1 Datasets

Here we describe the datasets employed in our experiments.

1.1 *Hospitalization* and *Severity*

Both datasets were extracted from the same data source containing the symptoms and comorbities declared when a citizen took a COVID test in the city of São José de Campos located in São Paulo state in Brazil. This data source was obtained through a partnership with the health Secretariat of São José dos Campos. The dataset encompasses the tests between March of 2020 and May of 2021. This database contains sensitive information and for ethical reasons can not be shared.

The *Hospitalization* dataset has three balanced classes: short (1:5 days), medium (6:10 days) and long (greater than 10 days) hospitalization. Table 6 summarize dataset information according to each class.

Table 1. Distribution of attributes per class in *Hospitalization* data set.

	Short	Medium	Long
Number of instances	1062	812	954
Fever	0.573	0.596	0.591
Cough	0.748	0.736	0.768
Sore throat	0.247	0.225	0.205
Dyspnea	0.707	0.778	0.752
Respiratory distress	0.529	0.499	0.534
Low oxygen saturation	0.569	0.638	0.666
Diarrhea	0.134	0.109	0.113
Vomit	0.065	0.578	0.055
Other symptoms	0.652	0.639	0.600
Chronic cardiovascular disease	0.034	0.298	0.0437
Immunodeficiency immunodepression	0.028	0.030	0.021
Chronic kidney disease	0.030	0.020	0.019
Diabetes mellitus	0.266	0.284	0.334
Asthma	0.034	0.026	0.035
Chronic neurological disease	0.024	0.039	0.026
Other chronic pneumopathy	0.034	0.030	0.034
Obesity	0.059	0.070	0.073
Other risks	0.328	0.372	0.410
Chronic respiratory disease	0.019	0.021	0.016
Sex (female)	0.455	0.432	0.413
Age (mean - sd)	53.3 - 19.7	58.5 - 16.5	61.3 - 14.7

Dataset *Severity* has two classes: severe and non-severe patients. Severe patients are those who progressed to death within days 30 of hospitalization or

were hospitalized for ten or more days. Non severe patients were not hospitalized or were hospitalized for less than 10 days. This dataset was very imbalanced, with only 0.9% of instances in the severe class. Since the class imbalance affects the instance hardness analysis, in this work we randomly removed part of the majority class before the experiments. Table 2 summarizes dataset information according to each class.

1.2 *Hospital 1* and *Hospital 2*

Both datasets were extracted from the same data source containing the laboratorial test results during the COVID pandemic for patients admitted at two private hospitals in the city of São Paulo. This data source is open and publicly available at [1]. The preprocessed datasets are available in the Supplementary material. In these datasets all patients are hospitalized and severity is assigned to patients who progress to death (within a period of 30 days from hospitalization) or for which the hospitalization period was larger than 14 days. Both datasets were imbalanced, and passed had instances randomly removed before the experiments. These datasets also contain missing values, which are input by the mean of the three nearest neighbors values. Table 4 summarizes some datasets information according to each class, after the imputation step.

2 Additional Results

Next we present additional results, divided by experiments.

2.1 Footprint Analysis

Figure 1 shows the how the hard footprints change according to some of the T values explored. As the T increases, we expect that more instances are considered hard (and easy too). This is true for the dataset *Hospital 1* that shows an intermediate hardness profile. It is possible to see that for each T value the footprints have their size increased to encompass new instances. In the other profiles only the hard footprints has a significant increase, although this happens for different reasons. In *Severity* all of the easy instances are already present from the first footprints generated. While for the *Hospitalization* dataset, even for low values of T the easy footprint is not formed.

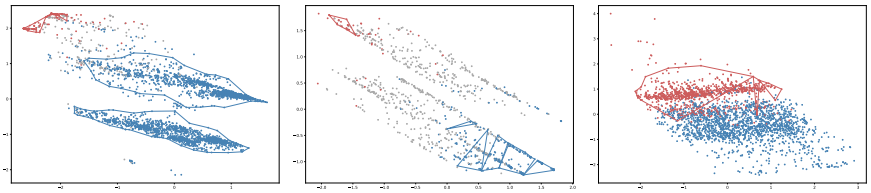
[1] https://repositoriodatasharingfapesp.uspdigital.usp.br/

Table 2. Distribution of attributes per class in *Severity* dataset.

