Interactive Visualization of Machine Learning Model Results Predicting Infection Risk

S. Schäfer[1] , T. Baumgartl[1,2] , Infection Control Study Group[3], A. Wulff[4,5] , A. Kuijper[6] ,
M. Marschollek[4] S. Scheithauer[7], and T. von Landesberger[1,2]

[1]TU Darmstadt, Germany
[2]University of Cologne, Germany
[3]HiGHmed Project Management Office, Health Data Science Unit, Heidelberg University Hospital
[4]Peter L. Reichertz Institute for Medical Informatics of TU Braunschweig and Hannover Medical School, Hannover, Germany
[5]Big Data in Medicine, Department of Health Services Research Carl von Ossietzky University Oldenburg, Oldenburg, Germany
[6]Fraunhofer IGD, Darmstadt Germany
[7] Department of Infection Control and Infectious Diseases, University Medical Center Göttingen, Georg-August University Göttingen, Germany

Abstract
We present a novel visual-interactive interface to show results of a machine learning algorithm, which predicts the infection probability for patients in hospitals. The model result data is complex and needs to be presented in a clear and intuitive way to microbiology and infection control experts in hospitals. Our visual-interactive interface offers linked views which allow for detailed analysis of the model results. Feedback from microbiology and infection control experts showed that they were able to extract new insights regarding outbreaks and transmission pathways.

CCS Concepts
• *Human-centered computing* → *Visual analytics;* • *Computing methodologies* → *Artificial intelligence;*

1. Introduction

A high occurrence of infectious diseases in a hospital is a thread for patients and hospital staff. A particular threat are pathogens which have developed resistance to multiple antibiotics as well as the new infections caused by SARS-CoV-2 as part of the worldwide pandemic. Infections occur in outbreaks in a temporally and spatially clustered manner. A promising strategy to reduce new infections is to detect high occurrence of pathogens at an early stage and to trace transmission routes. For clinicians and hygienists (for simplicity 'experts') it is currently very difficult to monitor the occurrence of infections. Relevant data is only available in tabular format and is neither visually processed nor meaningfully linked. This results in a high amount of time-expensive, manual labor. To help predicting infection risk of a patient, a machine learning model was created and used. The dataset contained over one million test results of patients collected from 2010 to 2014. In order to extract high-level patterns such as transmission pathways and high pathogen occurence (so-called "clusters") the data needs to be visualized in a compact view.

Related Work Previous work combines visual analysis of patient treatments, results, contagion and transmission paths [BPW*21]. First attempts at combining these data with machine learning results have been made [MPW*20]. However instead of focusing on individual patients, this work focuses on whole hospital. Visu-

alizing machine learning output results in new challenges for the viewer. Only displaying the output is not enough. For understanding the output, also the factors determining the output and their impact need to be shown. The field of explainalble AI (XAI) provides possible solutions to create explanations for machine learning models. [GMR*18] SHAP (SHapley Additive exPlanations) is a game theoretic approach to explain the output of any machine learning model. SHAP has shown promising results in the field of XAI and was already applied to medical applications [LL17] [LNV*18].

2. The Visual Interface and Use Case

The presented interactive interface consists of multiple linked views. It aggregates the data and shows prediction and patient location for different parts of the hospital in a treemap, which is enriched by the patient movements between wards. The averaged infection prediction is encoded by color and the patient count by size. Time series analysis is possible by interacting with the stream-graph. The visual interface is accessed via a recent webbrowser and implemented with React [GP15] and D3 [Dew12]. The SHAP values are calculated using the python library SHAP. We describe the interface's features with a typical workflow of an hospital's hygienist (cf. supplemental video)
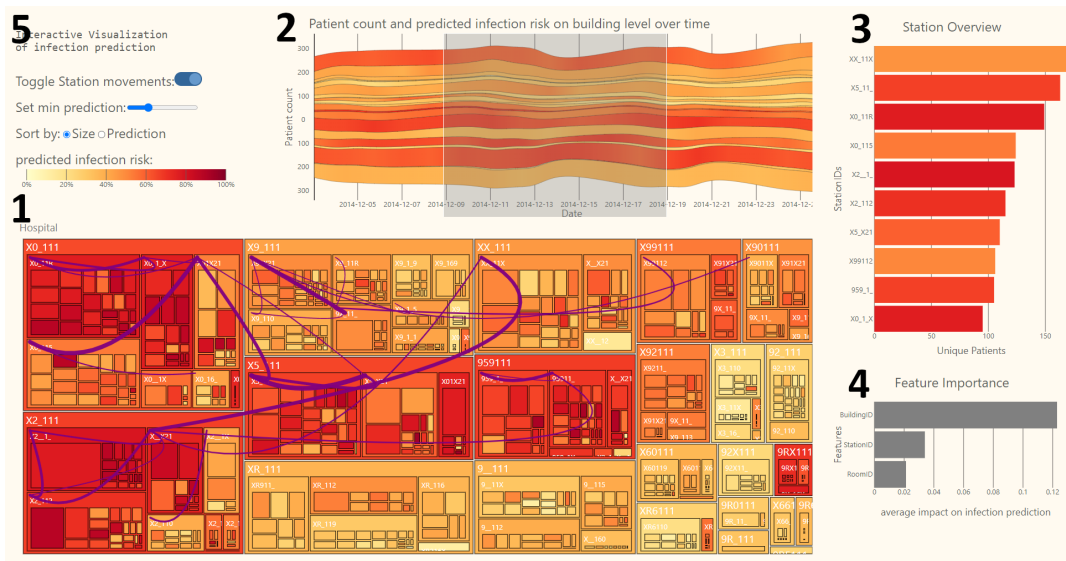
Figure 1: *The visual interface as a web application. The current state of the hospital can be seen on the treemap (1). Time series analysis can be done by selecting a time frame in the streamgraph (2). The overview shows the information in an compact manner (3). Importance of the three most important features is listed in (4). The menu can be used to sort the visualizations and set a minimal threshold (5)*

