Interpretable AI diagnostics for dyspnea in the emergency department by deep learning and a massive regional health care dataset

Tolestam Heyman, Ellen; Ashfaq, Awais; Ekelund, Ulf; Ohlsson, Mattias; Björk, Jonas; Dahlén Holmqvist, Lina; Khoshnood, Ardavan M.; Lingman, Markus

2023

Document Version:
Publisher's PDF, also known as Version of record

Interpretable AI diagnostics for dyspnea in the emergency department by deep learning and a massive regional health care dataset

ELLEN TOLESTAM HEYMAN[1,2], AWAIS ASHFAQ[3,4], ULF EKELUND[2,5], MATTIAS OHLSSON[2,4], JONAS BJÖRK[2], LINA HOLMQVIST[6,7], ARDAVAN KHOSHNOOD[2,5#], MARKUS LINGMAN[3,4,7#]

[1]HALLAND HOSPITALS; [2]LUND UNIVERSITY; [3]REGION HALLAND; [4]HALMSTAD UNIVERSITY; [5]SKÅNE UNIVERSITY HOSPITAL; [6]SAHLGRENSKA UNIVERSITY HOSPITAL; [7]UNIVERSITY OF GOTHENBURG

Conclusion

- We created an AI-tool for diagnosis in dyspneic adults at time of triage in the emergency department
- We analyzed complete data from an entire regional health care system
- The AI is interpretable for clinicians by placing data in its clinical context and time
- For today, we generate new, machine-derived understanding of unknown important diagnostic predictors
- For tomorrow, we glimpse future individualized medicine

Background & Aim

About half of dyspneic patients in the emergency department (ED) suffer from acute heart failure (AHF), chronic obstructive pulmonary disease exacerbation (COPD-E) and/or pneumonia, which are often misdiagnosed. Studies suggest that a third to a half of elderly patients with these conditions are mistreated during their ED stay. We aimed to design an artificially intelligent (AI) diagnostic support for adult dyspnea patients in ED triage. It should classify visits into AHF, COPD-E, pneumonia, and/or "other diagnoses". The AI-diagnostics should be clinically interpretable and individualized, i.e. based on patient-unique selected variables.

Figure 1. Included data from the complete Region Halland health care system:

- *Primary care*
- *Outpatient specialist care*
- *Ambulance service*
- *Emergency department care*
- *In-hospital care*
- *Self derived factors/other e.g. ordinary medication, blood samples, vital signs*

Methods

In this population-based cohort study, we included all patient visits older than 17 years at any of Region Halland's two EDs during July 1, 2017 to December 31, 2019. We included all collectable structured patient data from the visit´s previous five years, from the complete regional health care system (*figure 1*). Median and mean number of clinical events per patient visit was 1 095 (IQR 459 – 2 310) and 1 747 (STD 1 999). Diagnoses in the subsequent in-hospital or ED discharge notes were used as gold standard. For patients discharged from the ED, three emergency physician specialists reviewed the diagnoses of 1 070 ED visits. We developed a novel AI model, CARENET, with three hierarchically arranged recurrent neural networks.

Results

In all, 10 315 patient visits were included. Most prevalent diagnoses were AHF (15.2%), COPD-E (13.3%) and pneumonia (13.3%). Diagnostic performance had an AUROC of 0.8701 (STD 0.0079) (*Figure 3*). Each patient visit received a unique attention plot (*Figure 2*). Highly weighted earlier clinical events were graphically displayed in their time-period and clinical context. Additional attention profiles on cohort and subgroup level displayed an average group pattern, generating new ideas for important diagnostic variables. For all cohort, an earlier heart failure diagnosis in primary care had highest attention value, together with an earlier COPD diagnosis in ED or in-patient care. Diabetes type 1 was more important among women and smoking and atrial fibrillation among men.

