

# Fast prototyping of a local fuzzy search system for decision support and retraining of hospital staff during pandemic

Evgeny Bakin (✉ [eugene.bakin@gmail.com](mailto:eugene.bakin@gmail.com))

Pavlov University, Saint-Petersburg <https://orcid.org/0000-0002-5694-4348>

Oksana Stanevich

Pavlov University, Saint-Petersburg

Dmitry Lioznov

Pavlov University, Saint-Petersburg

Alexander Kulikov

Pavlov University, Saint-Petersburg

---

## Method Article

**Keywords:** Fuzzy search, COVID-19, Decision support algorithm, Patient-like-mine, Prototyping

**Posted Date:** October 19th, 2020

**DOI:** <https://doi.org/10.21203/rs.3.rs-93413/v1>

**License:**  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

**Version of Record:** A version of this preprint was published at Health Information Science and Systems on May 11th, 2021. See the published version at <https://doi.org/10.1007/s13755-021-00150-y>.

# Fast prototyping of a local fuzzy search system for decision support and retraining of hospital staff during pandemic

Evgeny A. Bakin · Oksana V. Stanevich · Dmitry A. Lioznov · Alexander N. Kulikov

Received: date / Accepted: date

**Abstract Purpose.** COVID-19 pandemic has shown an urgent need for decision support systems to help doctors in a period of stress and uncertainty. However, significant differences in hospital conditions, as well as skepticism of doctors about machine learning algorithms, limit their introduction into clinical practice. Our goal was to test and apply a principle of "patient-like-mine" decision support in rapidly changing conditions of a pandemic. **Methods.** In the developed system we implemented a fuzzy search that allows a doctor to compare their medical case with similar cases recorded since the beginning of the pandemic in their medical center. Various distance metrics were tried for obtaining clinically relevant search results. With the use of R programming language, we designed the first version of the system in approximately a week. A set of features for the comparison of the cases was selected with the use of random forest algorithm implemented in Caret. Shiny package was chosen for the design of GUI. **Results.** The deployed tool allowed doctors to quickly estimate the current conditions of their patients by means of studying the most similar previous cases stored in the local health information system. Extensive testing of the system during the first wave of COVID-19 has shown that this approach helps not only to draw a conclusion about the optimal treatment tactics and to train medical staff in real-time but also to optimize patients' individual testing plans. **Conclusions.** This project points to the possibility of rapid prototyping and effective usage of "patient-like-mine" search systems at the time of a pandemic caused by a poorly known pathogen.

**Keywords** Fuzzy search · COVID-19 · Decision support algorithm · Patient-like-mine · Prototyping

Pavlov First Saint Petersburg State Medical University  
L'va Tolstogo str. 6-8, Saint Petersburg, Russia, 197022  
+7 (812) 338-66-49  
E-mail: eugene.bakin@gmail.com

## 1 Introduction

Recent events revealed that despite considerable progress in health information systems (HIS), biostatistics, clinical pharmacology and evidence-based medicine in general, humanity still does not have sufficient tools to address sudden pandemic outbreaks. The spread of the disease caused by a new coronavirus infection (COVID-19) was accompanied by frequent changes in clinical guidelines, absence of unified standards for anti-epidemic measures, and a huge number of publications, sometimes conflicting with one another, appeared every day. Besides, a few cases of medical data falsification were detected. All this increased uncertainty, as well as a high workload of medical personnel in COVID-19 centers [1].

To assist physicians, research groups in IT started developing decision support systems (DSS) to mitigate different aspects of the pandemic. Thus, paper [2] investigates the issues of constructing a simple patient condition severity classifier based on remote survey data to adjust medical logistics. System [3] developed by Sapio Analytics is planned as a tool to optimize quarantine regime. Project Vida, proposed by a team of scientists from the MIT, aims to conduct a comprehensive analysis of the situation in a region (involving geographic information systems data) for the subsequent construction of SARS-CoV2 propagation models and assessment of its impacts on socio-economic indicators [4,5]. Study [6] investigates the dependence of pandemic spread velocity on the measures of epidemic control. Similarly, in [7], mortality predictions are made in real time basing on current epidemic data. Various platforms for the remote provision of medical services via the Internet are also offered [8].

At the same time, the accumulation of inpatients data while in hospital allowed for a variety of forecasting models for predicting the likelihood of infection and the severity

of disease basing on case history, demographic and clinical parameters as well as laboratory tests [9, 10, 11, 12, 13, 14].

These models were conceived as a way to help physicians (including non-infectious specializations), operating in centers for the treatment of COVID-19. However, in spite of all the above mentioned factors, the effectiveness of these tools in clinical practice remains very limited [15]. We consider the following main reasons for delaying the implementation of the decision-making systems that are so necessary during a pandemic.

#### 1. *Differences in hospital structure and equipment*

Patient management practices, the severity of cases admitted to a hospital, the possibility for conducting expensive laboratory tests on a regular basis, congestion in intensive care units (ICUs) have a strong impact on the main endpoints of the above studies. Thus, forecasting models built in the context of large, well-equipped hospitals may not be applicable to typical conditions in regional clinics or clinics in developing countries, and vice versa.

#### 2. *Significant percentage of non-infectionists working in COVID-19 treatment centers*

Refurbishment of non-infection hospitals was often carried out to cope with the outbreak of cases. Specialists in other medical fields who had no experience in the treatment of infectious diseases were rapidly involved in the work. In addition to the direct supervision of patients, they simultaneously faced the need to bridge a knowledge gap. The simple use of computer predictors that form a binary response (e.g. "death"/"recovery") doesn't contribute to the development of clinical skills, but, on the contrary, could introduce an element of formality in the doctor's work.

#### 3. *Lack of medications with a proven effectiveness*

History has shown that during a rapidly spreading previously unknown infection, the standard practice of conducting clinical trials of medications has significant lags. The off-label use of medications basing on results obtained in vitro, or in the absence of randomization and blinding often turned out to be ineffective, and, in some cases, even dangerous. An example is the widespread use of combinations of drugs based on hydroxychloroquine and ART at the initial stages of the pandemic, the expediency of which was then refuted [16,17]. Therefore, informing the physician about a patient's high risk of ICU admission (or death) in the absence of effective prevention therapy may cause unnecessary stress or, conversely, personnel passivity during treatment.

#### 4. *Skeptical attitude to the published information*

Unfortunately, the COVID-19 pandemic is also accompanied by a series of scandals related to unfair testing and falsification of results. The loudest of

these cases concerns the activities of the Surgisphere Company, which was found to have provided unreliable medical data for several studies [18]. As a result, a few articles published in well-established journals The Lancet and The New England Journal of Medicine, were subsequently withdrawn [19,20]. All this reduces the level of trust of medical professionals in the published data and makes them unwilling to rely on the experience of other organizations.

Thus, the attitude of clinicians to the published Decision Support Systems can be reduced to the formula "Abstract models based on irrelevant or falsified data urging to unknown actions". This inspired us to implement a different approach to medical informatics methods application for helping physicians.