(1) Overview

A treemap provides the expert with an overview of the current infection predictions and patient counts for all relevant hospital entities, such as rooms, wards and buildings. The treemap is very space efficient, allowing the whole hospital to be visualized. Thus direct comparisons between entities can be made. By clicking on an entity on the treemap, a zoomed-in version appears. Overlaid on top are the patients' movements between wards as a network diagram. The width of a curve encodes the number of patients. The movements are displayed bidirectional between wards by using different angles for each curve [FWD*03].

(2) Time-dependent View

A streamgraph is used to display the changes in patient count and predicted infection risk over time. Each bin represents a building in the hospital. The streamgraph and treemap are linked together. This allows the expert to select a time frame in the streamgraph and get real-time updates on the treemap for the selected time frame. On the other hand, if the expert selects a location in the treemap the streamgraph will highlight the corresponding data.

(3) Create a compact overview

While treemap and streamgraph allow to explore the complete hospital on any level at any given time-frame, we also want an easy to understand summary, so that anyone can detect high pathogen concentrations. For this case, a bar chart is used. It is linked with the streamgraph, meaning that it automatically updates when a time frame is selected. The expert can also click on the bar to get more information about the corresponding ward in the treemap.

(4) Explain and verify predictions

A bar chart is used to display the feature importance generated by calculating SHAP values. It shows the three most important pa-

rameters, which all belong to the location of the patient. Since the SHAP values are calculated beforehand, they are not updated by filtering.

(5) Interaction

The expert can set the minimal infection probability for filtering shown patients. It is also possible to order the nodes in the treemap and the bar chart by predicted infection risk or patient count.

3. Expert Feedback and Future Work

The visual interface was presented to a group of experts and a group of non-experts. They were asked for their opinion on the functionality, comprehensibility, and desired extensions. Both groups successfully identified transmissions patterns by using the movement network and sorting the nodes by prediction. The experts commented, that they could detect expected patterns in the streamgraph. In sum, they praised the layout and the methods of interaction. Having all views linked was very well received. However, both groups had difficulties identifying the direction of the curves in the movement network. Moreover, the experts wished to have a function exporting a list of selected patients from the treeemap. Since geographic data was not provided, future work may also try to sort wards and buildings in the treemap by physical distance. It is also possible to try different streamgraph layouts that feature a non-centered y-axis. The visualization can be integrated into different decision support applications to support domain experts in their infection surveillance.

References

[BPW*21] BAUMGARTL T., PETZOLD M., WUNDERLICH M., HOHN M., ARCHAMBAULT D., LIESER M., DALPKE A., SCHEITHAUER S., MARSCHOLLEK M., EICHEL V. M., MUTTERS N. T., CONSORTIUM H., LANDESBERGER T. V.: In search of patient zero: Visual analytics of pathogen transmission pathways in hospitals. *IEEE Transactions on Visualization and Computer Graphics 27*, 2 (2021), 711–721. `doi:10.1109/TVCG.2020.3030437`. 1

[Dew12] DEWAR M.: *Getting Started with D3: Creating Data-Driven Documents*. " O'Reilly Media, Inc.", 2012. 1

[FWD*03] FEKETE J., WANG D., DANG N., ARIS A., PLAISANT C.: Interactive poster: Overlaying graph links on treemaps. In *Proceedings of the IEEE Symposium on Information Visualization Conference Compendium (InfoVis 03)* (2003), Citeseer, pp. 82–83. 2

[GMR*18] GUIDOTTI R., MONREALE A., RUGGIERI S., TURINI F., GIANNOTTI F., PEDRESCHI D.: A survey of methods for explaining black box models. *ACM Comput. Surv. 51*, 5 (Aug. 2018). `doi:10.1145/3236009`. 1

[GP15] GACKENHEIMER C., PAUL A.: *Introduction to React*, vol. 52. Springer, 2015. 1

[LL17] LUNDBERG S. M., LEE S.-I.: A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30*, Guyon I., Luxburg U. V., Bengio S., Wallach H., Fergus R., Vishwanathan S., Garnett R., (Eds.). Curran Associates, Inc., 2017, pp. 4765–4774. 1

[LNV*18] LUNDBERG S. M., NAIR B., VAVILALA M. S., HORIBE M., EISSES M. J., ADAMS T., LISTON D. E., LOW D. K.-W., NEWMAN S.-F., KIM J., ET AL.: Explainable machine-learning predictions for the prevention of hypoxaemia during surgery. *Nature Biomedical Engineering 2*, 10 (2018), 749. 1

[MPW*20] MÜLLER M., PETZOLD M., WUNDERLICH M., BAUMGARTL T., HÖHN M., EICHEL V., MUTTERS N. T., SCHEITHAUER S., MARSCHOLLEK M., LANDESBERGER T. V.: Visual Analysis for Hospital Infection Control using a RNN Model. In *EuroVis Workshop on Visual Analytics (EuroVA)* (2020), Turkay C., Vrotsou K., (Eds.), The Eurographics Association. `doi:10.2312/eurova.20201090`. 1