MEDINFO 2023 — The Future Is Accessible
J. Bichel-Findlay et al. (Eds.)

doi:10.3233/SHTI231089

Equitable Machine Learning for Hypoglycaemia Risk Management

Jhordany RODRIGUEZ[a], Daniel PADILLA[a], Lenert BRUCE[b], Ben THOW[a], and Malcolm PRADHAN[c,1]
[a]*Alcidion, South Yarra, VIC, US*
[b]*Murrumbidgee LHD, NSW, Australia*
[c]*University of Sydney, Sydney, Australia*

Abstract. We developed a machine learning (ML) model for the detection of patients with high risk of hypoglycaemic events during their hospital stay to improve the detection and management of hypoglycaemia. Our model was trained on data from a regional local health care district in Australia. The model was found to have good predictive performance in the general case (AUC 0.837). We conducted subgroup analysis to ensure that the model performed in a way that did not disadvantage population subgroups, in this case based on gender or indigenous status. We found that our specific problem domain assisted us in reducing unwanted bias within the model, because it did not rely on practice patterns or subjective judgements for the outcome measure. With careful analysis for equity there is great potential for ML models to automate the detection of high-risk cohorts and automate mitigation strategies to reduce preventable errors.

Keywords. Machine learning, AI, equity, fairness, diabetes, hypoglycaemia, emr

1. Introduction

1.1. Background

According to the Australian Commission for Quality and Safety in Healthcare, a hospital acquired complication (HAC) is a clinical complication which may be prevented by appropriate risk mitigation strategies. Among the high-priority HACs identified, Hypoglycaemia is reported as part of Endocrine complications [1].

Hypoglycaemia has significant patient impacts which include an increase in mortality and morbidity. Ward-based patients exposed to hypoglycaemia had an increase of 4.1 days in their length of stay [2,3]. The mortality rates of diabetic ward-based inpatients doubled when they developed hypoglycaemia [2].

While risk factors for inpatient hypoglycaemia are well known, some of them evolve dynamically during the patient stay, which limits the effectiveness of a single risk assessment on admission. In addition, evidence supporting interventions to reduce inpatient hypoglycaemia is limited, with education interventions, audit and feedback shown to be ineffective, for the most part, at reducing inpatient hypoglycaemia [4]. Conversely, published evidence supports the use of predictive-informatics risk-alert

[1]Corresponding Author: Malcolm Pradhan, email: malcolm.pradhan@sydney.edu.au.

systems. Such a system, combined with trained nurse responders, managed to reduce the incidence of severe hypoglycaemia by 68% in a single centre [5].

In a busy clinical environment, identifying patients with high risk for hypoglycaemia and then notifying the appropriate staff to improve monitoring of the condition is a logistically difficult task to perform without technology support. This is particularly true in regional and remote hospital environments where there are relatively fewer specialists and staff resources.

1.2. Objective

Our objective for this work was to develop an equitable machine learning (ML) model based on data from the Electronic Medical Record (EMR) to improve the prediction of patients at risk of hypoglycaemia. We intend for this model to assist specialist diabetic nursing staff to identify high-risk patients as well as the clinical factors that present high correlations to hypoglycaemic events. This, in turn, will assist the clinical team in managing the patient's condition. This work was conducted at Murrumbidgee Local Health District (LHD), NSW, a regional healthcare service in Australia.

Equity with respects to ML is often called fairness in the ML literature. It refers to the ability of the model to not cause disadvantage to any specific sector of the population.

In the context of diabetes management, it is well known that several ethnic and social groups present a higher prevalence of the disease compared to others due to reasons including clinical, biological and social factors [6]. We aim to identify, minimise and document any outcome disparities among population subsets that the prediction model might produce in the context of these factors.

The broader goal of this work is to improve the detection and management of high-risk cohorts in real time. Real-time data from the EMR would trigger updates of the ML model which outputs a risk stratification score to identify high-risk patients for hypoglycaemia. At a specified risk threshold, a notification is triggered to diabetic nursing staff who would have a mobile interface to the list of patients along with relevant data. Ongoing monitoring would occur during the hospital admission.

2. Methods

2.1. Study Population

To train and test the model we used a de-identified dataset extracted from an EMR. A patient was included in the dataset if they had an ICD-10 code associated with diabetes or were on diabetic medications. The resulting dataset represented 10,020 diabetic inpatient visits over a 12-month period at Wagga Wagga Base Hospital, Murrumbidgee LHD in NSW.

