An interative dashboard for statistical analysis of COVID-19 data for PREMO project

Ruben Dias[(1)]
A46037@alunos.isel.pt

Artur Ferreira[(1),(2)]
artur.ferreira@isel.pt

Iola Pinto[(1),(3)]
iola.pinto@isel.pt

Carlos Geraldes[(1),(4)]
carlos.geraldes@isel.pt

Cristiana Von Rekowski[(1)]
crisvr66@hotmail.com

Luís Bento[(5),(6)]
luis.bento@chlc.min-saude.pt

[(1)] ISEL, Instituto Superior de Engenharia de Lisboa Instituto Politécnico de Lisboa, Lisboa, PORTUGAL

[(2)] IT, Instituto de Telecomunicações, Lisboa, PORTUGAL

[(3)] Center for Mathematics and Applications (NOVA Math), NOVA SST

[(4)] Centro de Estatística e Aplicações, CEAUL, Universidade de Lisboa, Lisboa, PORTUGAL

[(5)] Department of Intensive Care Medicine (Unidade de Urgência Médica), São José Hospital, Central Lisbon University Hospital, Lisboa, PORTUGAL

[(6)] NOVA Medical School, Universidade Nova de Lisboa, Lisboa, PORTUGAL

Abstract

COVID-19 caused a recent pandemic period, due to its ease of transmission and large number of infections. The pandemic had severe consequences for the mortality and morbidity of populations, especially the elderly. Several research projects have raised from the pandemic, such as the *Predictive Models of COVID-19 Outcomes for Higher Risk Patients Towards a Precision Medicine* (PREMO) project. In this paper, we report the development of a Web application providing a dashboard with COVID-19 statistical data analysis from patients of the six waves of COVID-19, in Portugal. The visualizations provided by the platform allow specific analyzes of the clinical profile for patients hospitalized in Intensive Care Unit (ICU), per wave. The dashboard is an useful tool to extract information and scientific knowledge about the disease evolution, providing also beneficial insights for future pandemics.

1 Introduction

COVID-19 is caused by a coronavirus identified for the first time in humans, in Wuhan, China, in November 2019. It led to a pandemic period due to its ease and speed of transmission, causing respiratory infections, flu-like symptoms, and pneumonia [6]. This disease affects populations in waves, with different levels of mortality and morbidity per wave. The knowledge about the disease and the introduction of vaccination at some point, led to changes in these indicators.

Hospitals and healthcare units collected many different types of data related to COVID-19 patients in ICU. To effectively study the behavior of the disease on its different waves, it is necessary to aggregate these dispersed data records on a single database and to provide a data visualization application running on a browser. The use of this application allows for better understanding of the disease, which helps to prepare the society for future pandemics. The *Predictive Models of COVID-19 Outcomes for Higher Risk Patients Towards a Precision Medicine* (PREMO) project, analyzes the temporal evolution of several clinical parameters of COVID-19 patients admitted to the ICU. It uses data gathered from different hospitals in Lisbon, Portugal, on the different COVID-19 waves.

In this paper, we propose a solution for the PREMO project allowing to aggregate data related to COVID-19 and to display different data views on a dashboard. It allows to analyze the evolution of biomarkers associated to different organs, related with the disease, performing statistical analysis and comparisons with different graphs and diagrams. One of the key aspects of the proposed solution is that it provides Kaplan-Meier survival curves, allowing to compare mortality rates according to cut-off points that separated the normal results from each of the biomarkers from the increased/decreased ones.

The remainder of this paper is organized as follows. Section 2 refers to related work on the development of similar dashboards. Section 3 describes the approach followed and the main aspects of the proposed solution. Section 4 provides examples of graphs generated by the system. Finally, Section 5 presents some general remarks about the solution.

2 Related Work

The development of vision boards, typically named as dashboards, is suitable in many applications. The proper data visualizing enables correct analysis and subsequent extraction of information and knowledge, allowing informed decisions.

Some dashboards show the number of infections, number of recovered patients, number of vaccinations, number of deaths, among other indicators, monitoring the spread of the virus. One such example is the dashboard developed by the Johns Hopkins University Center for Systems Science and Engineering [2], depicted in Figure 1. Another example of an interactive dashboard to monitor COVID-19 in the United States is the Wissel dashboard [7], shown in Figure 2.

