Semifinal Results of a Research Project Involving Algorithmic Complexity Estimation and Machine Learning

Victor Iapascurta [1,2]
[1] Technical University of Moldova, victor.iapascurta@doctorat.utm.md,
ORCID: 0000-0002-4540-7045 www.utm
[2] N.Testemitanu University of Medicine and Pharmacy, www.usmf.md

Keywords: Kolmogorov-Chaitin complexity, machine learning, sepsis, epilepsy

Abstract. In recent years, the intersection of algorithmic complexity and machine learning has opened new avenues for analyzing continuous biomedical data. This paper presents the semifinal results of a research project focused on estimating Kolmogorov-Chaitin Complexity (KCC) using the Block Decomposition Method. KCC serves as a core feature for machine learning models aimed at predicting sepsis and epileptic seizures. The results highlight the efficacy of these models, with promising performance metrics, and underscore the utility of algorithmic complexity measures in enhancing machine learning models for biomedical applications. By leveraging the inherent complexity in biomedical signals, these models achieve superior predictive performance.

Methodology. Kolmogorov-Chaitin Complexity (KCC) [1] is a measure of the randomness or information content of a data sequence. Estimating KCC for continuous biomedical signals involves the Block Decomposition Method, which breaks down data into manageable blocks to approximate complexity.

The KCC values derived from biomedical data are used as primary features in various machine learning models. The objective is to predict medical conditions—specifically, sepsis with a 4-hour horizon and epileptic seizures.

Data and Experimentation. Real-world biomedical data were sourced from international competitions, including time-series data relevant to sepsis and epilepsy. These datasets provided a robust foundation for training and testing machine learning models.

Several models, including neural networks, gradient boosting machines, generalized linear models, and others, were trained using the KCC features.

Results

Table 1. The machine learning models with the highest performance [2].

Data set	ML model	Performance by AUC
Epileptic EEG set	Word2Vec	96.8%
Sepsis set	Gradient Boosting Machine	95.3%

Note: AUC – area under the ROC curve

Discussion. The results underscore the utility of algorithmic complexity measures in enhancing machine learning models for biomedical applications. By leveraging the inherent complexity in biomedical signals, these models achieve superior predictive performance.

Conclusion. This research project demonstrates the effectiveness of combining algorithmic complexity estimation with machine learning to predict sepsis and epileptic seizures. With AUC scores of 95.3% and 96.8%, respectively, the models show significant promise for real-world medical applications. Further refinement and validation could enhance their utility in clinical practice.

References

[1] H. Zenil, A review of methods for estimating algorithmic complexity: options, challenges, and new directions. Entropy 22(6), (2020).

[2] V. Iapascurta, V. Estimation of Kolmogorov-Chaitin Complexity on Continuous Biomedical Data for Machine Learning Purposes. In: Intelligent and Fuzzy Systems. INFUS 2024. Lecture Notes in Networks and Systems, vol 1090. Springer, Cham. pp. 60-67, 2024.