## 2 Objectives

When developing a decision support system at Saint-Petersburg Pavlov University, we set the goal of creating an application with a graphical interface with the help of which a doctor, having entered a relatively small number of indicators of their patient, could find in the database (DB) the most similar cases recorded in the hospital earlier. Having studied the therapies applied earlier and what outcomes they have led to, the doctor can adjust the course of treatment and, if necessary, consult their colleagues who supervised previous patients. It is obvious that due to the uniqueness of each individual patient, the search based on the exact match of all clinically significant symptoms is hardly possible.

Hence, we decided to implement a fuzzy search system that allows incomplete data matching. Fuzzy search systems for medical data have been actively discussed recently in relation to computerization of medical institutions and the growth of stored information volume. Some of them are mainly focused on searching for patients based on incomplete or mistakenly entered identification data (full name, date of birth, sex, policy number, etc.) [21, 22, 23, 24]. The others utilize the more sophisticated approach named "patient-like-mine", which implies aggregating previous electronic health records for searching clinically-similar patients in a hospital history [25, 26, 27]. These systems may have quite a complex structure and require a deep integration with a local health information system (HIS). All this made it difficult to develop and introduce them into clinical practice immediately after the pandemic outbreak. To address the challenge, we set the following tasks:

- Select software tools which allow fast prototyping and deployment in the local network by work of a biostatistician unit without the involvement of third parties and without a modification in the local HIS.

- Create and pre-process a dataset containing data for all the patients in the COVID-19 treatment center, including case data, demography, anthropometry and daily results of laboratory tests.
- Select a relatively brief list of patient features that contain a sufficient amount of information about a patient’s current condition.
- Develop a fuzzy search system that takes in the features of the current patient entered by the doctor, and outputs a set of the most similar cases, indicating the ID numbers of past patients and the days of hospitalization when this similarity was noticed.
- Implement a graphical user interface (GUI) that represents the information found in a user-friendly form and allows the doctor to quickly navigate the cases under consideration.

### 3 Methods

#### 3.1 Software tools

R version 3.6.3 was chosen as the main programming language [28]. The list of the main R packages used in the project is given in Table 1.

Table 1: R packages used in the project

Purpose	Packages	Reference
Data pre-processing	<i>tidyr</i>	[29]
Application of machine learning methods for main features selection	<i>caret</i>	[30]
Vizualization of health information data	<i>ggplot2</i> <i>ggpubr</i>	[31, 32]
Building of GUI	<i>shiny</i>	[33]

#### 3.2 Generation of initial dataset

From the local health information system (HIS), we downloaded 6 separate datasets which stored data about the patients’ states:

1. Registration data: the date of intake, the number of bed-days in the hospital, the treatment outcome.
2. Demographic and anthropometric data: age, sex, height, weight (combined later into BMI).
3. Anamnestic data: the number of days of illness, the presence of chronic diseases, the smoking status.
4. Results of biochemical tests: ALT, AST, amylase, bilirubin, bound bilirubin, creatinine, SRP, D-dimer, fibrinogen, ferritin, glucose, glucose ABB, lactate ABB, potassium, potassium ABB, sodium, sodium ABB, total protein, troponin I, urea.

5. Results of a clinical blood test: total leukocytes, neutrophils, monocytes, platelets, lymphocytes, hemoglobin.
6. NEWS scale components: heart rate (HR), respiratory rate (RR), systolic blood pressure, temperature, saturation, need for additional oxygenation, AVPU score.

All the datasets contained the key field Patient ID. Datasets 4-6 were downloaded in a long format with reference to the date of the tests. Further, all the datasets were combined into a single array, one line of which corresponded to one day of hospitalization of one patient. The values in the columns corresponded to the results of tests taken on a particular day. If on one of the days a certain test was taken several times, the median of the measured values was entered in the table. The illustration of the process of forming the source dataset is shown in Figure 1. Next, missing values were filled in using last observation carried forward procedure (LOCF) [34].

#### 3.3 Features selection

For the selection of the main features that reflect the condition of the patient, we conducted training of a set of standard machine learning algorithms: logistic regression, random forest, k-nearest neighbors, support vector machine and gradient boosting. The results of the algorithms were checked using k-fold cross-validation. For the best classifier (in terms of area under the curve, AUC), the features were ordered according to their importance. Top 12 clinical and laboratory features were selected from this list, where sex and age were also added. Thus, the selection of similar cases was carried out basing on 14 features.

#### 3.4 Fuzzy search

To carry out the search, a shortened version of the dataset was prepared to contain only the patients’ registration numbers, the number of a day of hospital stay, and the 14 previously selected features. The sex was coded with numbers 0 (female) and 1 (male). All the feature columns were scaled by calculating the z-score statistics. As a measure of differences, the following distances were considered: Minkowski (including Manhattan and Euclidian distances as special cases), Machalanobis and Spearman [35]. As a result of the search, a list of the most similar patients was formed with indicating the day for which the degree of similarity was maximal.

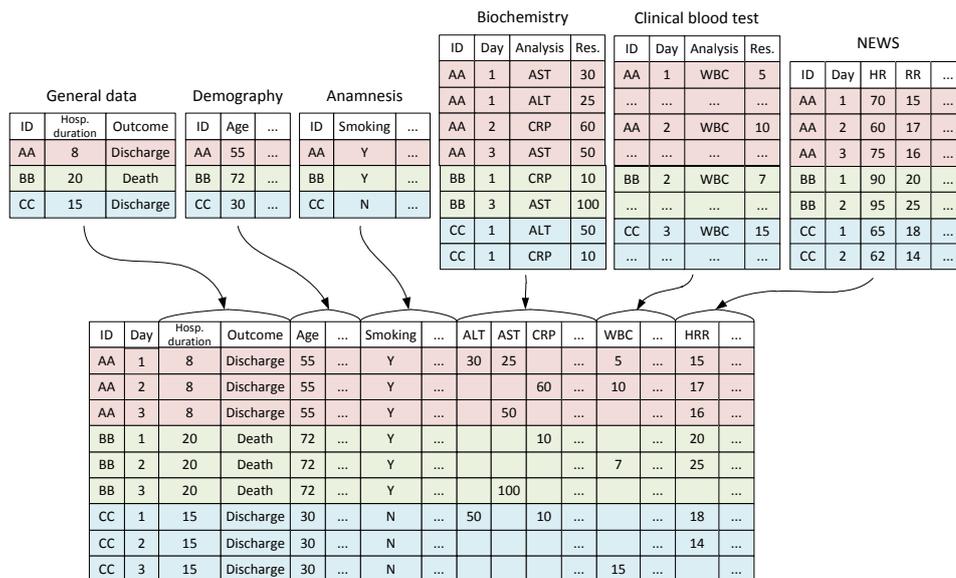


Fig. 1: The principle of formation of the initial dataset

## 4 Results

### 4.1 General description of DSS

The general structural diagram of the developed DSS is shown in Figure 2. An employee of the biostatistics department of the clinic downloads 6 raw datasets from the local HIS, which are further to be combined into a single dataset according to the principle described in Subsection 3.2. Next, the outliers are analyzed manually by finding abnormally large or small values in the data. The detected outliers are discussed with the clinicians for their correctness and the necessity to be excluded from the dataset. On the cleaned dataset, a list of the most important features is selected, so that doctors are able to search for analogies. Then the shortened version of the dataset is uploaded to the application server. On running the application, it loads the dataset and, based on the available columns, synthesizes the input fields for a graphical user interface (GUI). After entering the features of a patient whose condition the doctor doubts, the GUI displays brief information on similar cases recorded earlier. Having selected the analogies that deserve the most attention (for example, cases in which a complication has developed), the doctor can obtain a detailed diary of the cases from the HIS using the registration numbers of the patients.

