Background

[ref about drug misuse and new resistant strains of bacteria/virus]

[should also include some stuff about the ML methods we will use]

* WHAT is the problem?
  + Africa, asia, etc: deaths, infectious diseases, compare to what it is in the “West”, give numbers.
    - Detail which factors are probably responsible for this discrepancy between the two geographical zones.
      * Wealth
      * Climate: humid and dry seasons, humid + infectious disease = bank
        + Climate change might make it worse in some places
      * Access to medical care: facilities, clinicians per capita, state of medical infrastructures (and roads to get there), les wealth of knowledge + possible help from neighboring hospitals / else
      * In some cases, clinicians have less knowledge?
      * Less testing of some diseases and less equipment / resources to test
      * Less resources overall really
      * More contact to animals / less safety “measures”, animals often are reservoirs for some types of disease
  + Pandemics
    - Infectious diseases are probably worse and more easily started in these countries?
    - Although sub-Saharan and African countries are well prepared for epidemics and have protocols put in place, they have less resources so it’s still tougher to limit spread?
    - Detecting patterns in real time / soon enough is tough!
  + Non efficient and potentially dangerous use of antibiotics
    - Allocating resources where you don’t need them makes them scarce for other people
    - Erring on the side of Caution
    - Drug resistance
* WHAT is being done CURRENTLY to deal with it:
  + Static algorithms, WHO guidelines: not perfect and suffer from old age
    - Electronic stuff, eCDAs.
      * Still static for most of them but easier to use
  + Bring in computers, data science and ML
    - Tracking people’s health evolution, populations etc
    - Interactivity with user
* WHAT are WE doing that is new and great:
  + Addressing clinical needs specifically: “The machine learning algorithms aim to either 1) predict diagnostics either as well as e-POCT but with reduced resource consumption or 2) to improve the accuracy of the diagnosis according to a retrospectively collected gold standard. It also aims to eventually guide the clinician to the highest value features that would either 1) best to achieve these goals or 2) reduce uncertainty of missing values. The latter point is important given that medical data which often contains a high volume of systematically missing data.”
  + Explainability and interpretability boost
  + Active interactive and ease of use + sharing: web platform
  + What is NEW here: interpretability focus
  + Understanding edge cases and acknowledging them to better understand the model (also part of uncertainty measures)
* Obviously talk about the previous papers related to this one
* Talk about the project and Segway to aims / objectives
* ADD IN: Anything that needs explaining: context, terms, theories
  + ML
  + ML in clinical field
  + Interpretability and explainability in ML, and why it is needed (see above)

From papers I read about context:

* Global warming brings uncertainty and change to infectious disease evolution, some will re emerge so new may emerge[Bangladesh, chowdbury]
  + New diseases are super tough to spot [ref from Sahnoun thesis
  + Some infectious are more during summer, some more winter, some during autumn or spring; rain / dry season does not affect uniformly all infectious diseases; neither does humidity or rainfall
  + WHO: some are following an upward trend! In tropical countries e.g. pneumonia.
* A lot of parasites, viruses, bacterias are influenced by climate: geographic distribution, seasonality (e.g. flu) variability from year to year (dry year), or other trends [Thomson, climate drivers in Africa]
  + Arthropod diseases: a lot in low income countries
    - These especially: lag climate differences because mating, hatching, growth, infection of another host takes some time
  + EWS: early warning systems: climate data can be helpful, but usually it’s enough to know if there is an early seasonal rise in cases
  + Africa has many different climates
* WHO: Infectious diseases are caused by pathogenic microorganisms, such as bacteria, viruses, parasites or fungi; the diseases can be spread, directly or indirectly, from one person to another. [WHO, definition from their website]
* WHO: strong evidence that changing weather patterns associated with climate change are shifting the geographic range, seasonality and intensity of transmission of climate-sensitive infectious diseases. [WHO, 1.5 Health Report, synthesis of health content of Intergovernmental Panel on Climate Change - IPCC]
* Epidemics e.g. ebola was worse outbreak because of crowded urban centers, lack of public health structure [paules, what recent history]
* Pandemics = Spark and spread risk, spark in central and west Africa is high risk, and spread is less easily stopped also[madhav, disease control priorities]
  + pandemics have huge impacts: mortality, morbidity, economics, social disruption, potential violence and political stress
  + risk = dense pops, social inequalities, poverty, some environment things, malnutrition and other comorbidities, lack of clean water and sanitation make it more likely to propagate, etc… these factors spread it well
  + poorly prepared countries = political instability, weak public administration, etc, gaps in outbreak detection and response systems, etc…
* in 2001: healthcare in Niger 800x less investment than in US (AND WHAT ABOUT TODAY) [fenoliar, Africa 21st]
  + WHO: Africa: half of deaths are caused by infectious diseases, just 2% in europe (CHECK)
  + Less per capita spending on health, less availability of physicians
* 60% emerging infectious diseases = zonotopic, 72% of these are from wildlife [jones, global trends]
  + Now: multi-drug resistant tuberculosis, chloroquine-resistant malaria
  + NEED health monitoring
* [WHO]: antibiotics / others are losing effectiveness because of strain resistance: now need new, less efficient, more expensive drugs, which don’t always exist…
  + Economic burden, stumps economic growth if you don’t already have the infrastructure for it
* [infectious diseases world WHO report 2019] children in low income countries die 100x more than high income countries from infectious diseases (die more in general but still)
  + HIV + TB + Malaria + hepatitis + Neglected tropical diseases: 4.3 million deaths 2016 compared to 5.3 (2000): yet still very unfairly spread: low income African and south east asia most at risk (WHAT ABOUT 2019).
  + Risk of malaria greatest for children under 5
  + 2013-2018: 90% low income have less than 10 doctors per 10k; only 5% of high income countries are in that situation.

