Predicting adverse long-term neurocognitive outcomes after pediatric intensive care unit hospitalization

Felipe Kenji Nakano[1,2], Karolijn Dulfer[3], Ilse Vanhorebeek[4], Pieter J. Wouters[4], Sascha C. Verbruggen[3], Koen F. Joosten[3], Fabian Güiza Grandas [⋆4], Celine Vens[1,2], and Greet Van den Berghe[4]

[1] KU Leuven, Campus KULAK, Department of Public Health and Primary Care
[2] Itec, imec research group at KU Leuven
{felipekenji.nakano, celine.vens}@kuleuven.be
[3] Intensive Care Unit, Department of Paediatrics and Paediatric Surgery, Erasmus Medical Centre, Sophia Children's Hospital
{k.dulfer, s.verbruggen, k.joosten}@erasmusmc.nl
[4] Clinical Division and Laboratory of Intensive Care Medicine, Department of Cellular and Molecular Medicine, KU Leuven
{ilse.vanhorebeek, pieter.wouters, fabian.guiza, greet.vandenberghe}@kuleuven.be

Abstract. Critically ill children often suffer from impaired neurocognitive functions years after ICU (intensive care unit) discharge. To assess neurocognitive functions, children are subjected to a fixed sequence of tests. Undergoing all tests is, however, arduous, resulting in a possibly interrupted evaluation where several neurocognitive deficiencies might remain undetected. We propose a machine learning approach to predict the optimal order of tests for each child. More specifically, our approach groups the tests according to the neurocognitive functions and builds a multi-target random forest for each group. Further, we combine the outputs into a ranking that prioritizes the worse neurocognitive outcomes of each patient. Our experiments demonstrate that machine learning can be competitive or even superior to the current clinical practice.

Keywords: Neurocognitive outcomes · Pediatric ICU · Label ranking

1 Introduction and Method

Critically ill children often suffer from impaired neurocognitive functions years after PICU (paediatric intensive care unit) discharge [5]. In order to assess these impairments, children are subjected to a standard evaluation procedure which adopts a fixed sequence of tests which are grouped into neurocognitive domains, such as memory and executive functions, among others [5]. This sequence, however, does not consider the specific characteristics of each child. Furthermore,

⋆ These authors have equally contributed to this manuscript.

undergoing all tests is arduous, time-consuming and expensive, leading to interrupted evaluations and undetected neurocognitive deficiencies.

We propose to address this problem using machine learning. More specifically, our objective consists of predicting a personalized sequence of tests which prioritizes tests associated to neurocognitive functions (outcomes) that are expected to be affected. We propose a new label ranking method [4], namely Label Ranking per Group of Outcomes (LaRGO), that builds a multi-output random forest per group of outcomes and combines the output of each model into a ranking where worse outcomes are placed in higher positions[5].

2 Experiments

We employ data obtained in a 2 years follow-up from the PEPaNIC-RCT study, a multicenter, randomized and controlled trial [5]. It contains 23 features, related to demographics, socioeconomic status and PICU related features, and 23 neurocognitive outcomes measured from 786 previously hospitalized patients.

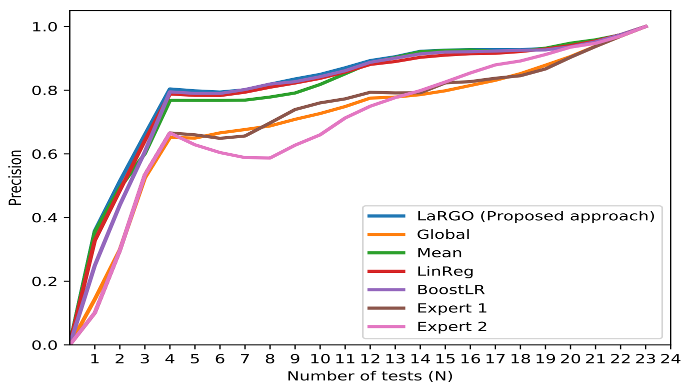

Fig. 1: Precision@N using the 10 × 5-fold cross validation for different values of N obtained using all comparison approaches.

As seen in Figure 1, all machine learning approaches yield better results, where LaRGO is slightly superior. When compared to the current clinical practice (Experts), a more visible difference is perceived, as LaRGO achieved approximately 80% of precision, whereas the experts fluctuate around 60%.

3 Conclusion

We have proposed a label ranking approach whose results surpass the current clinical practice, thus assisting physicians and other care providers in identifying a personalized sequence of tests that children should undergo to identify their worst neurocognitive deficiencies. In future work, we aim to incorporate data from the 4 years follow-up study [2], and also develop a new label ranking heuristic for the random forests, such as [1].

[5] Our complete work is available at [3].

References

1. Ziniu Hu, Yang Wang, Qu Peng, and Hang Li. Unbiased lambdamart: an unbiased pairwise learning-to-rank algorithm. In *The World Wide Web Conference*, pages 2830–2836, 2019.
2. An Jacobs, Karolijn Dulfer, Renate D Eveleens, José Hordijk, Hanna Van Cleemput, Ines Verlinden, Pieter J Wouters, Liese Mebis, Gonzalo Garcia Guerra, Koen Joosten, et al. Long-term developmental effect of withholding parenteral nutrition in paediatric intensive care units: a 4-year follow-up of the pepanic randomised controlled trial. *The Lancet Child & Adolescent Health*, 4(7):503–514, 2020.
3. Felipe Kenji Nakano, Karolijn Dulfer, Ilse Vanhorebeek, Pieter J Wouters, Sascha C Verbruggen, Koen F Joosten, Fabian Güiza Grandas, Celine Vens, and Greet Van den Berghe. Predicting adverse long-term neurocognitive outcomes after pediatric intensive care unit admission. *Computer Methods and Programs in Biomedicine*, 250:108166, 2024.
4. Shankar Vembu and Thomas Gärtner. Label ranking algorithms: A survey. In *Preference learning*, pages 45–64. Springer, 2010.
5. Sören Verstraete, Sascha C Verbruggen, José A Hordijk, Ilse Vanhorebeek, Karolijn Dulfer, Fabian Güiza, Esther van Puffelen, An Jacobs, Sandra Leys, Astrid Durt, et al. Long-term developmental effects of withholding parenteral nutrition for 1 week in the paediatric intensive care unit: a 2-year follow-up of the pepanic international, randomised, controlled trial. *The Lancet Respiratory Medicine*, 7(2):141–153, 2019.