The first test version of the software was based on data from the patients who were hospitalized at the Center for the treatment of the new coronavirus infection at hospital throughout May 2020. Subsequently, the system was updated bi-weekly as the patient data in the local HIS

were growing. The current version of the DSS database (as of September 1, 2020) is based on the information on 1572 patients with a median follow-up time of 8 days.

### 4.2 Selection of main patient features

For the selection of the most important attributes that describe the condition of a patient with COVID-19, we have trained a few machine learning models. The independent variables were the values of the features recorded on a particular day (see Subsection 3.3).

The dependent variable was the fact that the patient entered the intensive care unit within a week from the time of taking the tests. The resulting ROC - curves plotted on the basis of all analyzed attributes are shown in Figure 3a. As you can see, the random forest algorithm turned out to be the best in terms of AUC (AUC = 0.9). For this algorithm, the features were ranked according to their importance using the *VarImp* function. The results are shown in Figure 3b (the first 20 attributes are given).

Further on, the minimum number of attributes were selected for which the AUC remains at least 0.9 (their number was 12). The ROC-curves, built using these 12 features, are shown in Figure 3c. As one can see, in the random forest method, the AUC value remained at the level of 0.9, which indicates that the selected features contain a significant amount of information about the patient's condition. Later, age and sex were added to them as well. Thus, at the moment, the determination of the patient's current state is made according to the following 14 features:

- demography: age, sex;

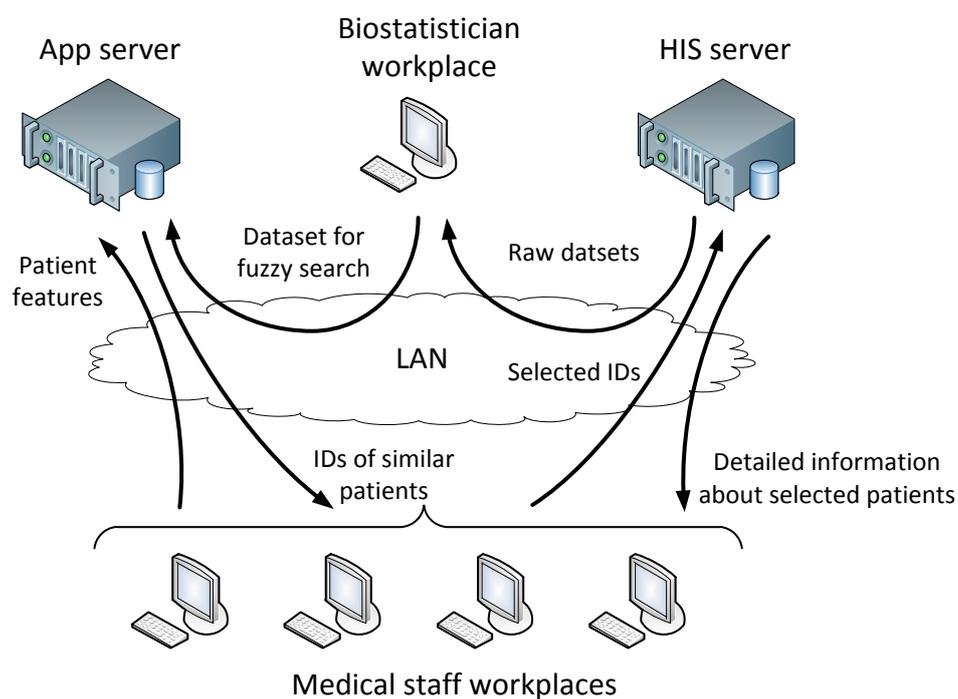


Fig. 2: The structure of the developed DSS

- clinical blood test: hemoglobin, as well as the absolute values of platelets, lymphocytes and neutrophils;
- biochemistry: SRP, procalcitonin, ferritin, creatinine, glucose, AST;
- vital indicators: SpO<sub>2</sub>, RR.

#### 4.3 Graphical user interface

Figure 4 shows an example of a graphical user interface. On the left panel there are elements for entering patient data, on the right - the search results in the database. For similar cases found, a brief summary is provided indicating the patient ID, full name, the date on which the entered data match as much as possible, some demographic and clinical laboratory parameters, as well as the outcome of hospital stay. There is also a histogram of outcomes for the detected similar cases, which allows us to quickly navigate for the severity of the patient's condition. Using the patient's ID obtained as a result of the search, we can then get a detailed diary of their stay in the hospital from the HIS.

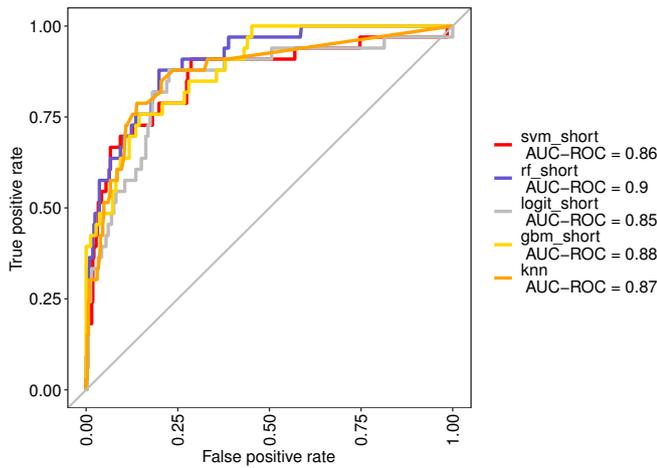
## 5 Discussion

Our experience of deploying the developed system in a hospital has shown great interest in such tools on the part

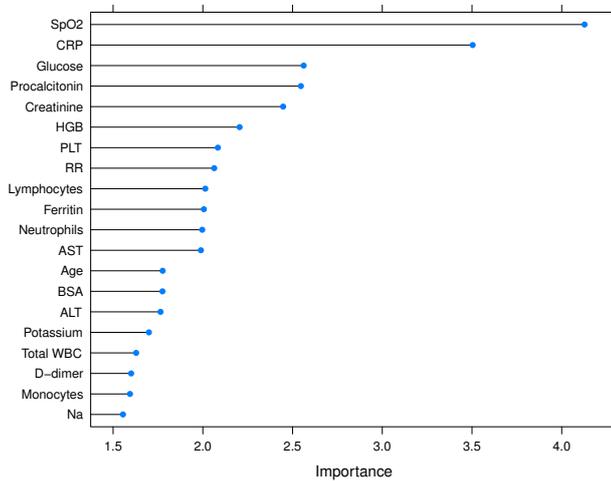
of medical workers. The lack of unambiguous instructions issued by the application helps reduce the level of skepticism and rejection among experienced doctors. At the same time, the need to independently draw conclusions on complex cases based on the study of several case histories improves qualifications and speed of immersion into a new subject area for doctors in related specializations.

