CONTENT

OVERALL AIMS

General objective

To develop machine learning algorithms that can predict critical illness events in Malawian children using continuous vital signs monitoring systems specifically updated for use in low-resource settings.

Specific objectives

1. To assess if sociodemographic data are associated with critical illness and if these data can improve the discrimination of machine learning algorithms and conventional models that predict critical illness events, by comparing the accuracy (sensitivity, specificity, false-positive and negative ratios, and area under the curve [AUC]) and the root mean square error (RMSE), and by calculating the net reclassification index (NRI).

2. To assess if a predefined set of biomarkers is associated with critical illness events; and assess if these biomarkers can improve the performance of the machine learning algorithms and conventional models to predict critical illness events by comparing accuracy and RMSE, and by calculating NRI. These markers will include lactate (POC-test), C-Reactive protein, RNA transcription markers (FAM89A and IFI44L) and other novel markers.7

3. To assess if machine learning algorithms and conventional risk prediction models to predict critical illness events differ in features and accuracy when stratified by location: the better-resourced centre (Blantyre) versus a less-resourced centre (Zomba), which may more closely resemble most sub-Saharan hospital settings.

4. To assess if machine learning algorithms and conventional risk prediction models to predict critical illness events differ in model composition and accuracy when stratified by age: infants (aged between 28 days and 12 months) versus older children (12 to 60 months).

5. To compare the accuracy of the conventional risk prediction models with previously published models and the machine learning models.

6. To confirm and better understand associations identified by the machine learning algorithms, by augmenting the conventional risk prediction models with these associations.

OUTCOMES

Critical Illness

The definition of critical illness is based on the WHO ETAT (paediatric Emergency and Triage Assessment and Treatment) updated Manual.[6] Critical illness is defined as a severe problem with the airway, breathing or circulation, or acute deterioration of conscious state.

Since the algorithm will be aimed at predicting deterioration to allow interventions, the chosen endpoints or *critical illness events* are primarily therapies aimed at preventing deterioration or restoring the normal function of these systems. (P)ICU admission and death were further added as these are commonly accepted endpoints in prediction models.

Critical Illness Events

1.	Respiratory	Start or increase of respiratory support: oxygen or CPAP (Non)Invasive ventilation: bag & mask ventilation or intubation Start or increase of bronchodilator support
2.	Circulatory	Transfusion of blood (products) Intravenous fluid bolus of 10ml/kg or more Start or increase of continuous/intermittent inotropic support (IV/IM adrenalin)
3.	Neurological	Decrease in Blantyre Coma Score of 1 point or more Convulsions requiring anticonvulsants
4.	Sepsis	Clinical suspicion of sepsis that has led to the collection of a new blood culture and/or start or change in antibiotic treatment[2]
5.	Other	Start of anti-malarial treatment Objectified hypoglycaemia requiring correction (IV or enteral) Unplanned admission to the (P)ICU Unplanned surgical procedure (including chest drains)
6.	CPR/death	Cardio-pulmonary resuscitation (CPR): resuscitation setting involving chest compressions Death

Primary outcome

The primary outcome will be any critical illness event, excluding death[1]. Hence, the primary outcome will be set to 1 if any of the above critical illness event proxies were observed (i.e., any of the respiratory, circulatory, neurological, sepsis, other or CPR events), and otherwise set to 0.

Secondary outcomes will be indicators (1/0) for whether any of the respiratory, circulatory, neurological, or sepsis events were observed, respectively.

Other parameters collected

- Clinical diagnosis on admission and discharge
- Medical and surgical interventions (including medication)

[1] In sensitivity analysis we will investigate the consequences of including CPR events in the primary outcome, as many CPR events are likely strongly related to the child dying a little later.

- Clinical observations during admission (including manual vital signs)
- IMPALA vital signs monitor alarm logs and system data
- Haematological, biochemical, radiological, and microbiological results
- Overall outcome and duration of hospital admission

MACHINE LEARNING APPROACH

Machine learning approach

To develop robust machine learning driven models that are able to recognise and predict critical illness based on the data collected using the novel monitor combined with other relevant patient data is challenging. This is for two reasons: (1) the type of data (time series data combined with static patient characteristics), and (2) the high number of patients that are available in the dataset. In the end, the goal is to deliver machine learning models that can be applied to the broad group of patients and generalise well to all settings where we envision to apply the novel approach. To address these challenges, we will exploit a variety of state-of-the-art machine learning solutions.