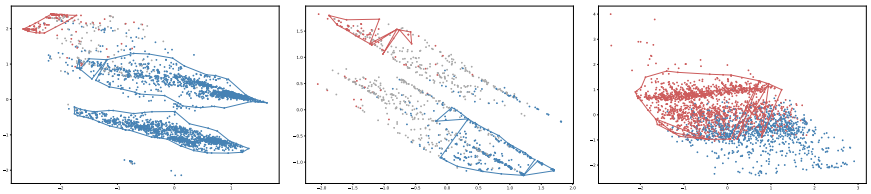
	Severe	Non-severe
Number of instances	1710	1710
Fever	0.551	0.680
Cough	0.735	0.791
Sore throat	0.184	0.546
Dyspnea	0.784	0.308
Respiratory distress	0.568	0.041
Low oxygen saturation	0.692	0.046
Diarrhea	0.091	0.012
Vomit	0.056	0.007
Other symptoms	0.556	0.750
Chronic cardiovascular disease	0.470	0.150
Immunodeficiency immunodepression	0.030	0.008
Diabetes mellitus	0.367	0.115
Obesity	0.074	0.009
Other risks	0.438	0.029
Chronic respiratory disease	0.017	0.036
Sex (female)	0.414	0.536
Age (mean - sd)	64.7 - 15.7	42.0 - 16.0

Table 3. Summary of the datasets *Hospital 1* including number o patients, period of data collection, percentage of male patients, age statistics and blood tests.

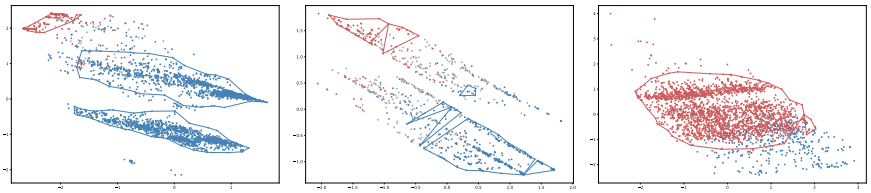
	Severe	Non-severe
Period	03/2020 - 05/2021	
Number of instances	529	529
Sex (female)	0.357	0.340
Age (mean - sd)	68.9 - 12.9	59.4 - 15.4
Blood tests	sodium, potassium, C-reactive protein, GOT, GPT, urea, D-dimer, troponin, creatine kinase, blood count, creatinine	



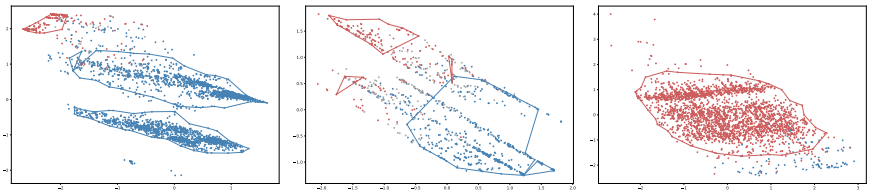
(a) *Severity* - $T = 0.032$ (b) *Hospital 1* - $T = 0.032$ (c) *Hospitalization* - $T = 0.032$



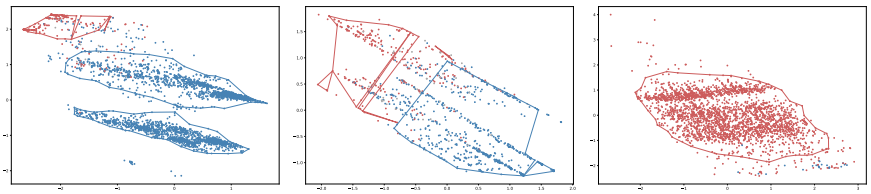
(d) *Severity* - $T = 0.036$ (e) *Hospital 1* - $T = 0.036$ (f) *Hospitalization* - $T = 0.036$



(g) *Severity* - $T = 0.040$ (h) *Hospital 1* - $T = 0.040$ (i) *Hospitalization* - $T = 0.040$



(j) *Severity* - $T = 0.044$ (k) *Hospital 1* - $T = 0.044$ (l) *Hospitalization* - $T = 0.044$