Since the application deployment, more than 200 queries were addressed to the DSS. Their analysis has shown that our system turned out to be particularly useful in the following three use-cases as stated by physicians.

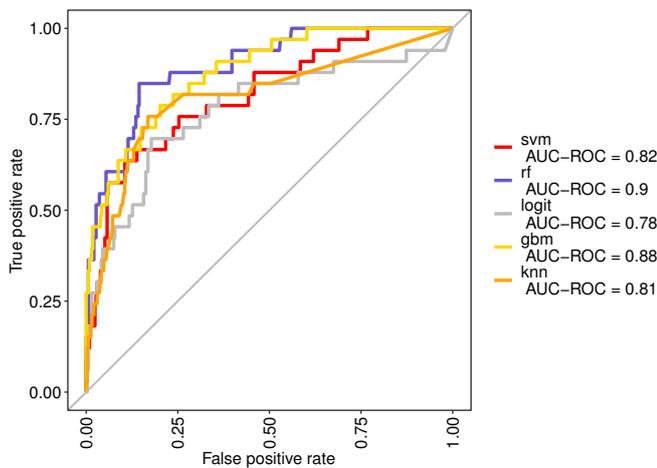
1. *Hesitation about the application of a medicine with controversial effects.* Particularly, we have noticed that early application of dexamethasone doesn't prevent patients with moderate conditions from developing severe complications. This finding was later rigorously proved in RECOVERY project [36].
2. *Optimization of a testing plan for a particular patient.* The study of similar cases may help foresee a rapid deterioration of a patient's condition, which provides a good reason for making particular medical tests more frequently.
3. *Forming a pool of patients for a further rehabilitation program.* Observing in HIS long-term consequences for similar patients who were discharged earlier, we may



(a) ROC curves for 5 classifiers built with the use of all predictors



(b) The features importance plot



(c) ROC curves for 5 classifiers built with the use of top-12 predictors only

Fig. 3: Results of a features selection procedure

add the analyzed patient to the invitation list for a consequent rehabilitation program.

In terms of choice of tools for DSS development in a pandemic, our conclusions are consistent with those of the authors of work [37]: in an unexpectedly changing environment, the choice of tools is determined by their potential for a rapid prototyping using ready-made modules. In this sense, *R* and *Shiny* are one of the best combinations of software. In our case, the development and deployment of the first version of the DSS took no more than a week, which made it possible to begin intensive testing of the proposed approach shortly after the opening of the COVID-19 Treatment Center.

## 6 Conclusion

The study showed a great potential for the application of DSS based on fuzzy search of information in conditions of uncertainty at the time of a pandemic caused by a poorly known pathogen. Our experience points to the possibility of the use of such systems both to support decisions taken by doctors, and to train medical staff. Besides, optimization of a testing plan for the patients may decrease a financial burden on the hospital during a high stress period.

**Acknowledgements** The authors would like to acknowledge medical staff of Pavlov University who participated in fruitful discussions and the testing of the developed system, in particular, Vassily Belash, Alexandra Novikova and Alexandra Lebedeva.

## Declarations

## Funding

None.

## Conflict of interest

The authors declare that they have no conflict of interest.

## Ethics approval

The study was performed in compliance with the World Medical Association Declaration of Helsinki Ethical Principles for Medical Research Involving Human Subjects, and was reviewed by Saint-Petersburg Pavlov University Institutional Review Board.

## Availability of data and material

Not applicable.

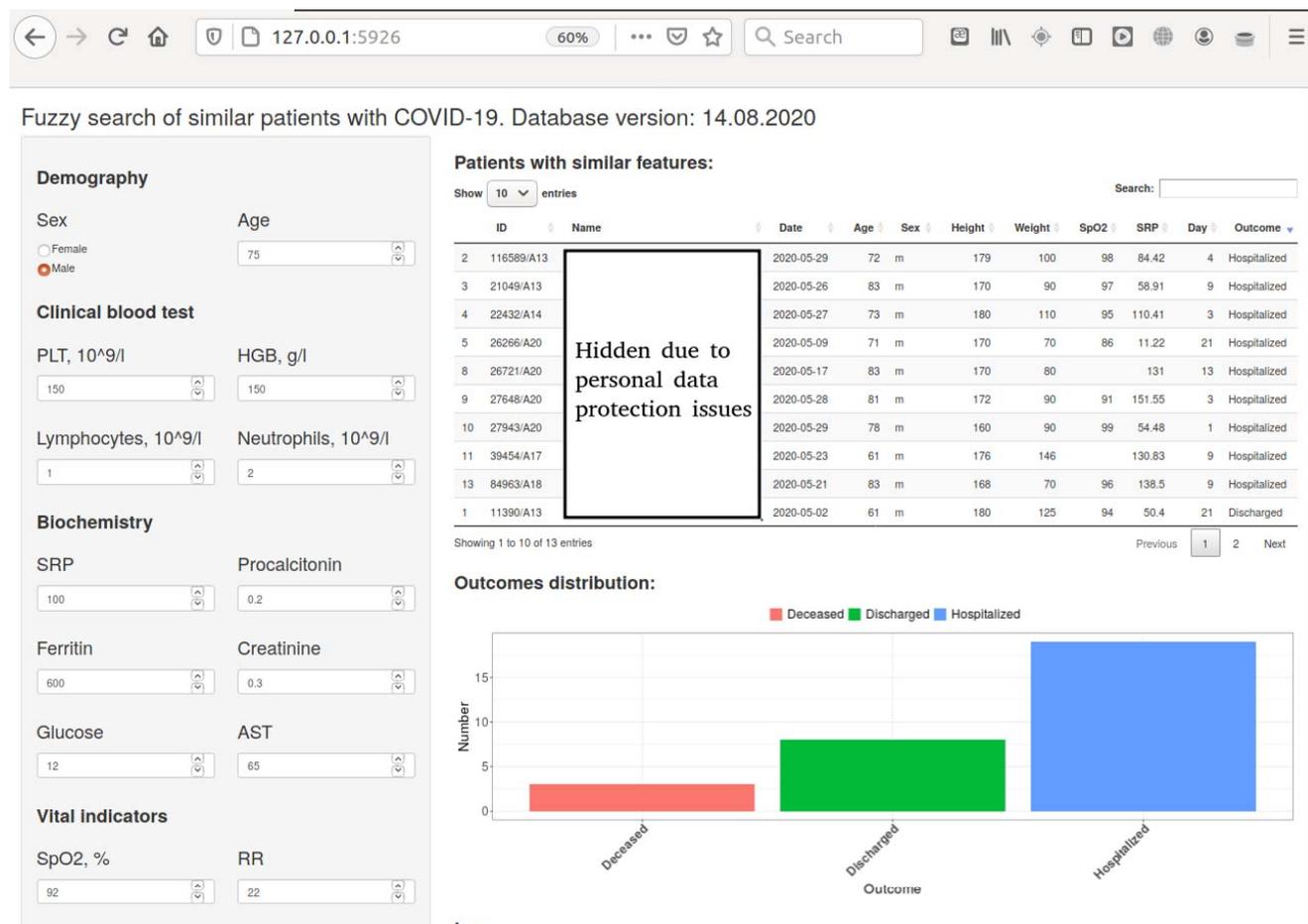


Fig. 4: Graphical user interface

## Code availability

Not applicable.

