02/10/2020

**To do:**

* In subject part: Need to redefine Alg-E (look at L. Bifano’s thesis)
* In subject part: did not include “management” objective
* In aims and objectives: need to be more precise about the methods used especially in interpretability (but can’t really fill that in just yet)
* In both: did not include the possibility of having “real-time” data from the e-POCT results in Tanzania (will know if available soon enough)
* In both: merge the first sentence of “subject” and the “aim” part

**Title:**

Interpretability of Machine Learning algorithms using e-POCT data for clinicians and validation of the Alg-E platform models

**Subject:**

This thesis aims to better the interpretability of Machine Learning (ML) models – including black box models – in order to generate results and behaviors that are more easily understood by clinicians and therefore generalize, widen and give more credibility to the use of ML models in the medical field (especially in the precise field of medical diagnostic).

Interpretability of ML models will be explored at different levels; at the data level, at the model level and at the output level and will be done through different interpretability and visualization techniques (exploring probabilistic space of the model, intelligent descriptive statistics, confidence intervals for results).

This thesis will also provide a validation test of the Alg-E platform through the use of a dataset pertaining to Ebola infection cases in Tanzania. The models used will be carefully analyzed in order to find insights on a list of important (non-exhaustive) questions) and therefore validate the platform. The models and results themselves will be validated and benchmarked using another set of models and results.

The Alg-E platform is a web-based platform that allows users to upload tabular data in order to output clinical decision algorithms.

The end goal is for clinicians, the end-users, to be able to give criticism concerning the platform and alter it to evolve in that direction.

**Aim:**

This thesis aims to improve the interpretability of the ML models used in the Alg-E web platform at different levels in order to widen and give more credibility to the use of ML models for e-POCT data, then to use these methods as well as comparison to state-of-the-art models to validate the Alg-E platform models and their clinical interpretation.

**Objectives:**

* Review the literature of interpretable methods in the field of ML applied to medicine and other available tools and demonstrate the need for improvement in that domain.
* Manage the Alg-E web platform and ensure integration of different team members’ work into the platform and map the communication and development flows.
* Improve interpretability of the platform results through visualization techniques, understand the clinicians’ requirements for interpretability of data, models and results then implement changes addressing these. There are two main layers to this approach. In a first part a survey of end-users’ expectations will be conducted in order to understand the necessary changes. In a second part, the system itself will be probed and analyzed, using descriptive statistics, confidence intervals, visualization techniques and quantification of uncertainty.
* Validate the platform models using a dataset pertaining to Ebola, then validate the results using another set of models used on the same dataset and finally validate the clinical aspect of results through clinicians’ expertise.

**Timeline:**

Get in touch with Alg-E team (Asap, Tuesday 6th)

Research interpretability (finish book + still 3 papers to read, then will be ongoing during the thesis) – Tuesday 6th

Finish reviewing literature – Tuesday 6th (although more papers will be coming in)

Take notes + write results on these in an organized (not totally formal) way (Tuesday 13th)

Interpretability methods investigation (start on Tuesday 6th)

Get in touch with clinicians (After interpretability shenanigans / during)

Build ML model for Ebola dataset (start on Tuesday 20th ?)

Validate other models for Ebola dataset (in parallel of previous task? Tuesday 20th?)

Write the thesis (start 1 month prior to deadline to get a feel of what needs to be done still)

Issue with the platform need to check what is happening: probably wsl2 ubuntu subsystem port occupied in windows supersystem ? otherwise everything runs smooth on wsl2, on windows nothing works (2 issues: 1 package is not available on windows, there are workarounds but they are filthy, and another package is also sort of not available under the version required).