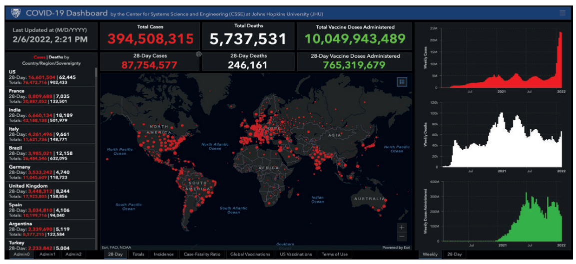

Figure 1: Dashboard to monitorize the spread of the virus [2].

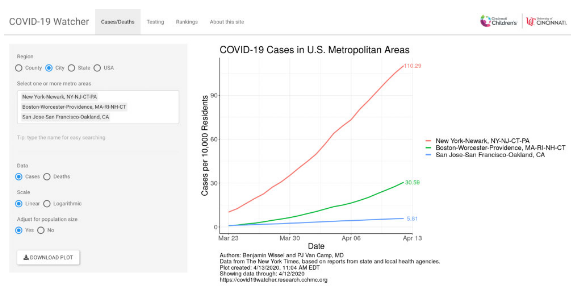

Figure 2: The Wissel interactive dashboard [7].

A dashboard to monitor ventilation equipment (types of machines, operating modes, and ventilated patients) and the status of patients (laboratory results) was also proposed [3]. A software package for analysis and visualization of COVID-19 data on a dashboard is also available [5].

3 Proposed Solution

On the first step of this work, we found that the collected data from two Lisbon hospitals were scattered across different databases and files. Thus, to gather all the data, we devised an Extract, Transform, and Load (ETL) process composed by three phases as described in Figure 3. The first

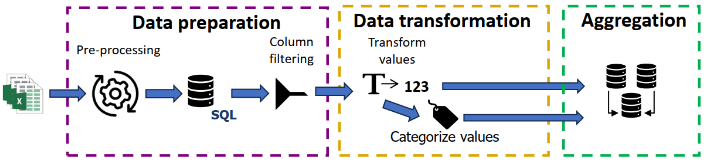

Figure 3: Extract, Transform, and Load (ETL) process with three phases.

phase imports the data from several sources into a single database, removes duplicates and out of range values. The second phase is composed by data transformation and labelling. Finally, on the third phase we perform patient data aggregation and description.

We devise a 3-layer architecture [1][4], depicted in Figure 4. These layers are Presentation, Logic, and Database. The Presentation layer is

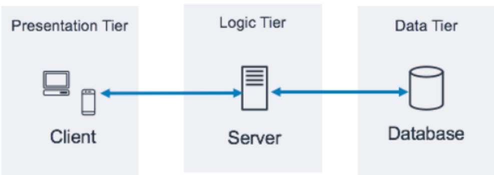

Figure 4: The 3-layer architecture [1] of our Web application.

implemented in Javascript using React, making the application (dashboard) available through a browser. To input the generated data to the dashboard, we resort to the Nginx tool. The solution is deployed with Docker containers.

4 Application Interface and Graphs

The construction of graphs is done at the Presentation layer level, with the Plotly library. Figure 5 displays the dashboard interface as seen from the browser. On the left-hand side we have graphic controls for choosing patient data. In the central part and on the right-hand side, the graphic views are displayed according to the parameters indicated by the user. Our solution displays different graphs required by the clinical staff: *circular*, to analyze nominal variables; *dispersion*, to check correlations among quantitative variables; *box-plot*, since parallel boxplots compare the distribution of quantitative indicators performing statistical comparisons between groups defined by the categories of nominal variables; *survival curves* to analyze the survival probabilities between groups of patients, via Kaplan-Meier curves; *line-plot* to assess the evolution of the different parameters of ICU patients. Figure 6 provides an example of a dispersion graph while Figure 7 contains a parallel box-plot to compare pH levels. Finally, Figure 8 shows an example of survival curves.

