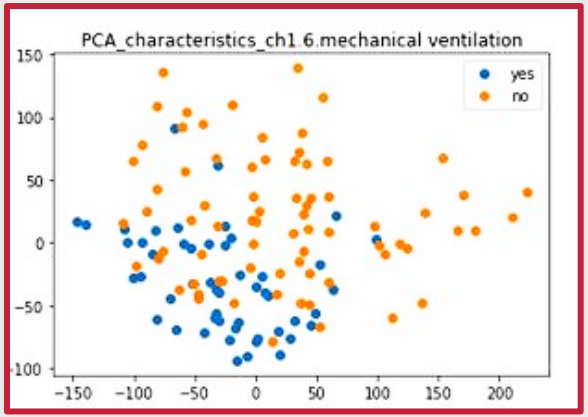
```
Teste U de Mann-Whitney
MannwhitneyuResult(statistic=70.0, pvalue=0.000562692682917295) characteristics_ch1.10.ferritin (ng/ml)
MannwhitneyuResult(statistic=117.0, pvalue=0.09047272300856335) characteristics_ch1.12.ddimer (mg/l_feu)
MannwhitneyuResult(statistic=98.5, pvalue=0.022068584395312O3) characteristics_ch1.15.fibrinogen
ttest
Ttest_indResult(statistic=1.559438709104518, pvalue=0.12452661754344743) characteristics_ch1.11.crp (mg/l)
Ttest_indResult(statistic=0.5402356178160989, pvalue=0.5911766977694919) characteristics_ch1.13.procalcitonin (n
g/ml)
Ttest_indResult(statistic=-3.2745173761918296, pvalue=0.001868200032359405) characteristics_ch1.14.lactate (mmol
/l)
```

