Challenges of Trustable AI and Added-Value on Health B. Séroussi et al. (Eds.) © 2022 European Federation for Medical Informatics (EFMI) and IOS Press. This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0). doi:10.3233/SHTI220529

Using Explainable Artificial Intelligence Models (ML) to Predict Suspected Diagnoses as Clinical Decision Support

 Zully RITTER^{a,1}, Stefan VOGEL^a, Frank SCHULTZE^b, Kerstin PISCHEK-KOCH^a, Wiebke SCHIRRMEISTER^c, Felix WALCHER^{c,e}, Rainer RÖHRIG^{d,e}, Tibor KESZTYÜS^a, Dagmar KREFTING^a and Sabine BLASCHKE^{b,e}
^a Institute of Medical Informatics, University Medicine Göttingen, Georg-August University, Göttingen, Germany
^b Central Emergency Department, University Medicine Göttingen, Georg-August University, Göttingen, Germany
^c Department of Trauma Surgery, Otto-von-Guericke University Magdeburg, Magdeburg, Germany
^d AKTIN-Research Group, Germany
^e Institute of Medical Informatics, Medical Faculty of RWTH Aachen University, Aachen, Germany

> Abstract. The complexity of emergency cases and the number of emergency patients have increased dramatically. Due to a reduced or even missing specialist medical staff in the emergency departments (EDs), medical knowledge is often used without professional supervision for the diagnosis. The result is a failure in diagnosis and treatment, even death in the worst case. Secondary: high expenditure of time and high costs. Using accurate patient data from the German national registry of the medical emergency departments (AKTIN-registry, Home - Notaufnahmeregister (aktin.org)), the most 20 frequent diagnoses were selected for creating explainable artificial intelligence (XAI) models as part of the ENSURE project (ENSURE (umg.eu)). 137.152 samples and 51 features (vital signs and symptoms) were analyzed. The XAI models achieved a mean area under the curve (AUC) one-vsrest of 0.98 for logistic regression (LR) and 0.99 for the random forest (RF), and predictive accuracies of 0.927 (LR) and 0.99 (RF). Based on its grade of explainability and performance, the best model will be incorporated into a portable CDSS to improve diagnoses and outcomes of ED treatment and reduce cost. The CDSS will be tested in a clinical pilot study at EDs of selected hospitals in Germany.

> Keywords. Clinical Decision Support, Explainable Artificial Intelligence, Machine Learning, Diagnoses Prediction, Emergency Department

1. Introduction

The number and complexity of emergency cases in emergency departments (EDs) have increased substantially in Germany. High pressure for correct diagnoses in limited time and the reduced number of specialized medical staff in EDs are catalysts leading to

¹ Corresponding Author, Zully Ritter, University Medicine Göttingen, Georg-August University, Institute of Medical Informatics, D-37075, Germany; E-mail: zully.ritter@med.uni-goettingen.de.

erroneous diagnoses. Frequently unproven medical knowledge used for diagnosis identification leads to diagnosis failure or even death in the worst scenery. The German interdisciplinary research project ENSURE (grant ZMVI 2520DAT803) addressed this problem (ENSURE (umg.eu)). The aim is to develop an explainable artificial intelligence machine learning model, to predict suspected diagnoses for patients in the ED. The nationwide collected ED-patient data from the AKTIN registry has been used. The diagnoses suggested by the XAI are to be used as support decisions. Thus, the patient's diagnosis and treatment are to be performed only by the medical doctor.

2. Materials and Methods

On the Data: The AKTIN-Data registry from ED patients (2017-2020) was used to create explainable ML models (eXML) [1]. The most 20 frequent diagnoses were used to train (80%) and test the eXML-models. 137.152 samples matching the target diagnoses and 51 features (vital signs and symptoms), and 1 column with the associated diagnosis (ICD-10 codes) were analyzed. *On the eXML-Models*: Different eXML models were created and compared. Performance was evaluated using a multiclass confusion matrix, ROC curves, and predictive accuracy. Shap was used for explainability. Sensitivity was preferred over specificity. These processes were already tested using open data [2].

3. Results

Following ICD-10 codes were the most frequent diagnoses in Germany: (I21, I26, J44, I10, I20, E87, I63, S06, K56, J18, F10, I48, F10, I48, E11, G45, T78, R55, A41, K80, I50, J16, M54, J15, S72, A40, J17). These diagnoses were confirmed by ED experts matching with those selected in ENSURE. The best values from the eXML logistic regression (LR) and random forest model (RFM) were: predictive accuracy of 0.927 (LR) and 0.99 (RFM). The area under the curve (AUC, "one vs. rest (ovr)" mean) was: 0.98 (LR) and 0.99 (RFM). Feature reduction, grid search, and data balance using data sets of the E.Care ED within the clinical study centers will be made in future analysis.

4. Conclusions

1) Secondary ED data from the German AKTIN-registry database was proven usable to develop explainable machine learning models. 2) Physiological filters for vital signs and symptoms variables were a key factor for achieving high accuracy and grade of explainability. 3) Random Forest, due to the higher accuracy, AUC, and explainability, is recommendable to be integrated into a CDSS for diagnosis prediction in ED.

References

- [1] Gilpin LH, Bau D, Yuan BZ, Bajwa A, Specter M, Kagal L. Explaining Explanations: An Overview of Interpretability of Machine Learning. In: 2018 IEEE 5th International Conference on data science and advanced analytics (DSAA). Turin, Italy; 2018. p. 80–9.
- [2] https://share.streamlit.io/zritter2050/cormeum/main/hrfApp.py