5 Conclusions

The monitoring of COVID-19 variables allows to study the pandemic. The developed solution allows to analyze these variables, since the user can select different combinations of parameters, isolate the data by wave or to aggregate waves, using the browser with a simple-to-use interface. This visualization yields knowledge about the behavior of different waves of COVID-19. Using statistic analysis, we differentiate the behavior of the variables in each wave of COVID-19. The platform contributes to the understanding of this disease and to prepare us for future pandemics.

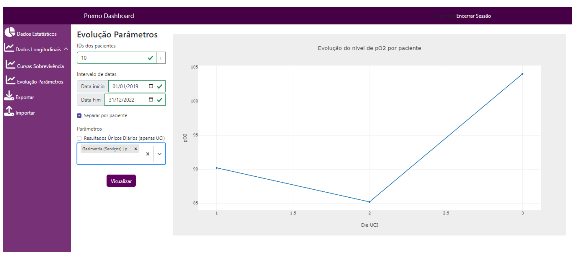

Figure 5: Dashboard interface of the proposed solution.

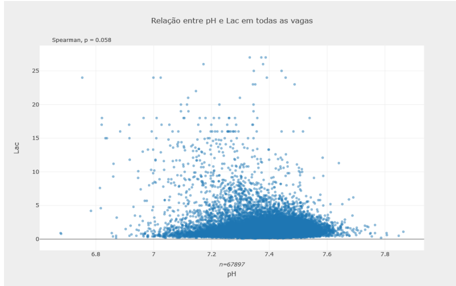

Figure 6: Scatter plot to check the correlation between the quantitative variables pH and Lac, with Spearman correlation coefficient.

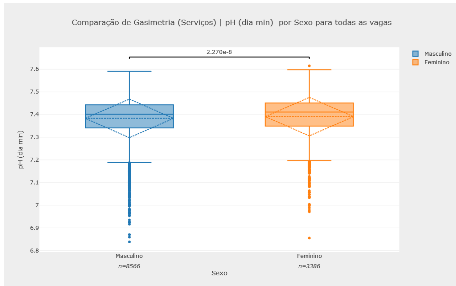

Figure 7: Box-and-whisker plots to compare the pH level by gender.

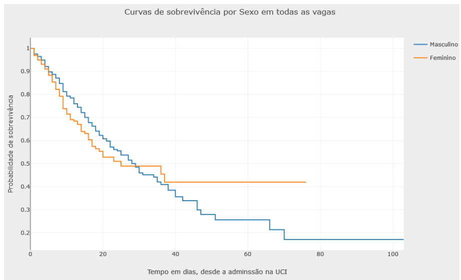

Figure 8: Survival curves for all COVID-19 waves, with the probability of survival, as a function of the number of days in the ICU.

Acknowledgements

The authors thank the Predictive Models of COVID-19 Outcomes for Higher Risk Patients Towards a Precision Medicine (PREMO) project, by Fundação para a Ciência e Tecnologia (FCT), grant DSAIPA/DS/0117/2020.

References

[1] AWS. Three-tier architecture overview, February 2023. URL `https://docs.aws.amazon.com/whitepapers`.

[2] Ensheng Dong et al. The Johns Hopkins University Center for Systems Science and Engineering COVID-19 Dashboard: data collection process, challenges faced, and lessons learned. *The lancet infectious diseases*, 22(12):e370–e376, 2022.

[3] Randeep Jawa et al. A reference guide to rapidly implementing an institutional dashboard for resource allocation and oversight during covid-19 pandemic surge. *JAMIA open*, 3(4):518–522, 2020.

[4] IBM. What is three-tier architecture?, February 2023. URL `https://www.ibm.com/topics/three-tier-architecture`.

[5] Marcelo Ponce and Amit Sandhel. covid19. analytics: An r package to obtain, analyze and visualize data from the coronavirus disease pandemic. *arXiv preprint arXiv:2009.01091*, 2020.

[6] SNS. Sns24 - covid-19, February 2023. URL `https://www.sns24.gov.pt/tema/doencas-infecciosas/covid-19/#o-que-e-a-covid-19`.

[7] Benjamin D Wissel, PJ Van Camp, Michal Kouril, Chad Weis, Tracy A Glauser, Peter S White, Isaac S Kohane, and Judith W Dexheimer. An interactive online dashboard for tracking covid-19 in us counties, cities, and states in real time. *Journal of the American Medical Informatics Association*, 27(7):1121–1125, 2020.