NON-SUPERVISED ANALYSIS

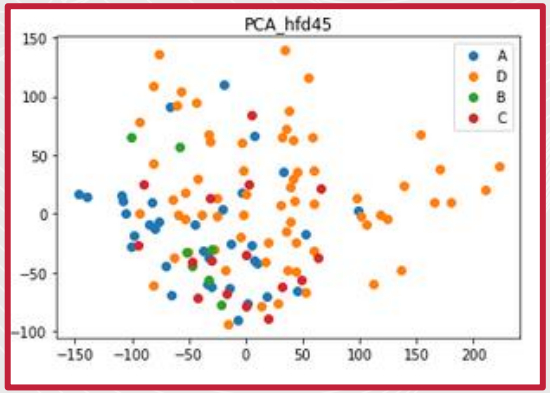
PCA – PRINCIPAL COMPONENT ANALYSIS

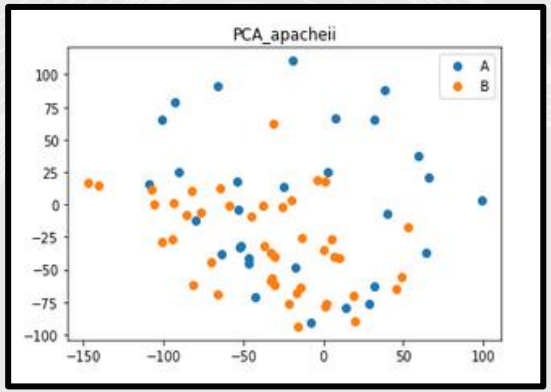


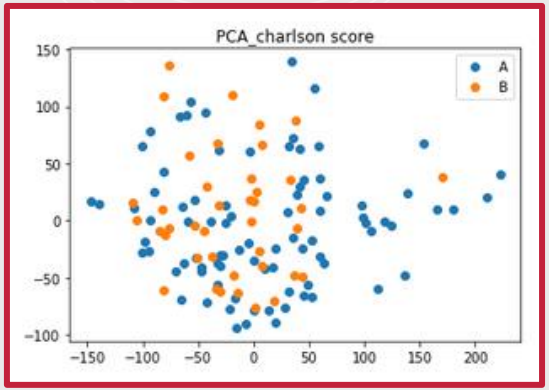


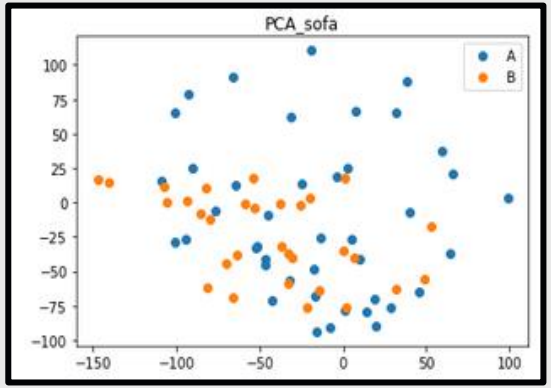


NON-SUPERVISED ANALYSIS

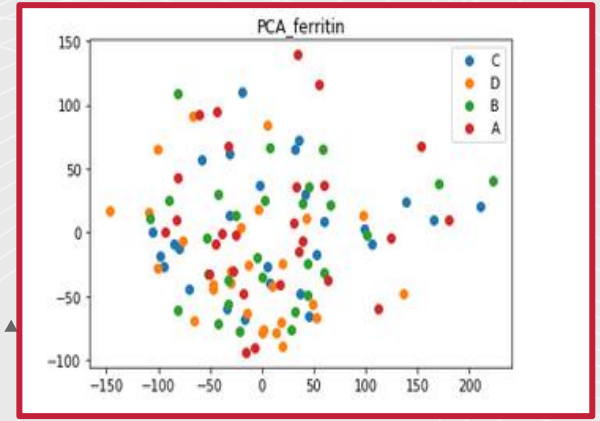


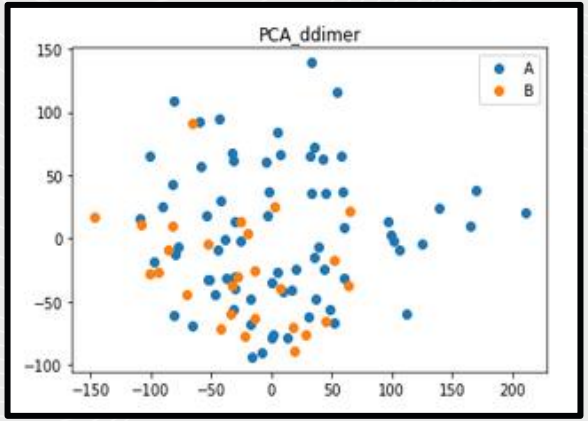


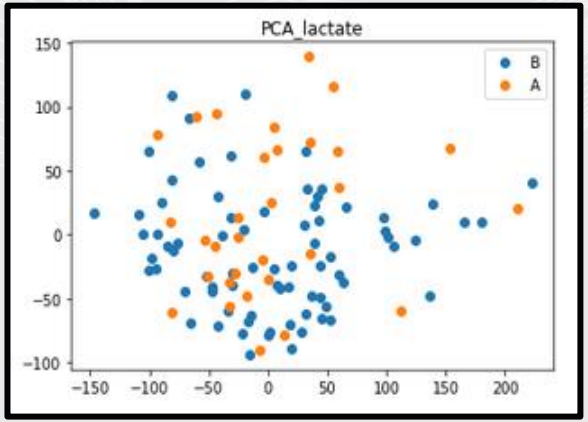


