

## Preface

This volume contains the papers presented at KDH-2019: The 4th International Workshop on Knowledge Discovery in Healthcare Data held on August 10-16, 2019 in Macao.

The Knowledge Discovery in Health care Data (KDH) workshop series was established in 2016 to bring together AI and clinical researchers, fostering collaborative discussions and presenting AI research efforts to solve pressing problems in health care. This is the workshop's fourth year; held along with IJCAI in Macao, China.

There were 17 submissions. Each submission received at least 1 review, and on the average 3, by program committee members. The committee decided to accept 10 papers. Papers are ordered in the proceedings according to the workshop schedule.

The first session included two papers setting the scene on broad themes of frameworks and repositories. Firstly exploring the validity and authenticity of using crowdsourced annotations for healthcare data and the need for a framework that can automatically correct incorrectly captured annotations of outcomes. The authors argue that such functionality is particularly relevant for Evidence based Medicine (EBM) primarily focusing on health outcomes using a rule based chunking algorithm to recognise and fix errors/flaws. The second paper is directed at the idea of creating a national UK repository of phenotyping algorithms in such a way that a common standard representation is adopted. The authors draw from 70 existing phenotyping algorithms and set out to identify what the hallmarks of the envisioned standard representation should be and argue that they should be based on five criteria: source, terminology, validation, format and implementation.

In the machine learning and classification session, the focus was on explainability and classification as covered by four papers using data from time series to accelerometer to images and noisy medical records. First of the explainability papers presented a framework for deep classification models that can learn prototypical representations during training with time series data. The authors introduce a regularising mechanism to enable direct control over whether the learned prototypes are few and diverse, or many and granular. The next paper provides a qualitative comparison of two Convolution Neural Net (CNN) based feature distillation techniques on a Diabetic Retinopathy image dataset. For many medical fields, the distillation, and hence the explainability of machine vision methods is of great importance. The authors use DenseNet to classify images for identifying the stage of Diabetic Retinopathy and extract the feature maps to identify the regions of focus for a given classification instance. Of the two classification applications the first paper focuses on text classification on veterinary narratives to identify tick parasitism using an ensemble architecture. Here the focus is to combine domain-specific and general word embeddings to overcome challenges with textual data. The last paper explored how mobile accelerometer data can be used to recognise heavy drinking episodes with a view to providing just-in time adaptive interventions to promote healthy behaviors.

The afternoon session focused on KDH applications in clinical trials, medical negligence claims and ICU mortality prediction. The first paper in this session introduces a multi arm bandit technique for adaptive clinical trials with the idea of an explore/exploit approach to maximise patient gains while finding out about new treatments. The authors argue that traditional randomised approaches to clinical trials are often too slow for a rapidly changing modern world. They provide an optimisation to include variation in results as a measure of success, so that treatments that provide consistent results are preferred over those that are inconsistent even if the mean value may be poorer. Next the application of NLP to automatically identify information necessary for medical negligence claims is presented as a means for quickly identifying relevant information among a large volume of longitudinally collected electronic medical record

information. The third paper in this section presents a Bidirectional LSTM mechanism that incorporates both prior medical knowledge and intensive Care Unit (ICU) data for mortality prediction.

The final position paper presents a vision on how to create a decision process for Self-Management and chronic patient support. Here, the authors argue that despite the growing prevalence of multimorbidities, current digital self-management approaches still prioritise single conditions. Specifically, a model-aware and data-agnostic platform is presented on the basis of a tailored self-management plan with three integral concepts - Monitoring (M) multiple information sources to empower Predictions (P) and trigger intelligent Interventions (I).

We very much appreciate the support of the workshop chair, David Sarne, Amal El Fallah Seghrouchni, as well as this year's conference chair Thomas Eiter, and program chair Sarit Kraus. Further we would like to thank Zhiguo Gong for the local arrangements.

We sincerely hope that the participants enjoyed this year's workshop program and that this collection of papers will inspire and encourage more AI-related research for and within healthcare in the future.

July 29, 2019  
Aberdeen

Nirmalie Wiratunga  
Frans Coenen  
Sadiq Sani