## Authors' contributions

EB designed the proposed system, OS and EB implemented software modules of DSS, DL coordinated testing of the system, AK encouraged the the team to prototype the DSS and supervised the project.

## References

- Grange, E.S., Neil, E.J., Stoffel, M., Singh, A.P., Tseng, E., Resco-Summers, K., Fellner, B.J., Lynch, J.B., Mathias, P.C., Mauritz-Miller, K., Sutton, P.R., Leu, M.G.: Responding to COVID-19: The UW Medicine Information Technology Services Experience. *Applied Clinical Informatics* **11**(02), 265–275 (2020). DOI 10.1055/s-0040-1709715. URL <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0040-1709715>
- Govindan, K., Mina, H., Alavi, B.: A decision support system for demand management in healthcare supply chains considering the epidemic outbreaks: A case study of coronavirus disease 2019 (COVID-19). *Transportation Research Part E: Logistics and Transportation Review* **138**, 101967 (2020). DOI 10.1016/j.tre.2020.101967. URL <https://linkinghub.elsevier.com/retrieve/pii/S1366554520306189>
- Decision Support System by Sapio Analytics. URL <https://www.sapioanalytics.com/covid-19-decision-support-system/>
- Vida Decision Support System for COVID-19. URL <https://www.media.mit.edu/projects/vida-decision-support-system/overview/>
- Reid, J., Zeng, C., Wood, D.: Combining Social, Environmental and Design Models to Support the Sustainable Development Goals. In: 2019 IEEE Aerospace Conference, pp. 1–13. IEEE, Big Sky, MT, USA (2019). DOI 10.1109/AERO.2019.8741623. URL <https://ieeexplore.ieee.org/document/8741623/>
- Li, L., Yang, Z., Dang, Z., Meng, C., Huang, J., Meng, H., Wang, D., Chen, G., Zhang, J., Peng, H., Shao, Y.: Propagation analysis and prediction of the COVID-19. *Infectious Disease Modelling* **5**, 282–292 (2020). DOI 10.1016/j.idm.2020.03.002. URL <https://linkinghub.elsevier.com/retrieve/pii/S2468042720300087>
- Wang, L., Li, J., Guo, S., Xie, N., Yao, L., Cao, Y., Day, S.W., Howard, S.C., Graff, J.C., Gu, T., Ji, J., Gu, W., Sun, D.: Real-time estimation and prediction of mortality caused by COVID-19 with patient information based algorithm. *Science of The Total Environment* **727**, 138394 (2020). DOI 10.1016/j.scitotenv.2020.138394. URL <https://linkinghub.elsevier.com/retrieve/pii/S0048969720319070>