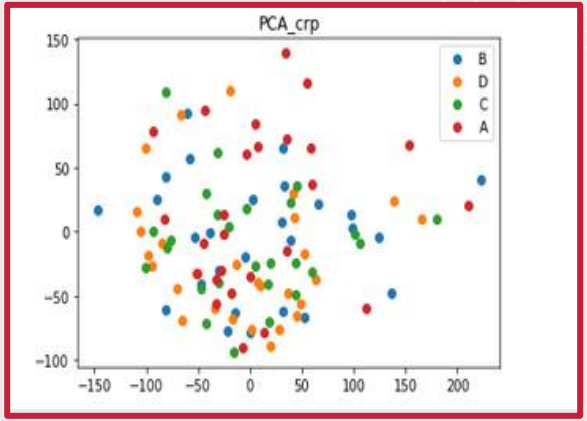


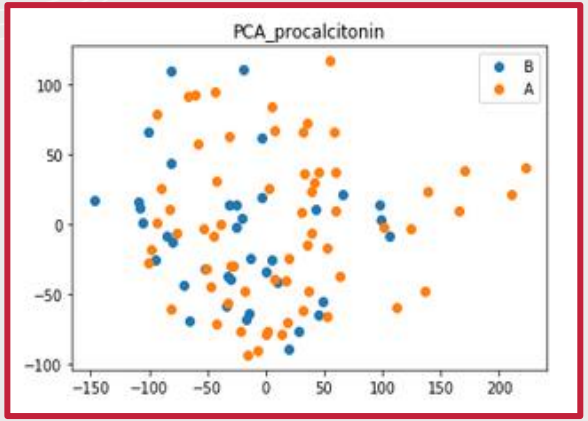
NON-SUPERVISED ANALYSIS

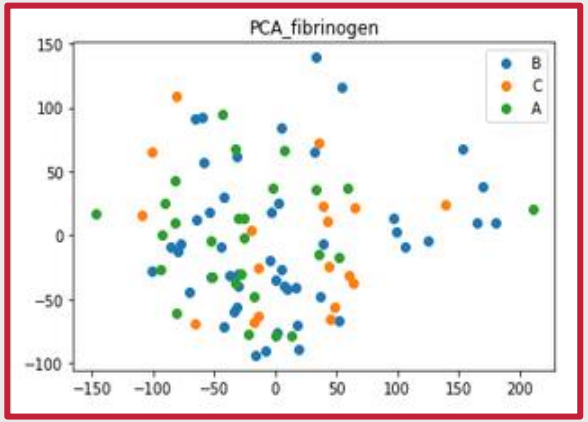






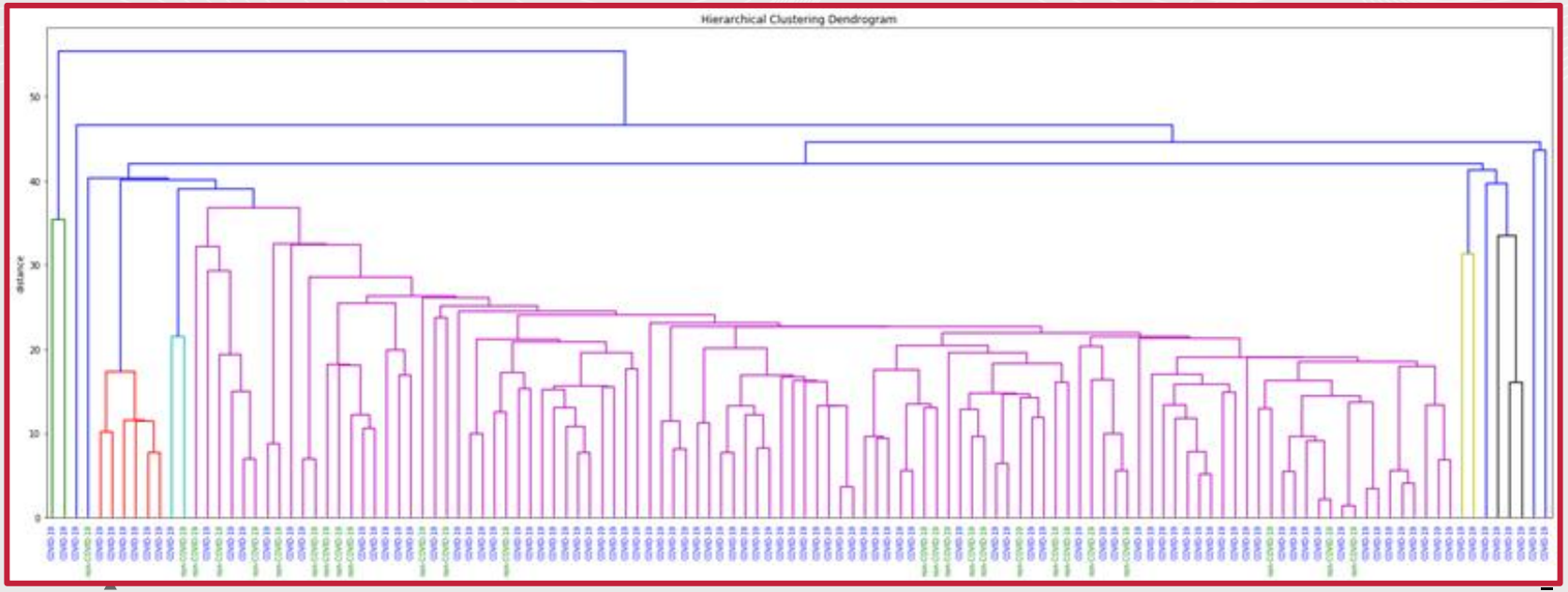






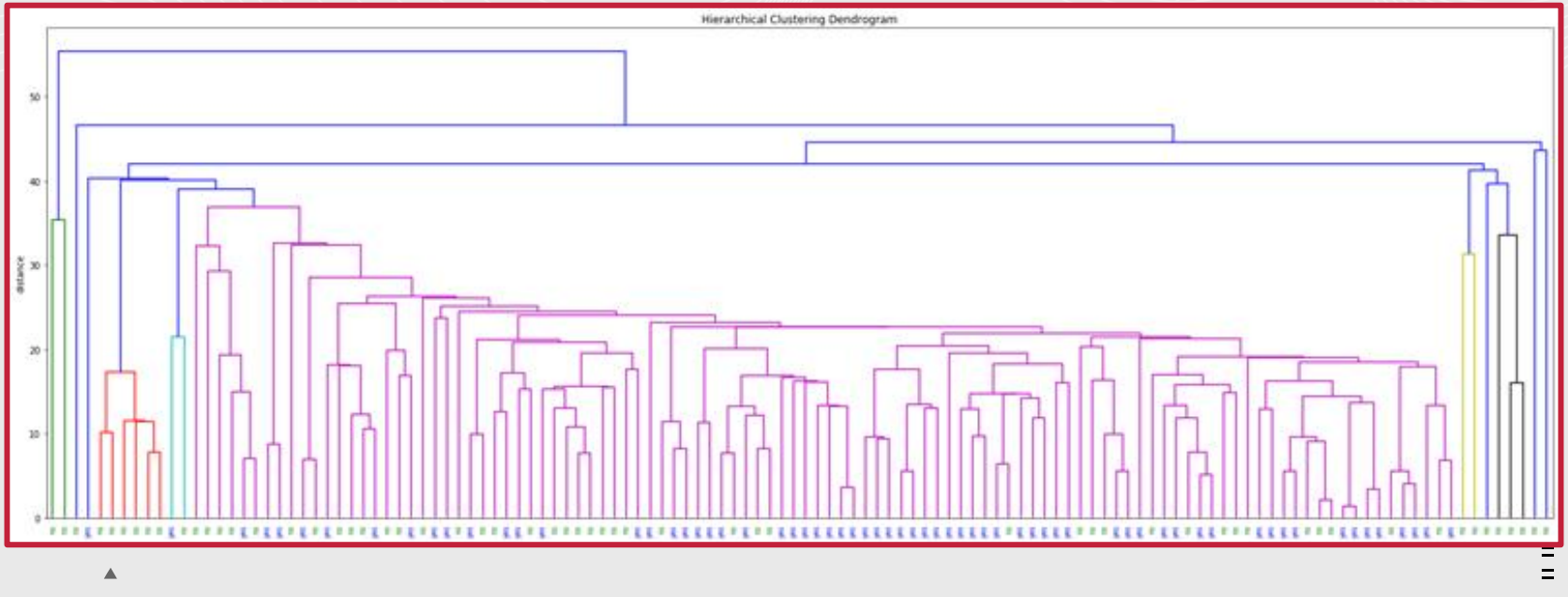
NON-SUPERVISED ANALYSIS

HIERARCHICAL CLUSTERING - DISEASE STATE



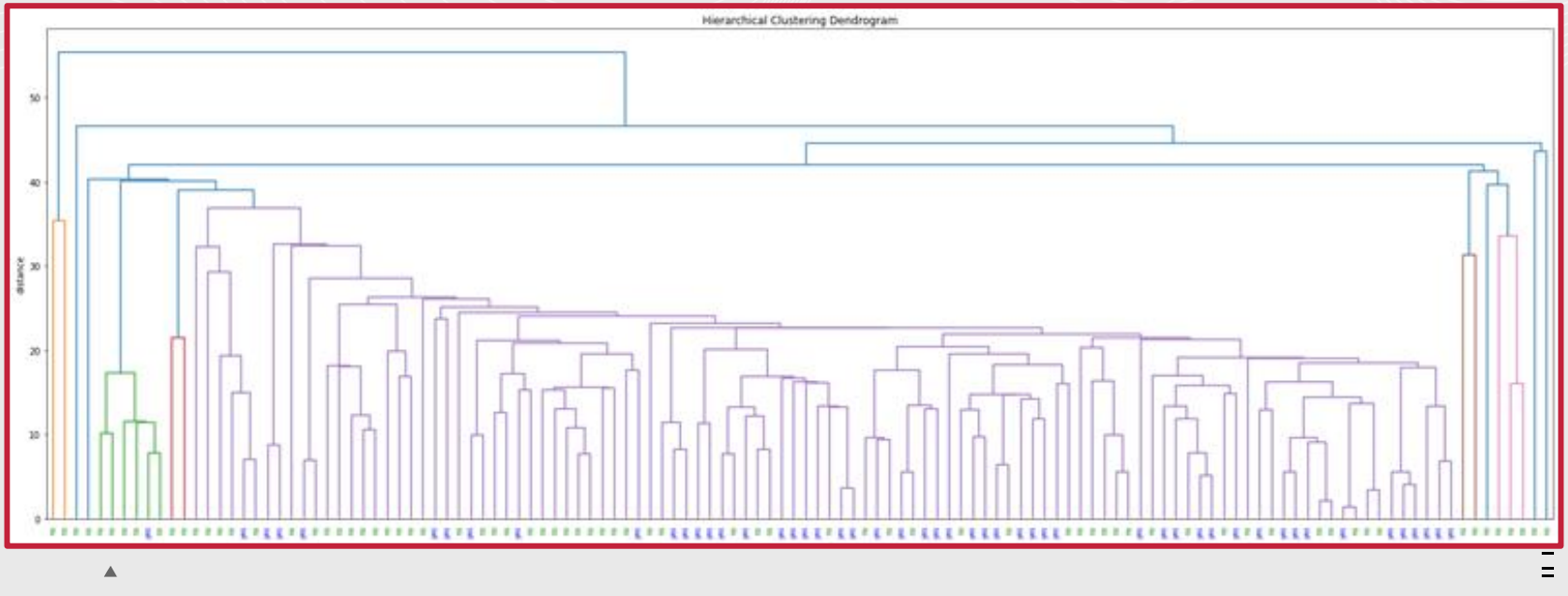
NON-SUPERVISED ANALYSIS

HIERARCHICAL CLUSTERING – ICU