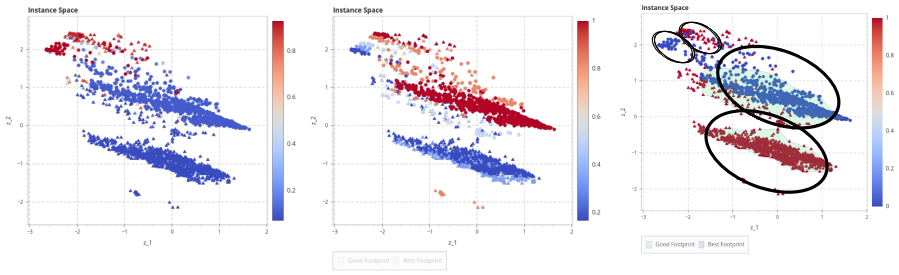
(m) *Severity* - $T = 0.048$ (n) *Hospital 1* - $T = 0.048$ (o) *Hospitalization* - $T = 0.048$

Fig. 1. Hard and easy footprints in datasets of distinct hardness profile. Footprints consider different values of T, increasing with a step = 0.04.

2.2 Visually inspecting the hardness embedding

In this section we extend the analyse inside subgroups of the instance space for *Severity* and *Hospital 1* datasets.

***Severity* dataset** Figure 2a and 2b presents the IS colored by some of the meta-features employed in the framework and Figure 2c shows the IS colored by the original class as well as the four groups delineated.



(a) *Severity* IS colored by feature CL

(b) *Severity* IS colored by feature TD_P

(c) *Severity* IS colored by original class.

Fig. 2. IS of *Severity* dataset colored according to meta-feature values and original class values. In (c) we highlight four interest groups formed in the IS.

Next we filtered instances inside each group using their meta-feature values. Table 5 presents meta-features used in this step, as well as a summary of feature values in each group.

Group 1 contains severe patients that are easy do classify. Analysing feature values we can see that three attributes are strongly correlated with the severe class: low oxygen saturation, respiratory distress and other risks. Indeed, all instances inside group 1 contain at least one of these symptoms or comorbidities. This is also reveled by meta-feature values since this group presents low TD_P values, meaning that when one of these attributes are present the models can easily classify the instance as severe, on the other hand if they are not present another attributes should be take on account. This is why instances in group 4 show a higher TD_P.

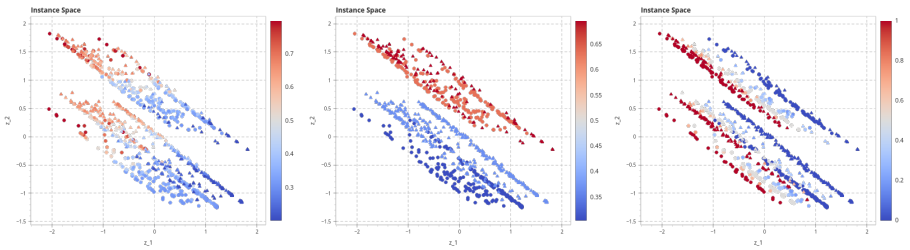
For group 2 we add to the previous analysis an interpretation of the meta-features. This group is caracterized by high values of CL meaning they have a low probablity of belonging tho their own classes, which in some cases proved to be true after further investigations. The TD_P are low for these instances, since they have attributes strongly correlated with the opposite class, in this way they are classified in the opposite class, showing a high hardness level.

Group 3 individuals developed a severe condition although they did not present the most correlate symptoms to severity. This is why this group presents

high values of TD_P, when these attributes are not present, the decision tree includes other ramifications to classify such instances. Analysing the original data base we looked for possible reasons why this patients developed a severe condition without presenting some important severe symptoms. Three observations are mislabeled again, while 27 were hospitalizad between 10 -14 days. This last group can be considered instances in the overlap regions since they are between the two classes, and would be considered non-severe if we had adopted an slightly different proxy to define severity. A group of 78 patients progressed to death after hospitalization, for them the number of days between test and hospitalization is in average 5, so they could have developed the symptoms after testing. We do not discharge the possibility of mistakes in data collection since the forms were first filled by hand when the patient went to the pharmacy or hospital for testing.

Inspecting Group 4, we see that these patients do not present any symptom correlated with severity and consequently they are correctly classified.

***Hospital 1* dataset** Figure 3 presents the IS colored by some of the meta-features employed in the framework.



(a) *Hospital 1* IS colored by feature CL

(b) *Hospital 1* IS colored by feature DCP

(c) *Hospital 1* IS colored by feature $N1$.