8. Solutions that relieve healthcare services in the covid-19 crisis. URL <http://healthtechnordic.com/covid19/>
9. Yan, L., Zhang, H.T., Goncalves, J., Xiao, Y., Wang, M., Guo, Y., Sun, C., Tang, X., Jing, L., Zhang, M., Huang, X., Xiao, Y., Cao, H., Chen, Y., Ren, T., Wang, F., Xiao, Y., Huang, S., Tan, X., Huang, N., Jiao, B., Cheng, C., Zhang, Y., Luo, A., Mombaerts, L., Jin, J., Cao, Z., Li, S., Xu, H., Yuan, Y.: An interpretable mortality prediction model for COVID-19 patients. *Nature Machine Intelligence* **2**(5), 283–288 (2020). DOI 10.1038/s42256-020-0180-7. URL <http://www.nature.com/articles/s42256-020-0180-7>
10. Jehi, L., Ji, X., Milinovich, A., Erzurum, S., Rubin, B., Gordon, S., Young, J., Kattan, M.W.: Individualizing Risk Prediction for Positive COVID-19 Testing. *Chest* p. S0012369220316548 (2020). DOI 10.1016/j.chest.2020.05.580. URL <https://linkinghub.elsevier.com/retrieve/pii/S0012369220316548>
11. Song, L., Dong, Y., Xu, M., Shi, D., Guo, W., Li, Y., Zhang, X., Ma, X.: Analysis of prediction and early warning indexes of patients with COVID-19. *Expert Review of Respiratory Medicine* pp. 1–4 (2020). DOI 10.1080/17476348.2020.1793674. URL <https://www.tandfonline.com/doi/full/10.1080/17476348.2020.1793674>
12. Liu, Y., Wang, Z., Ren, J., Tian, Y., Zhou, M., Zhou, T., Ye, K., Zhao, Y., Qiu, Y., Li, J.: A COVID-19 Risk Assessment Decision Support System for General Practitioners: Design and Development Study. *Journal of Medical Internet Research* **22**(6), e19786 (2020). DOI 10.2196/19786. URL <http://www.jmir.org/2020/6/e19786/>
13. Ji, D., Zhang, D., Xu, J., Chen, Z., Yang, T., Zhao, P., Chen, G., Cheng, G., Wang, Y., Bi, J., Tan, L., Lau, G., Qin, E.: Prediction for Progression Risk in Patients With COVID-19 Pneumonia: The CALL Score. *Clinical Infectious Diseases* p. ciaa414 (2020). DOI 10.1093/cid/ciaa414. URL <https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa414/5818317>
14. Wu, G., Yang, P., Xie, Y., Woodruff, H.C., Rao, X., Guiot, J., Frix, A.N., Louis, R., Moutschen, M., Li, J., Li, J., Yan, C., Du, D., Zhao, S., Ding, Y., Liu, B., Sun, W., Albarello, F., D’Abramo, A., Schinina, V., Nicastrì, E., Occhipinti, M., Barisione, G., Barisione, E., Halilaj, I., Lovinfosse, P., Wang, X., Wu, J., Lambin, P.: Development of a Clinical Decision Support System for Severity Risk Prediction and Triage of COVID-19 Patients at Hospital Admission: an International Multicenter Study. *European Respiratory Journal* p. 2001104 (2020). DOI 10.1183/13993003.01104-2020. URL <http://erj.ersjournals.com/lookup/doi/10.1183/13993003.01104-2020>
15. Wynants, L., Van Calster, B., Collins, G.S., Riley, R.D., Heinze, G., Schuit, E., Bonten, M.M.J., Dahly, D.L., Damen, J.A.A., Debray, T.P.A., de Jong, V.M.T., De Vos, M., Dhiman, P., Haller, M.C., Harhay, M.O., Henckaerts, L., Heus, P., Kreuzberger, N., Lohmann, A., Luijken, K., Ma, J., Martin, G.P., Andaur Navarro, C.L., Reitsma, J.B., Sergeant, J.C., Shi, C., Skoetz, N., Smits, L.J.M., Snell, K.I.E., Sperrin, M., Spijker, R., Steyerberg, E.W., Takada, T., Tzoulaki, I., van Kuijk, S.M.J., van Royen, F.S., Verbakel, J.Y., Wallisch, C., Wilkinson, J., Wolff, R., Hooft, L., Moons, K.G.M., van Smeden, M.: Prediction models for diagnosis and prognosis of covid-19: systematic review and critical appraisal. *BMJ* p. m1328 (2020). DOI 10.1136/bmj.m1328. URL <http://www.bmj.com/lookup/doi/10.1136/bmj.m1328>
16. Geleris, J., Sun, Y., Platt, J., Zucker, J., Baldwin, M., Hripcsak, G., Labella, A., Manson, D.K., Kubin, C., Barr, R.G., Sobieszczyk, M.E., Schluger, N.W.: Observational Study of Hydroxychloroquine in Hospitalized Patients with Covid-19. *New England Journal of Medicine* **382**(25), 2411–2418 (2020). DOI 10.1056/NEJMoa2012410. URL <http://www.nejm.org/doi/10.1056/NEJMoa2012410>
17. Boulware, D.R., Pullen, M.F., Bangdiwala, A.S., Pastick, K.A., Lofgren, S.M., Okafor, E.C., Skipper, C.P., Nascene, A.A., Nicol, M.R., Abassi, M., Engen, N.W., Cheng, M.P., LaBar, D., Lother, S.A., MacKenzie, L.J., Drobot, G., Marten, N., Zarychanski, R., Kelly, L.E., Schwartz, I.S., McDonald, E.G., Rajasingham, R., Lee, T.C., Hullsiek, K.H.: A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19. *New England Journal of Medicine* p. NEJMoa2016638 (2020). DOI 10.1056/NEJMoa2016638. URL <http://www.nejm.org/doi/10.1056/NEJMoa2016638>
18. Coronapod: The Surgisphere scandal that rocked coronavirus drug research. URL <https://www.nature.com/articles/d41586-020-01790-y>
19. Mehra, M.R., Ruschitzka, F., Patel, A.N.: Retraction—Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *The Lancet* **395**(10240), 1820 (2020). DOI 10.1016/S0140-6736(20)31324-6. URL <https://linkinghub.elsevier.com/retrieve/pii/S0140673620313246>
20. Mehra, M.R., Desai, S.S., Kuy, S., Henry, T.D., Patel, A.N.: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *New England Journal of Medicine* **382**(25), e102 (2020). DOI 10.1056/NEJMoa2007621. URL <http://www.nejm.org/doi/10.1056/NEJMoa2007621>
21. Karr, A.F., Taylor, M.T., West, S.L., Setoguchi, S., Kou, T.D., Gerhard, T., Horton, D.B.: Comparing record linkage software programs and algorithms using real-world data. *PLOS ONE* **14**(9), e0221459 (2019). DOI 10.1371/journal.pone.0221459. URL <https://dx.plos.org/10.1371/journal.pone.0221459>
22. Zhang, Y., Guo, S.L., Han, L.N., Li, T.L.: Application and Exploration of Big Data Mining in Clinical Medicine. *Chinese Medical Journal* **129**(6), 731–738 (2016). DOI 10.4103/0366-6999.178019. URL <http://journals.lww.com/00029330-201603200-00015>
23. Rastegar-Mojarad, M., Kadolph, C., Ye, Z., Wall, D., Murali, N., Lin, S.: A Fuzzy-Match Search Engine for Physician Directories. *JMIR Medical Informatics* **2**(2), e30 (2014). DOI 10.2196/medinform.3463. URL <http://medinform.jmir.org/2014/2/e30/>
24. Liu, Z., Weng, J., Li, J., Yang, J., Fu, C., Jia, C.: Cloud-based electronic health record system supporting fuzzy keyword search. *Soft Computing* **20**(8), 3243–3255 (2016). DOI 10.1007/s00500-015-1699-0. URL <http://link.springer.com/10.1007/s00500-015-1699-0>
25. Gombor, S., Callahan, A., Califf, R., Harrington, R., Shah, N.H.: It is time to learn from patients like mine. *npj Digital Medicine* **2**(1), 16 (2019). DOI 10.1038/s41746-019-0091-3. URL <http://www.nature.com/articles/s41746-019-0091-3>
26. Li, P., Yates, S.N., Lovely, J.K., Larson, D.W.: Patient-like-mine: A real time, visual analytics tool for clinical decision support. In: 2015 IEEE International Conference on Big Data (Big Data), pp. 2865–2867. IEEE, Santa Clara, CA, USA (2015). DOI 10.1109/BigData.2015.7364104. URL <http://ieeexplore.ieee.org/document/7364104/>
27. Longhurst, C.A., Harrington, R.A., Shah, N.H.: A ‘Green Button’ For Using Aggregate Patient Data At The Point Of Care. *Health Affairs* **33**(7), 1229–1235 (2014). DOI 10.1377/hlthaff.2014.0099. URL <http://www.healthaffairs.org/doi/10.1377/hlthaff.2014.0099>
28. R: A language and environment for statistical computing. Tech. rep., R Core Team, Vienna, Austria (2019). URL <https://www.R-project.org/>
29. Wickham, H., Henry, L.: tidy: Tidy Messy Data (2019). URL <https://CRAN.R-project.org/package=tidy>
30. Kuhn, M.: caret: Classification and Regression Training (2020). URL <https://CRAN.R-project.org/package=caret>

31. Wickham, H.: *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York (2016). URL <https://ggplot2.tidyverse.org>
32. Kassambara, A.: *ggpubr: 'ggplot2' Based Publication Ready Plots* (2019). URL <https://CRAN.R-project.org/package=ggpubr>
33. Chang, W., Cheng, J., Allaire, J.J., Xie, Y., McPherson, J.: *shiny: Web Application Framework for R* (2020). URL <https://CRAN.R-project.org/package=shiny>
34. Last Observation Carried Forward. In: *Encyclopedia of Research Design*. SAGE Publications, Inc., 2455 Teller Road, Thousand Oaks California 91320 United States (2010). DOI 10.4135/9781412961288.n211. URL <http://methods.sagepub.com/reference/encyc-of-research-design/n211.xml>
35. Abu Alfeilat, H.A., Hassanat, A.B., Lasassmeh, O., Tarawneh, A.S., Alhasanat, M.B., Eyal Salman, H.S., Prasath, V.S.: Effects of Distance Measure Choice on K-Nearest Neighbor Classifier Performance: A Review. *Big Data* 7(4), 221–248 (2019). DOI 10.1089/big.2018.0175. URL <https://www.liebertpub.com/doi/10.1089/big.2018.0175>
36. The RECOVERY Collaborative Group: Dexamethasone in Hospitalized Patients with Covid-19 — Preliminary Report. *New England Journal of Medicine* p. NEJMoa2021436 (2020). DOI 10.1056/NEJMoa2021436. URL <http://www.nejm.org/doi/10.1056/NEJMoa2021436>
37. Hron, J.D., Parsons, C.R., Williams, L.A., Harper, M.B., Bourgeois, F.C.: Rapid Implementation of an Inpatient Telehealth Program during the COVID-19 Pandemic. *Applied Clinical Informatics* 11(03), 452–459 (2020). DOI 10.1055/s-0040-1713635. URL <http://www.thieme-connect.de/DOI/DOI?10.1055/s-0040-1713635>

