**CASE STUDY: DEVELOPMENT OF A MACHINE LEARNING BASED CLINICAL DECISION SUPPORT SYSTEM TO PREDCIT CLINICAL DETERIORATION IN PATIENTS VISITING THE EMERGENCY DEPARTMENT**

**ABSTRACT**

This case study reviews the application of Machine learning (ML) in the healthcare sector with a focus on the development of a machine learning decision support system. We identified key problems in this space and explored the methodology used by the authors to address some of these problems. Finally, we provide a critical appraisal with recommendations for future work.

**INTRODUCTION**

Clinical Decision Support Systems (CDSS) represent a pivotal technological advancement in healthcare, designed to influence and enhance clinician decision-making processes concerning individual patients in real-time. The primary objective of CDSS is to assist healthcare professionals in minimizing errors and ensuring a consistent approach through the utilization of available data. (Berner, 2007). Within the medical domain, CDSS plays a crucial role in supporting prognosis and diagnosis, with the rapid integration of machine learning models further amplifying its capabilities in recognizing patterns, image classification, and extracting meaningful inferences from textual data. Particularly, in the exigent environment of Emergency Departments (ED) within hospitals, CDSS proves indispensable due to the imperative to make prompt decisions in response to the substantial influx of patients.(Choi et al., 2023)

*Key Issues*

The current state of EDs faces significant challenges, prompting the necessity for the integration of Machine Learning-enhanced Clinical Decision Support Systems (ML-CDSS) to address key issues. Challenges includes:

1. ED overcrowding, leading to a surge in patient influx, emphasizing the critical role of ML-CDSS in facilitating efficient decision-making.
2. Prolonged decision delays in emergency care, which can worsen patient conditions, underscore the importance of ML-CDSS in mitigating delays and improving the timeliness of decisions(Choi et al., 2023)
3. Data Limitations for CDSS: The effectiveness of CDSS is contingent upon the availability and quality of data.

The effectiveness of CDSS is hindered by.

* challenges arising from rapid population growth versus healthcare improvements,
* insufficient financial support for healthcare infrastructure
* concerns about the trustworthiness and interpretability of CDSS outputs(Perrott & Holland, 2005; Berner, 2007).

Overcoming these challenges is vital for realizing the full potential of ML-CDSS in EDs, ultimately leading to improved patient outcomes and enhanced healthcare efficacy.

**APPLICATION OF MACHINE LEARNING**

*Data collection*

Patient data, encompassing individuals aged 18 and above, was systematically collected from the ED of a teaching hospital over 4 years. The utilization of the Clinical Research Analysis Portal System ensured the credibility and reliability of the amassed data. The dataset comprises crucial patient attributes resulting in a comprehensive set of 120 features.

To facilitate temporal analysis, the temporal labels of the data were resampled at one-hour intervals following the patient's arrival in the ED. In instances where no observations were recorded within the initial hour for a particular patient, the preceding observation was imputed forward.

To address the inherent class imbalance within the dataset, under-sampling techniques were applied during the training phase. However, to maintain the model's applicability to real-world scenarios, the test data was deliberately left unaltered. A brief description of the process flow is shown in figure 1.

A diagram of a patient's flow

Description automatically generated

Figure . Data processing flow chart

*Model development*

XGBoost, chosen for its robust predictive capabilities and accelerated processing time compared to alternative machine learning algorithms, was employed in constructing the model. Renowned for its versatility across various applications, XGBoost is recognized for its resilience against error bias and underfitting(Chen & Guestrin, 2016). Through this algorithmic approach, ten predictors demonstrating significant associations with outcomes, alongside sixty-nine subfactors, were identified as shown in figure 2. The output was to forecasting four potential patient outcomes after their arrival in the ED: In-Hospital Cardiac Arrest (IHCA), inotropic use, intubation, and ICU admission with a time lag and lead of 1, 2, 3, and 6 hours.

A group of graphs showing different types of drugs

Description automatically generated with medium confidence

Figure . 10 best predictors for the 4 outcomes

*Personal evaluation of methods*

* Due to the under sampling used to curb the imbalance nature of the dataset, the test set is between 100 to 250 times the training size. However, the imbalance nature of field gotten data cannot be avoided because of the variability in the factors in the ED in real life scenarios. So, there is possibility the model would be bias based on this.
* Over the period of the data collation, there might be cases where patients visit the ED more than once for similar reasons of entirely different reasons. There was no mention of how this was handled in the dataset collection phase.
* The outputs measured are very limited as there are many other outcomes that could arise from the ED like the case could less serious and require the patient to be transferred to a different unit.
* The 10 best predictors were likely gotten using statistical and mathematical analysis and more research is needed to back that these predictors are best to decide if a patient outcome.
* Many of the outcomes are results gotten from the laboratory which could cause a time delay and the patients’ outcome could have varied.

**ANALYSIS AND RESULTS**

To access the model’s performance, they authors used specificity, sensitivity, precision and F1 score AUROC and AUPRC. As seen in figure 3, The model with the lagging 6 and leading 0 displayed the best results with the highest AUROC for all four outcomes. The number of false positives with a fixed sensitivity of 95%, 99% and 100 % was used to determine the accuracy of the model. For inotropic use a false positive of 11.5% with sensitivity of 95% increased to 61.4 percent when the sensitivity was 100 percent. The false positive increased from 10 to 23.9 %, 39.5 to 819% and 18.8 to 86.9 % for ventilation, cardiac arrest, and ICU admission respectively when the sensitivity was increased from 95% to 100 percent.

A graph of different colored lines

Description automatically generated

Figure . AUROC score outcome for lagging and leading.

A graph of a number of patients

Description automatically generated

Figure . False positives and sensitivity

**AUTHORS RECOMMENDATION FOR FUTURE WORK**

1. There is a need for more experiments and training using different machine learning algorithms. The sensitivity could then be compared across the models to see which performed better.
2. The retrospective design of the study can present potential bias. Especially in the data collation. This could be improved by using real time patient information acquisition systems.
3. Management of patients in the ED was not represented in the model. The model needs to be expanded to include efforts by physicians in the ED.

*Personal Appraisal and critic*

From figure 4 we can observe that a little change in the sensitivity increases the number of false positives exponentially. Approaches recommended by (Forstmeier et al., 2017) could be explored to limit the effects of false positives. All humans are different and different parameters can lead to vastly different results, hence the use of generic parameters for all future cases could be controversial and further leads to more false positives which wasn’t accounted for in the model creation. Models should be selected on the ability to generalize e.g., neural networks due to the limitations in the dataset.

**BIBLIOGRAPHY**

Berner, E. S. (2007) *Clinical decision support systems*, *233*.Springer.

Chen, T. & Guestrin, C. (2016) Xgboost: A scalable tree boosting system, *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*.

Choi, A., Choi, S. Y., Chung, K., Chung, H. S., Song, T., Choi, B. & Kim, J. H. (2023) Development of a machine learning-based clinical decision support system to predict clinical deterioration in patients visiting the emergency department. *Scientific Reports*, 13(1), 8561.

Forstmeier, W., Wagenmakers, E. J. & Parker, T. H. (2017) Detecting and avoiding likely false‐positive findings–a practical guide. *Biological Reviews*, 92(4), 1941-1968.

Perrott, G. S. J. & Holland, D. F. (2005) Population trends and problems of public health. *The Milbank Quarterly*, 83(4), 569.