From papers I read:

* Negatives as positives classification = dangerous. Caution principle gone wrong, usually more in resource poor environments. Also important in some cases to have a stratified triage: separate people based on severity / symptoms etc. [Hartley, malaria-sensitive triage]
  + Also: important to have differential risks in epidemics / dynamic setting
  + Also interests us to have a “threshold” for risk as in figure 2 of our notes: derivation of a malaria-sensitive triage scoring system.
* What is ePOCT, POCT, IMCI, ALMANACH [Keitel, novel electronic algorithm]
  + Why they suck
* Febrile illnesses consultation number, eCDA definition, why useful and why limited .[Keitel, electronic clinical decision algorithms review]
  + Tree based
  + Clinical efficacy, prospect for futures
* Data, disparity in ebola deaths in high vs low income countries, availability of guides and why they are limited w.r.t. Ebola [Hartley , Colubri, ML prognostics]
  + Static
  + Epidemics
  + Mobile application and visualization
  + Basically where we are getting our data from
* CDS issues, new types of CDS[Baron, ML and other tools]
  + Why ML fits this
  + 3 branches: knowledge discovery, curation , application
  + Implementation of new algorithms and link to EHR / clinicians
* Adopting ML for clinicians: need interpretability, transparence as much as possible, explainability, need for working with clinicians and understanding what they need [Tonekaboni, what clinicians want]
  + Algorithms should be very specific in the need they are addressing
  + Criteria for explainability and acceptability
  + Reliable clinical ML design
    - Feature importance; instance level explanations; uncertainty: each have their pros and cons and in some contexts are more / less important
    - More details on some methods of explainability
* Workflow, data, model behavior, shortcomings, goals, example of interface [hilda, black box classifiers]
* Connecting to phone is useful; black box, class imbalance especially in rare diseases [Sullivan, transduction connectivity]
* Statistics vs M, inference vs prediction, why complexity can’t be dealt with statistics [bzdok, stats vs ml]
* Why AI is not adopted in clinics and why some deep learning studies are completely off [nagendran, Ai vs clinicians]
  + Cant be reproduced, everything is “safeguarded”
  + Not tested in real world environments
  + Not enough humans doing the comparator group
  + Straight up lying on title and promises
  + Not peer reviewed enough especially for a fast growing field
  + Media headlines are total bs
* Where are the peer reviews, randomized controlled trials? tough to do [Kelly, challenges]
  + Use metrics that capture real clinical applicability and are understandable
  + Fast growing field but watch out for potential harm
  + 4 goals: AI for medicine
    - Improve experience of care
    - Improve health of populations
    - Reduce per capita costs
    - Improve work life of healthcare providers
  + TRIPOD: transparent reporting of a multivariable prediction model for individual prognosis or diagnosis
  + Comparing algorithms is difficult, but needed for model to be benchmarked, etc
  + Watch out for dataset shift
  + Watch out for accidentally fitting confounders vs true signal
  + Watch out for generalization being impossible because of geographic differences / else
  + Watch out for algorithmic bias: unbalanced classes 🡪 intrinsic bias
  + Outcome noise
  + Of course tradeoff: performance explainability
  + Need for the interaction between computers and clinicians to be beneficial, it is not sufficient to have a great model that is explainable and performant, you need it to be helpful to the task the clinician is undergoing.
* Pandemics: articles are hastily written, disease evolves rapidly, conclusions from different countries and a few months ago don’t necessarily hold up [futoma, myth of generalizability]
  + Geographic: some things you can’t generalize: some clinicians have local practices, some diseases are local, so is the climate, patients and clinicians both. Also, clinicians constantly adapt their behavior with regards to their patient and what they are dealing with, the availability of resources, the local practice norms 🡪 so tough to generalize
  + What about reinforcement learning
* Basically SHAP [lundberg, unified approach]
* Few ML papers have actually helped medicine although so many are published [deo, ML in medicine]
  + Not many competitions 🡪 way less “open sourcing” and sharing of knowledge, mostly because data is sensitive / secret / private but still this is tough and in part explains why models don’t hold up in the real world and why the advance is so slow.
* Treatment, diagnosis, monitoring, transfer learning [tayarani, applications of AI]
  + Monitoring helps mortality and recovery rate (better resource allocation)
  + Images are harder to deal with than tabular data (pipeline is more advanced usually)