The outcome (target) was a binary variable indicating 1 if the patient had any hypoglycaemic event during their hospital stay (blood glucose < 4.0).

2.2. Feature Engineering

We selected 28 patient attributes, laboratory results and coded data (both current and historical) that potentially contribute to them experiencing a hypoglycaemic event during their inpatient stay. Attribute selection was based on literature review.

We created two predictive models. The first was based on all information available prior to admission to the ED (historical data) [**admit** model]. To account for the dynamic nature of hypoglycaemia and to integrate the current degree of illness of the patient, we trained a second model [**admit+4h** model] that incorporates data from the first 4 hours after admission to hospital as well as the historical data of diabetic control. The second model included factors such as blood pressure, respiratory rate and $Sp0_2$ as markers of how acutely sick a patient may be, not because of their relation to diabetes specifically.

The factors, including data collected in the first 4-hour after admission time range [admit+4h], were: age of the patient at the time of admission; body mass index (BMI); the type of diabetes; whether the patient is on insulin or not; whether the patient is on any diabetes management medication or not; the indigenous status; gender; a count of comorbidities known at the time of admission; number of blood glucose level (BGL) tests taken [admit+4h]; number of BGL test taken in the 30 days leading to the admission date; number of hypoglycaemic events [admit+4h]; number of hypoglycaemic events recorded in the 30 days leading to the target datetime; result of the last HbA1c test taken; last abnormal heart rate [admit+4h]; last abnormal respiratory rate [admit+4h]; last abnormal temperature [admit+4h]; last abnormal oxygen saturation [admit+4h]; number of BGL tests, average, standard deviation, min and max values in the week and month leading to the target datetime.

2.3. Modelling and Evaluation

We split the 10,020 patient encounters into training and validation sets containing stratified (with respect to the target variable), randomised samples, comprising 70% and 30% of the data, respectively. To compare predictive performance, we trained a logistic regression model and a XGBoost model [7].

We quantified the predictive capability of the models via ROC curves, where a higher Area Under the Curve (AUC) correlates with a more accurate prediction ability overall, while allowing the implementor to trade-off sensitivity and specificity.

3. Results

3.1. Descriptive Analysis

Patients represented in the extracted datasets had the following characteristics: 11.8% indigenous, 88.2% non-indigenous; 57.7% male, 42.3% female.

3.2. Model Performance

Performance of the [admit+4h] model was significantly better than the [admit] model. The [admit+4h] XGBoost model achieved a score of **0.837** AUC which is a good result

for a risk stratification model. The [admit+4h] logistic regression model followed with a **0.801** AUC score (Figure 1). It is worth noting that the [admit] XGBoost model, using data only prior to admission, scored 0.753 AUC.

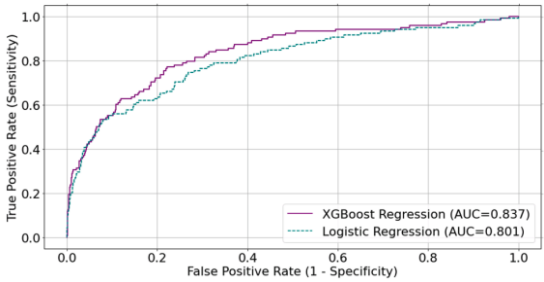

Figure 1. ROC for XGBoost vs Logistic Regression models.

3.3. Testing for Fairness

We assessed the predictive performance of the model by gender and found no substantial difference with AUC scores of 0.835 (male) vs 0.839 (female) (Figure 2). Testing for indigenous status revealed that there was no significant difference with AUC curves 0.831 (non-indigenous) vs 0.898 (indigenous), despite having fewer indigenous patients in the dataset.

Despite the reasonable overall performance of the logistic regression model we found it showed significant demographic disparity with AUC scores: 0.826 (male) vs 0.753 (female) (Figure 3) and 0.802 (non-indigenous) vs 0.754 (indigenous).

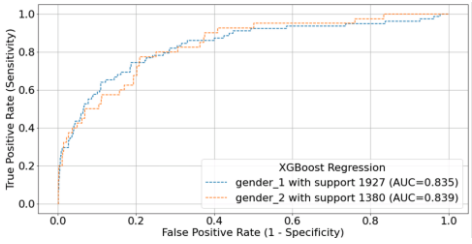

Figure 2. XGBoost ROC male vs female.