# Figures

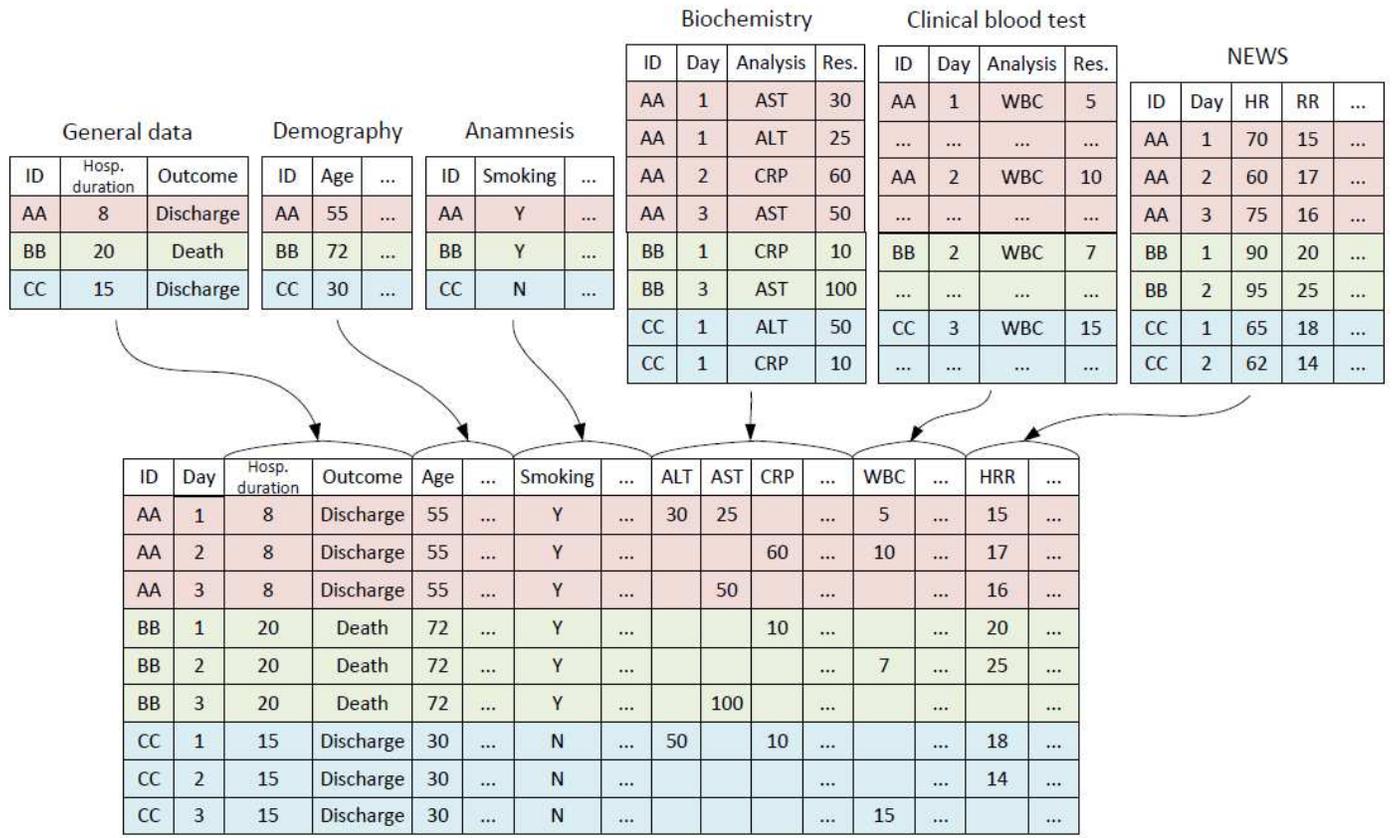
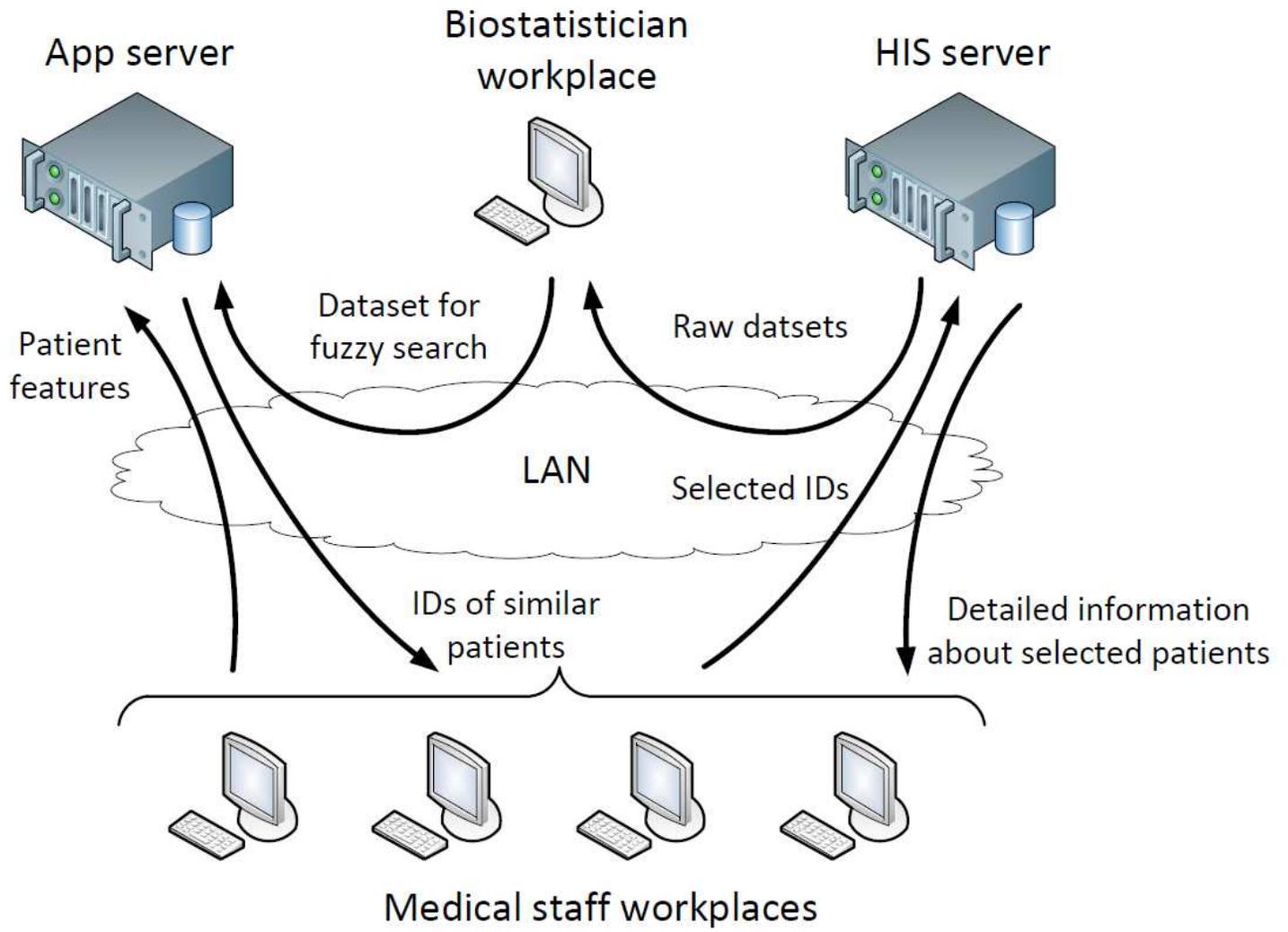


