2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Unsupervised Machine Learning Models for Characterization of Risk for Pediatric Several Critical Events from Anesthesia Using the Wake-Up Safe Registry

Geoffrey M. Gray, Sharon Crabtree, Allison Fernandez, Anna Varughese, Luis M. Ahumada

***Abstract*— characterization and risk prediction of adverse events from anesthesia is an important challenge. This study uses unsupervised machine learning with the wake up safe database to create risk categories to characterize patients at risk of experiencing critical events.**

***Clinical Relevance*— This study characterizes models risk for severe critical events from anesthesia using the wake up safe database.**

I. INTRODUCTION

Characterization of severe critical events remains an ongoing challenge due to the relative infrequency of the events resulting in a paucity of data. The Wake-Up Safe (WUS) registry represents one effort to overcome this impediment by collecting multicenter data. However, the registry suffers from deficiencies preventing full utilization of the data blunting the scope of application. This study applies unsupervised machine learning methods to create models characterizing risks for several critical events from anesthesia to overcome these issues and provide useful tools for clinical decision making and risk stratification. These tools help facilitate better use of the WUS registry in a clinical setting.

II. METHODS

Two datasets are made available from WUS, the billing dataset and the events dataset. The billing dataset provides billing information on anesthesia procedures and contains nearly 4 000 000 entries. Despite the large collection of data, the features collected are relatively sparse (Age, ASA Score, ICD, CPT, ASA Emergency Status). Contrastingly, the events dataset contains detailed information on each critical event. Unfortunately, the absence of a direct link between the events dataset and the billing dataset, as well as the substantial discrepancy in the features contained in each dataset, prohibit directly combining them into a single dataset for predictive modeling. To overcome this issue, we devise a novel strategy to indirectly link the detailed critical events in the registry to the large dataset in the billing file. The approach is as follows: first the billing dataset is clustered using the kprototypes algorithm.[1] A fuzzy record matching algorithm is then applied to match the entries from the events dataset to clusters in the billing dataset. Then, the probability of different events is calculated for each cluster. This can be used to predict patient risk given diagnosis, procedure and demographics.

III. RESULTS

Cluster cost by total number of clusters is shown in Fig. 1. The elbow of the curve occurs at 4 clusters, indicating that this is the optimal number of clusters. Comparison of the distances between the centroids of the cluster (data not shown) supported this, as well.

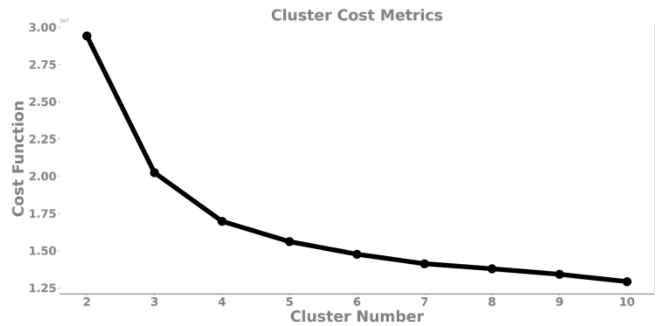

Figure 1. Cluster cost (x1e7) metrics.

TABLE I. Percentage of total PAEs by cluster.

Cluster	Percentage of PAE
1	24.16%
2	28.32%
3	22.20%
4	25.32%

Table I shows the percentage of total PAEs matched to each cluster. Cluster 2 shows the highest number of adverse events, while cluster 3 has the smallest number.

IV. DISCUSSION & CONCLUSION

This work demonstrates how to apply machine learning to data that is not collected for this application. This application has led to a model of risk for critical events in anesthesia.

REFERENCES

1. Huang ZX. Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery.* 1998;2(3):283-304.

*Research supported by Johns Hopkins All Children's Hospital.

Luis M. Ahumada is with the *Center for Pediatric Data Science and Analytics Methodology, Johns Hopkins All Children's Hospital, St. Petersburg, FL* 33701 USA, phone: 727-767-2349; e-mail: ahumada@jhmi.edu.