Type of data: time series and static patient characteristics

To analyse temporal sensory data, feature engineering or using temporal machine learning algorithms are commonly used[15] and will both be applied in this project. Feature engineering approaches create features that summarise the sensory signal over time. Such features can act as input to well-known machine learning approaches such as XGBoost[16], Random Forest[17], Decision trees[18] or logistic regression. These models have the advantage that they can usually be easily understood. There is however inherently a bias in the design of such features.

Alternatively, we will also study the usage of tTemporal machine learning models. Such models learn the aforementioned features themselves. We will use LSTM networks[19] and TCNs[20], the latter being a specific variant of convolutional neural networks designed to handle time series. These approaches do not suffer from the disadvantage of the bias in feature design, however, do often require more data and usually lack insight. We will therefore combine these approaches with methods to gain insight into the concepts that have been learned, including LIME[21] and SHAP[22]. Our consortium has previous experience with both approaches in medical settings (see[23, 24]).

Big (amount of) data

While the amount of data collected is enough to develop models based on logistic regression (as calculated before), more advanced machine learning models primarily thrive when data is abundant as applies to this study. While for low resource settings and for our novel measurement device relatively limited data is available, there are rich and extensive datasets available from different (yet similar) settings that can act as a starting point for machine learning. These include the MIMIC dataset[25] and data collected in the Netherlands, Europe and three African datasets. We will start with the development of models based on these large datasets, and use it to: (1) identify the most important variables to incorporate in the models, also based on the constraints of the measurement device we will use; (2) identify the best features we can derive from these variables, and (3) act as a basic predictive model that we will later refine. To be able to reuse models derived from such datasets for our Malawian setting (for the selection of variables and engineering of features this is obvious) we will exploit the concept of transfer learning[26]. Hereby, we will take the pre-trained models and refine/improve them based on the data collected with our developed device.

Ultimately, we are interested in the generalisability of our models across the entire potential population of patients targeted with our approach. While a complete assessment is beyond the scope of this project, we will study generalisability in the following way: (1) we will develop models using data from one center, and apply the found models to data from the second center, and (2) we will study the generalisability in a setup where we do use patients from the same center, but splitting the patients in a training and test portion, leaving the data of 30% of the patients out acting as an independent test set. Overall, we are confident that these approaches will deliver us a generalisable model. We will select the final model based on a combination of performance criteria as well as the level of insight that can be gained from the model.

CONVENTIONAL ANALYSIS APPROACH

Although the primary aim is to develop a machine learning algorithm to predict critical illness events, additionally a conventional analysis approach will be performed and compared against both previous literature (e.g., PEWS, BqSOFA, FEAST, etc.) and the machine learning algorithms. The performance of the conventional analysis approach will be compared to the machine learning approach and previous risk prediction scores by comparing the accuracy (sensitivity, specificity, false-positive and negative ratios, and area under the curve [AUC]), c-statistics, and by decision curves and calibration plots.

To identify predictors associated with the primary outcome (any critical illness event) and secondary outcomes (respiratory, circulatory, neurological, sepsis, other or CPR events, respectively) we will employ logistic regression models using stepwise backward feature selection ($p<0,1$).

Some children may have several critical illness events, to account for this we will also model the primary and secondary outcome as time to event, using survival analysis with multiple events. Initially, Cox proportional hazard regression model will be assessed for its proportional hazard assumption using the Schoenfeld residuals method and ability to deal with sampling weights, control for site shared unobserved characteristics and lack of independence of observations within sites. Then, parametric multilevel survival analysis regression models will be considered by assessing their fit with specific parametric probability distribution functions and their robustness to deal with sampling weights, control for site shared unobserved characteristics and lack of independence of observations within sites. The common survival analysis parametric models (exponential, gamma, Weibull, and log-normal distribution) fit with the underlying probability distribution will be examined by comparing their predicated Cox–Snell residuals to assess concordance with the failure time or log failure time distribution and identify outlier observations.

Potential predictors for the conventional analysis approach

The conventional analysis will not make use of all the data collected in this study as conventional methods are not suited for big data. Therefore, we will have to reduce the amount of data first. Potential predictors will be selected by first selecting a subset of potential predictors based on previous literature and expert opinion and next by limiting the analysis to predictors (1) observed at admission and (2) observed at 4 hour intervals (as opposed to continuous data as generated by the vital signs monitoring system[2]).

The subset of potential predictors will be limited to features from the following domains: socio-demographic (age, sex), socio-economic (educational level mother, wealyh status household), vital/clinical signs (HR, RR, BP, saturation, increased WOB, CRT, cold periph, weak rad pulse, pallor, BCS, AVPU, temp, passed urine, patient able to drink/eat, condition according to nurse, and condition according to caretaker), biomarkers (x, y, z), underlying conditions (HIV, chronic conditions, including malnutrition) and previous admission.