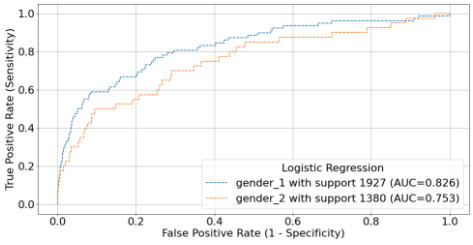

Figure 3. Logistic Regression ROC male vs female.

4. Discussion

While the ML community has created various metrics that aim to quantify bias in models [8,9], we did not extensively rely on these for 3 reasons: (1) the metrics have very little meaning to non-ML researchers and in particular clinicians so we chose to use a more visual method of ROC analysis; (2) our analysis was not reliant on the prevalence of diabetes in each population as the model is not diagnostic but trained on those already diagnosed with diabetes; (3) the outcome measure, hypoglycaemia, is an objective measure and not subject to bias of opinion or judgement.

5. Conclusions

The ultimate promise of AI and clinical decision support is to improve equity in healthcare. Monitoring and accounting for fairness in ML models is a new and active area of research [10]. We know that datasets represent the world where there are systematic biases, inequity, and underrepresented populations which we must consider when creating ML models, acknowledging the potential trade-off between accuracy and fairness [11].

We did not find any demographic disparity in the XGBoost model most likely due to the characteristics of the dataset and the additional flexibility of the XGBoost model. However, the simpler logistic regression model did show disparity in demographic performance despite an overall reasonable AUC score. It is up to both algorithm developers and users to be sensitive to the issue of fairness in AI/ML models and to continue assessing their performance.

We have found that key design choices such as targeting a population that have an established diagnosis (diabetes), using an objective outcome measure (BGL level), and incorporating objective laboratory and vital signs data assisted us in creating an equitable model and reducing the potential for bias.

Acknowledgements

This work was supported by Alcidion in partnership with Murrumbidgee Local Health District, NSW, Australia.

References

[1] Australian Commission on Safety and Quality in Healthcare. (2020, January). Hospital-acquired complications (HACs). Retrieved from Australian Commission on Safety and Quality in Healthcare: https://www.safetyandquality.gov.au/our-work/indicators/hospital-acquired-complications.

[2] Lake A, Arthur A, Byrne C, Davenport K, Yamamoto JM, Murphy HR. The effect of hypoglycaemia during hospital admission on health-related outcomes for people with diabetes: a systematic review and meta-analysis. Diabet Med. 2019 Nov;36(11):1349-59, doi: 10.1111/dme.14115.

[3] Ruan Y, Moysova Z, Tan GD, Lumb A, Davies J, Rea R. Inpatient hypoglycaemia in older people is associated with a doubling in the increased length of stay compared with the younger population. Age Ageing. 2021 Feb;50(2):576-80, doi: 10.1093/ageing/afaa212.

[4] Gray J, Roseleur J, Edney L, Karnon J. Pragmatic review of interventions to prevent inpatient chypoglycaemia. Diabet Med. 2022 Feb;39(2):e14737, doi: 10.1111/dme.14737.

[5] Kilpatrick CR, Elliott MB, Pratt E, Schafers SJ, Blackburn MC, Heard K, McGill JB, Thoelke M, Tobin GS. Prevention of inpatient hypoglycemia with a real-time informatics alert. J Hosp Med. 2014 Oct;9(10):621-6, doi: 10.1002/jhm.2221.

[6] Spanakis EK, Golden SH. Race/ethnic difference in diabetes and diabetic complications. Curr Diab Rep. 2013 Dec;13(6):814-23, doi: 10.1007/s11892-013-0421-9.

[7] Chen T, Guestrin C. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining 2016 Aug. p. 785-94, doi: 10.48550/arXiv.1603.02754.

[8] Corbett-Davies S, Pierson E, Feller A, Goel S, Huq A. Algorithmic decision making and the cost of fairness. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2017 Aug. p. 797-806, doi: 10.48550/arXiv.1701.08230.

[9] IBM Corp. AI Fairness 360 Toolkit. https://aif360.mybluemix.net.

[10] Microsoft Corp. Fairlearn. https://github.com/fairlearn/fairlearn.

[11] Pessach D, Shmueli E. A review on fairness in machine learning. ACM Comput Surv. 2022 Feb;55(3):1-44, doi: 10.1145/3494672.