Fig. 3. IS of *Hospital 1* IS colored according to meta-feature values.

Contrasting meta-features values we identified six groups inside the hardness profile and proceed a further inspection to obtain insights about how and why instances in each group are difficulty to classify. We filtered instances inside each group using their meta-feature values. Table 6 presents meta-features used in this step.

In this analysis all groups have hardness level >0.6, the same is true for feature CL values. Groups 1 and 3 have already been discussed in the main paper. For other groups the explanations are not so clear, although interesting questions came out through IS analysis and raw data base inspection. For group 2 patients are hospitalized without an aggravated clinical condition, and their condition worse on the following days. The question here is why patients are

hospitalized if they do not have an aggravated clinical condition? One possible reason may be the group age, since they are older the hospitalization decision may have been made since the doctor was already expecting a severe outcome. Here it is important to point out that this data base comes from a private hospital in the city of São Paulo where more resources were available during the pandemic.

Inside group 4 an interesting phenomenon is that some patients are hospitalized without their blood tests pointint out an aggravated condition, as in the group 2. In some days they have a significant worsening and are moved to the ICU. From the data we have access to it is impossible to guess the reason. Maybe they can receive a similar explanation of group 2 although they are younger in group 4. We also can prospect the possibility that they were hospitalized due to other health conditions, like a surgery for example, being the COVID a consequence of the hospitalization and not its cause.

For groups 5 and 6 we could not grasp a clear patter between instances. These instances are characterized by high values of feature $N1$, meaning that they are not completely surrounded by instances on the opposite class, but are probably located in the border of the classes.

Table 4. Summary of the datasets *Hospital 2* including number o patients, period of data collection, percentage of male patients, age statistics and blood tests.

	Severe	Non-severe
Period	03/2020 - 01/2021	
Number of Instances	67	67
Sex (female)	0.328	0.537
Age (mean - sd)	65.8 - 14.6	55.2 - 17.0
Blood tests	sodium, potassium, C-reactive protein, GOT, GPT, urea, D-dimer, troponin, creatine kinase, blood count, creatinine	

Table 5. Summary of subgroups from dataset *Severity* with meta-features threshold used to filter data and feature values inside each group.

	Group 1	Group 2	Group 3	Group 4
Number of instances	1380	90	124	1529
Original class	Severe	Non-severe	Non-severe	Severe
Hardness level	<0.4	>0.6	>0.6	<0.4
Feature CL	<0.4	>0.6	>0.6	<0.4
Feature TD_P	<0.4	<0.4	>0.6	<0.6
Fever	0.53	0.53	0.69	0.69
Cough	0.72	0.78	0.85	0.81
Sore throat	0.17	0.18	0.35	0.59
Dyspnea	0.83	0.81	0.50	0.28
Respiratory distress	0.64	0.62	0	0
Low oxygen saturation	0.84	0.87	0	0
Diarrhea	0.09	0.11	0	0
Vomit	0.06	0.07	0	0
Other symptoms	0.53	0.63	0.80	0.76
Chronic cardiovascular disease	0.5	0.42	0.27	0.12
Immunodeficiency	0.03	0.02	0	0
Diabetes mellitus	0.40	0.37	0.19	0.09
Obesity	0.08	0.11	0	0
Other risks	0.54	0.42	0	0
Chronic respiratory disease	0.01	0.01	0.04	0.04
Sex (female)	0.41	0.46	0.38	0.54
Age (mean)	66.0	54.54	52.59	40.39
Age (min - max)	6 - 108	2 - 92	2 - 94	0 - 86

Table 6. Summary of subgroups from dataset *Severity* with meta-features threshold used to filter data and feature values inside each group. The value *0* for the original class corresponds to *non-severe* class and the value *1* to the *severe* class.

	Group 1	Group 2	Group 3	Group 4	Group 5	Group 6
Number of instances	8	12	30	30	12	7
Original class	0	1	0	1	0	1
Feature CL	>0.6	>0.6	>0.6	<0.6	>0.6	>0.6
Feature DCP	<0.4	<0.4	>0.6	<0.6	>0.6	<0.6
Feature $N1$	>0.6	<0.6	>0.6	<0.6	<0.4	<0.4
Age (min - max)	52 - 58	62 - 83	62 - 87	27 - 61	62 - 89	42 - 61