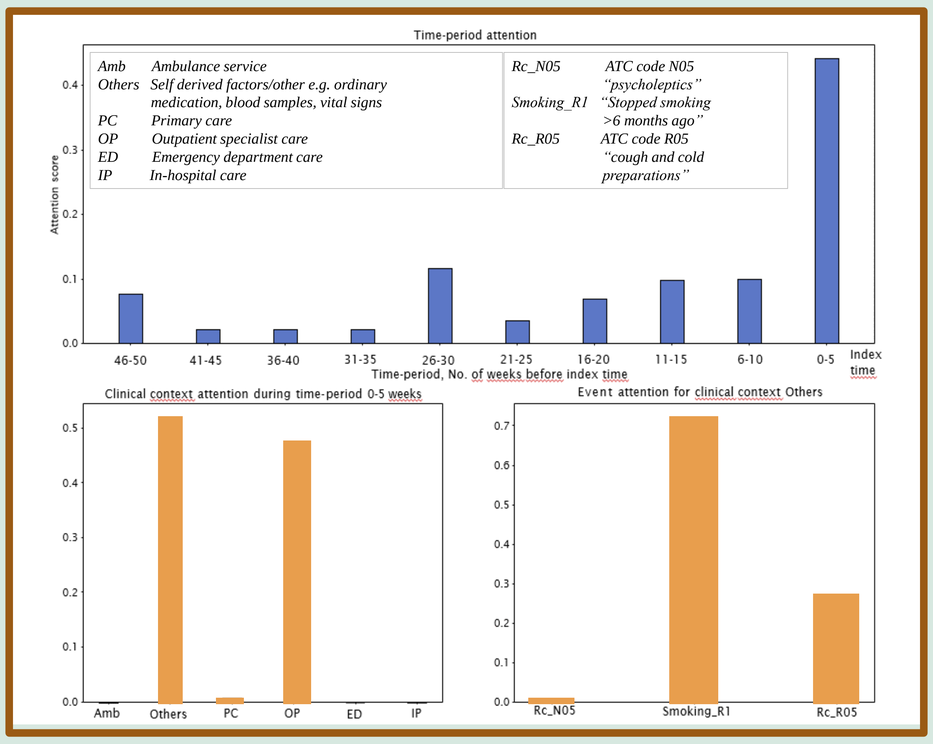
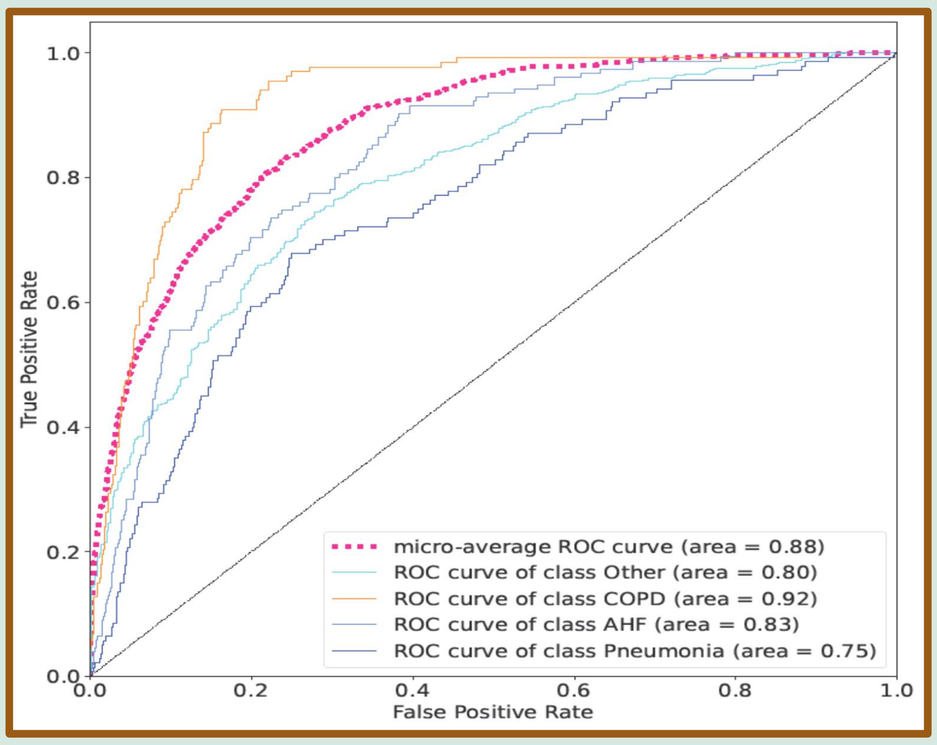


Figure 2. An individual attention plot for a patient with pneumonia. The previous five weeks are most weighted, and its attention is highest for "stopped smoking >6 months ago" and collected medications "R05 cough and cold preparations" and "N05 psycholeptics".

Figure 3. Diagnostic performance at time of triage; overall and each of the diagnosis classes using one year of data. The performance did not improve when using five years of data. Performance was equal in women and men and equal with and without any data from the present ED visit.



Contact Ellen Tolestam Heyman at *ellen.tolestam-heyman@regionhalland.se*,

Acknowledgement Lola Arnarsdottir, David Edberg, Philip Holker and Johanna Nordmark, members of the expert jury; # shared last authorship