NON-SUPERVISED ANALYSIS

HIERARCHICAL CLUSTERING – MECHANICAL VENTILATION



NON-SUPERVISED ANALYSIS

K-MEANS CLUSTERING

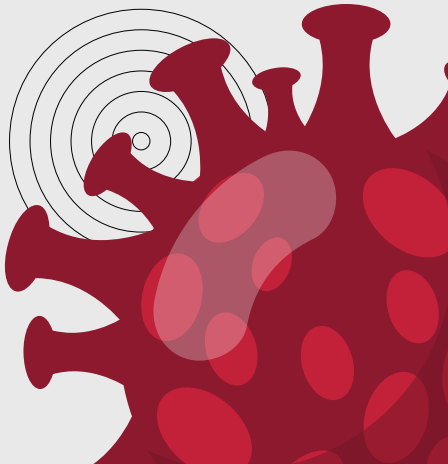
characteristics_ch1.0.disease state	COVID-19	non-COVID-19
clusters		
0	61	22
1	39	4

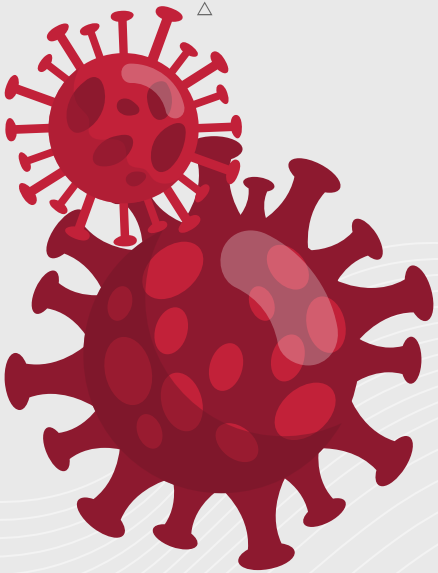
characteristics_ch1.3.icu	no	yes
clusters		
0	25	54
1	35	12

characteristics_ch1.6.mechanical ventilation	no	yes
clusters		
0	39	44
1	36	7

FUTURE WORK

PREDICT SEVERITY AND OUTCOME BASED ON OMICS, CLINICAL AND LABORATORY DATA IN ORDER TO PRIORITISE PATIENTS AND TREATMENTS





THANKS!

MSc in Bioinformatics
Sistemas Inteligentes em Bioinformática
Professor Miguel Rocha

Cátia Mendes, pg30435
Joana Ribeiro, pg38272
Maria João Lopes, pg40965
Miguel de Sousa Castro, pg40970

With the collaboration of Marta Sequeira