NORA2022 Submission 28

Submission information updates are disabled.

For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.

Submission 28	
Title	Improving accuracy of machine learning models for diagnosing irritable bowel syndrome through integration of domain knowledge in data preparation phase
Paper:	(Mar 28, 19:02 GMT)
Author keywords	Machine learning data science feature engineering irritable bowel syndrome
Abstract	Artificial intelligence models are everywhere around us, providing useful suggestions in everyday tasks, assisting experts in decision making process and forecasting critical risks in various areas of application, thanks to the advancements of the past decade in machine learning (ML) and access to the technology through publicly available open source software. Today, the development of highly accurate and sophisticated ML models for a variety of purposes only require suitable sized representative datasets along with few lines of computational code to program high level ML libraries [1]. However, despite of these notable developments, modelling of complex multivariate phenomenon such as biological processes and associated health disorders is often not a trivial task. The curse of high dimensionality and low sample size often associated with medical datasets combined with intrinsic complexity and non-linear nature of underlying associations between variables, challenge normal practices in machine learning based modelling. In such scenarios, the understanding and integration of domain knowledge in the model preparation can help to improve model performance. Our research is focused on ML based modelling of irritable bowel syndrome (IBS) using a medium sized dataset collected from 9 different organs / systems of 100 participants, which include IBS patients and healthy controls (HC) [2]. IBS is known as a functional gastrointestinal disorder, which affects about 10% of global population by altering bowel habits of the patients through diarrhea or constipation (or both) combined with symptoms of increased visceral sensitivity, abdominal pain and bloating. The etiology and pathophysiology of IBS is not yet fully understood and consequently its clinical diagnosis is often complex, subjective and challenging [3]. We aim to address this problem by developing a novel diagnostic criteria for IBS through identification of unique and exclusive patterns, within our dataset using ML techniques, which can also serve as reliable biomarkers in clinics. In our experiments, which are aimed to develop optimally performing models for classifying between IBS and HC, the critical role of understanding and integrating domain / clinical knowledge at various stages of modelling process was observed. Those ML models which incorporated domain knowledge e.g. clinical subgrouping of the disease, patent demographics, co-morbidities etc. in modelling process, outperformed comparable models which did not take domain knowledge into account. On average, 15-20% improvement in model accuracy was achieved merely by integration of domain knowledge, suggesting its critical importance in modelling complex phenomenon in limited sized data.
Submitted	Mar 28, 19:02 GMT
Last update	Mar 28, 19:02 GMT

Authors

first name	last name	email	country	affiliation	Web page	corresponding?
Noman	Haleem	nomanhaleem@gmail.com	Norway	Helse Bergen / Haukeland University Hospital	https://braingut.no/braingutanalytics/	✔