\section{Interpretation Pipeline}

\paragraph{Building a contextual and comprehensive ML interpretation tool}

We're doing this because we want to integrate some new tools in the platform - but first we need to make sure they're ok

After having outlined the tool functionalities, one can now start implementing these. To get the most explanation value out of results, the idea is to take the end user from overall analysis and allow them to thoroughly investigate individual points, as well as global tendencies.

In order to be able to understand the issue at hand, time was spent figuring out what were the issues reported and commonly seen.

In order to overcome these issues, various methods were implemented.

These were implemented in code, using Python and libraries: pyplot, etc\textbf{ (ADD)}.

\subsection{Interpretation Tools}

A list of methods were looked into in order to help resolve some of the issues aforementioned.

Pertaining to interpretability of machine learning models – especially “black-boxes”, several tools have be found to be useful and efficient. The tools which have been investigated are LIME, Shapley values, SHAP, \textbf{influential instances, outlier investigation -- only through SHAP really}, \textbf{PDP plots, ALE plots}, and some other tools (\textbf{MENTION}).

Some visualization tools were thought about in order to convey the maximum information in a concise and clear way. In medical contexts, just like many applications of ML, the models might deal with a considerable amount of features and data points. Therefore an emphasis was put on the selection of data to be visualized and the way data can be selected in order to convey the most useful information. The plotting technique used are all based on PDP plots \textbf{meaning that they follow the same objective but really they're not based on it at all lol}.

\paragraph{Medical vs. Machine Learning Metrics}

\paragraph{Model Interpretation Tools}

\paragraph{Distribution investigation}

\paragraph{Severity Score}

Now that the tools that can be applied to any data and any model have been implemented, a more use-specific tool is constructed. It is based on a paper by \textbf{[annie + ? INCLUDE AUTHORS]} which details the construction of a severity score for malaria patients.

The severity score is built here using the model predictions and their associated probability. The only prediction we are considering is the malaria positive score – therefore the prediction being equal to 1. To each prediction we attach its probability, then bin predictions according to their probability. In the figure above, there are ten bins, which each are separated by thresholds at equal distances; from a probability of 0 to 10\%, to 10\% to 20\%, and onward up to 90\% to 100\%. A high associated probability reflects a high possibility of being malaria positive, whereas a low associated probability reflects a low possibility of being malaria positive – in effect a high probability of being malaria negative. These amount respectively to a high severity score and a low severity score. Once these bins have been created, a sensitivity and specificity curve is drawn, with the thresholds used being the ones used for the bin grouping. Once the figure is completed, the user can choose to move a threshold which helps select a total number of patients / data points to include. The threshold represents the total number of people that are at least as high-risk as the bin it is located at. Therefore if there are only a certain of treatment doses available, the user can choose to select the 250 people or so most at risk according to the model.

The bins themselves contain an indication of the labels / ground truth within each bin. The color coding indicates the true values of the predictions. It is thus possible to also quickly identify if it is possible to trust the model, if the bin is split between two colors, it shows the model is indecisive at that threshold.

\subsection{Clinicians Feedback}

In parallel to the previous tasks, a demonstration of the platform and follow-up discussion plus a survey was written and conducted on \textbf{(SPECIFY)} a certain number of clinicians and doctors with various backgrounds in statistics and machine learning – or none. This approach is necessary because of the need of machine learning to be interwoven with the medical context, and because the platform on which the data can be used and train needs \textbf{(RUN ON)} to be well integrated into the workflow of clinicians\textbf{ (REMEMBER ITS ALSO MOSTLY FOR ON-SITE).} There are many machine learning tools that have already been developed but few are tailored to the needs of the end-users in the medical field. The way the questionnaire was built was to start from reading papers related to clinicians’ needs in the field of ML, then figuring out which of these needs can be addressed via the platform. Once a first version of the questionnaire was constructed, it was given to several doctors and clinicians with different backgrounds in both mathematics, and machine-learning. Feedback from different doctors was reported. Since the questionnaire focuses on the use of the platform, a demonstration of its functionalities and an explanation of its use was performed before giving out the questionnaire. During this whole process, an emphasis was put on the fact that the platform should be suited to clinicians – therefore features should be added or removed with this in mind. Once the feedback was collected, an analysis on which features to collect was done. These chosen features will be discussed and an analysis of their feasibility and interest will be done before potentially integrating them into the platform over time.

The design of the questionnaire is done in a thorough way, and the choice of people answering the survey was also through. The goal of the survey is to make sure that the prioritization of the team’s concerns is correct according to the doctors. The reporting of results was done in a tabloid manner, with open-ended questions’ answers categorized into different groups. The categorization was made in order to be able to regroup answers and gain insight in a more general manner. However, specific results were still conserved, to be make sure the granularity of answers is preserved.

\paragraph{Platform Use and Tutorials}

In order to make sure that ht platform can be used and that the tools provided are understood, some introductions and guides were developed for the end-users.

Another key component of this approach is to understand the model itself and the ways in which it might be appropriate or not for the data – and problem – at hand. Although doctors are very well suited in describing what is needed for an algorithm to be trusted and useful, the understanding of the algorithm’s internal mechanisms and output might be difficult to understand without prior explanation. Since ML understanding is as important as understanding of the clinical context, the knowledge should be shared in both directions. In an effort to simply the considerable amount of knowledge required to fully understand these models, a guide was created. This guide takes a less mathematical, and more general approach of the machine learning topic and its application in the medical field. Images will be used to exemplify the use of these tools, and a table will be added in order to compare interpretation models on a global level. The table includes fields such as “name”, “output”, an image of its output on a dataset, “usability”, etc… which will help identify if the method interprets a single result, a row of result, a feature, if it is a robust statistic or more of an intuition, etc… The general applicability, computational complexity and overall time to run will also be added into the table.

Methods and model types will also be explained, with an image of how they generally divide the probabilistic space; for example for an SVM, have an image of how the classifier behaves so end-users can get an intuition of what type of dataset it can be used on, and what type of dataset could potentially make it behave oddly. Regularization techniques such as Lasso, and feature choice techniques such as Boruta will also be explained with both texts and images.

Another aspect of this was to design a how-to guide related to the platform use. The guide is split into different parts. A first part pertains to navigating the platform and being able to use all of its functionalities; which amounts to creating a step-by-step user guide. A second part is focused on the explanation of the features and their context and use. Any end-user with a basic background in medical and mathematical sciences should be able to get the most out of the platform , therefore there will be effort put into creating a knowledge base – mostly ML and mathematics oriented, but applied to medical contexts – which can be used in parallel with platform’s algorithms. This guide should help the end-user understand the models but also perhaps understand which model is best suited to what kind of dataset and how some models contain design flaws and should be interpreted considering this. A third and final part will be the creation of medical user cases. These user cases should be focused on the aspects previously discussed. Another aspect of this guide is to have end-users be able to understand their model outputs. Therefore the guide should contain information on what results to expect, what can go wrong, what types of models are known to fail in which cases, etc… Visualization tools and metrics will also be explained and their use will be detailed.

\paragraph{Platform Demonstration}

The platform and its features was demonstrated to clinicians live in order to get their feedback and get some insight on what would be needed or what is unnecessary.

The clinicians know what the platform is doing and get some feedback on what is done.

qualitative assessment by clinicians: what people thought of the platform, what they were looking for, features that were requested to be added, what things weren't fully easy to explain / etc to do list

\paragraph{Tools Demonstration}

The clinicians were shown the interpretation tools in order to get an idea of their ease of use and the overall information they convey.