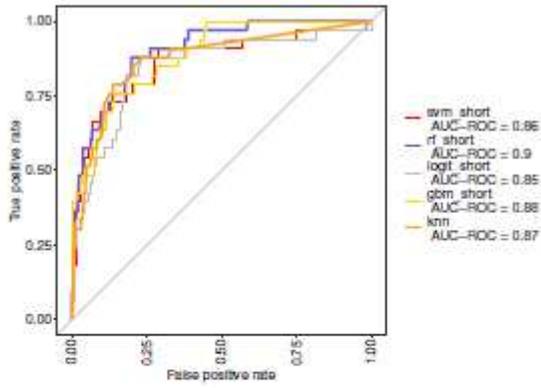
Figure 1

The principle of formation of the initial dataset

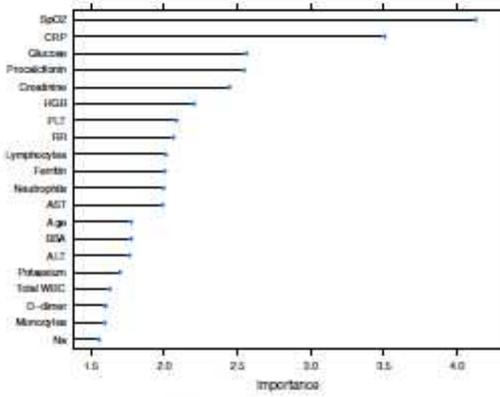


**Figure 2**

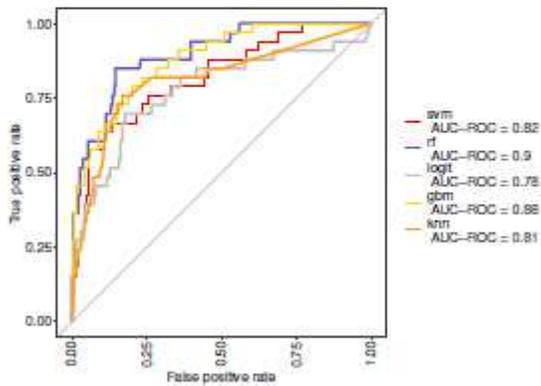
The structure of the developed DSS



(a) ROC curves for 5 classifiers built with the use of all predictors



(b) The feature's importance plot



(c) ROC curves for 5 classifiers built with the use of top-12 predictors only

### Figure 3

Results of a features selection procedure

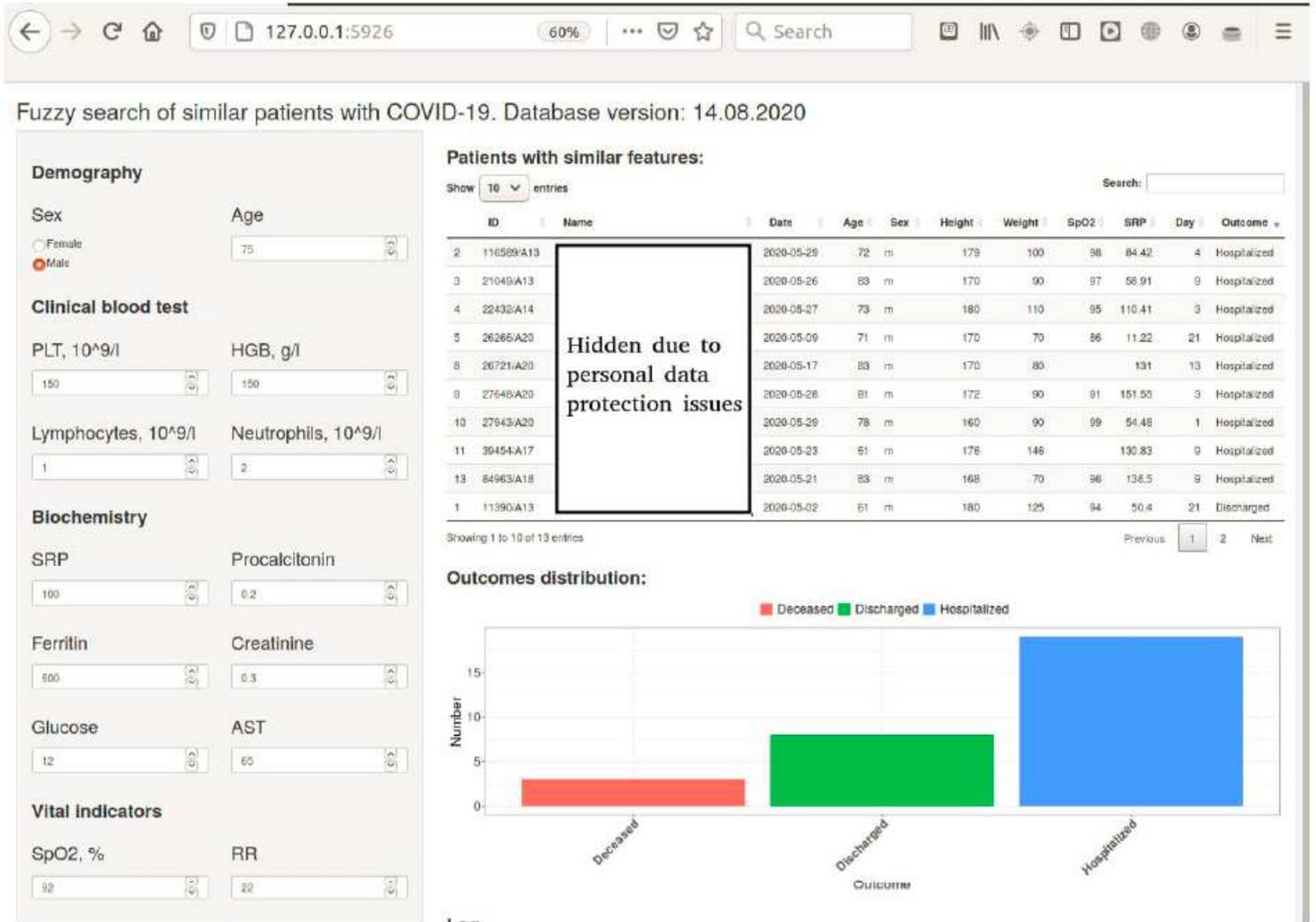


Figure 4

Graphical user interface