[2] If data were not observed 4 hourly, then we extract this data from the vital signs monitoring system at the same time point as it would have been observed by the nurse during her 4-hourly ward rounds.

REFERENCES

6. Guideline: updates on paediatric emergency triage, assessment, and treatment: care of critically ill children [Internet]. Apps.who.int. 2016 [cited 6 August 2020]. Available from: https://apps.who.int/iris/bitstream/handle/10665/204463/9789241510219_eng.pdf
15. Hoogendoorn M, Funk B. Machine Learning for the Quantified Self. Cognitive Systems Monographs. 2018. DOI: 10.1007/978-3-319-66308-1.
16. Friedman J. machine. The Annals of Statistics. 2001;29(5):1189-1232. DOI: 10.1214/aos/1013203451.
17. Ho TK. Random Decision Forest. Proceedings of the 3rd International Conference on Document Analysis and Recognition, Montreal, 14-16 August 1995, 278-282.
18. Quinlan J. Induction of decision trees. Machine Learning. 1986;1(1):81-106. DOI: 10.1007/BF00116251.
19. Hochreiter S, Schmidhuber J. Long Short-Term Memory. Neural Computation. 1997;9(8):1735-1780. DOI: 10.1162/neco.1997.9.8.1735.
20. Bai S, Kolter J, Koltun V. An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling. arXiv:1803.01271.
21. Ribeiro M, Singh S, Guestrin C. "Why Should I Trust You?". Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2016. DOI: 10.1145/2939672.2939778.
22. Lundberg S, Lee S. A Unified Approach to Interpreting Model Predictions [Internet]. Papers.nips.cc. 2020 [cited 7 August 2020]. Available from: https://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.
23. Hoogendoorn M, el Hassouni A, Mok K, Ghassemi M, Szolovits P. Prediction using patient comparison vs. modeling: A case study for mortality prediction. 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). 2016. DOI: 10.1109/EMBC.2016.7591229.
24. Kop R, Hoogendoorn M, Teije A, Büchner F, Slottje P, Moons L et al. Predictive modeling of colorectal cancer using a dedicated pre-processing pipeline on routine electronic medical records. Computers in Biology and Medicine. 2016; 76:30-38. DOI: 10.1016/j.compbiomed.2016.06.019.
25. Johnson A, Pollard T, Shen L, Lehman L, Feng M, Ghassemi M et al. MIMIC-III, a freely accessible critical care database. Scientific Data. 2016;3(1). DOI: 10.1038/sdata.2016.35.
26. Pan S, Yang Q. A Survey on Transfer Learning. IEEE Transactions on Knowledge and Data Engineering. 2010;22(10):1345-1359. DOI: 10.1109/TKDE.2009.191.
27. Secka F, Herberg J, Sarr I, Darboe S, Sey G, Saidykhan M et al. Bacteremia in Childhood Life-Threatening Infections in Urban Gambia: EUCLIDS in West Africa. Open Forum Infectious Diseases. 2019;6(9). DOI: 10.1093/ofid/ofz332. PMID: 31660408.
28. Maitland K, Kiguli S, Opoka R, Engoru C, Olupot-Olupot P, Akech S et al. Mortality after Fluid Bolus in African Children with Severe Infection. New England Journal of Medicine. 2011;364(26):2483-2495. DOI: 10.1056/NEJMoa1101549.
29. Calis J, Phiri K, Faragher E, Brabin B, Bates I, Cuevas L et al. Severe Anemia in Malawian Children. New England Journal of Medicine. 2008;358(9):888-899. DOI: 10.1056/NEJMoa072727.
30. Kanyuka M, Ndawala J, Mleme T, Chisesa L, Makwemba M, Amouzou A et al. Malawi and Millennium Development Goal 4: a Countdown to 2015 country case study. The Lancet Global Health. 2016;4(3):e201-e214. DOI: 10.1016/S2214-109X (15)00294-6.
31. Malawi Private Health Sector Assessment [Internet]. Hanshep.org. 2012 [cited 7 August 2020]. Available from: http://www.hanshep.org/resources/shops-country-private-sector-health-assessments/malawi-private-health-sector-assessment.
32. Harris C, Mills R, Seager E, Blackstock S, Hiwa T, Pumphrey J et al. Paediatric deaths in a tertiary government hospital setting, Malawi. Paediatrics and International Child Health. 2018;39(4):240-248. DOI: 10.1080/20469047.2018.1536873.