Learning a Patient Representation using Tabular Transformers

This master's thesis project focuses on leveraging the power of Tabular Transformers to learn comprehensive patient representations from electronic health records (EHR) and other healthcare-related tabular data. The central research question is: How can Tabular Transformers be adapted and employed to capture complex patient profiles, including their medical history, demographics, and treatment outcomes, in a manner that is both interpretable and efficient? This research seeks to address the critical need for more effective patient representations in healthcare, with potential applications in diagnosis, treatment planning, and medical research.

The project entails the following key objectives:

1. Tabular Transformer Adaptation: Customizing the architecture and training procedures of Tabular Transformers to handle healthcare tabular data efficiently.
2. Feature Engineering: Exploring feature engineering techniques to preprocess EHR and related tabular data, including handling missing values, encoding categorical variables, and scaling numerical attributes.
3. Representation Learning: Developing methods to learn informative and interpretable patient representations from the adapted Tabular Transformer, capturing both individual and population-level insights.
4. Evaluation and Validation: Evaluating the learned patient representations through tasks such as patient similarity analysis, prediction of medical outcomes, and visualization for interpretability.
5. Ethical Considerations: Addressing ethical considerations regarding data privacy, bias, and fairness in healthcare representation learning.

Expected outcomes include a novel approach to patient representation learning, insights into the potential benefits of Tabular Transformers in healthcare analytics, and advancements in the field's ability to harness EHR data for improved patient care and medical research.

Contact: **Fabian Schmidt** <fschm@kth.se>
Distributed Computing Group, SCS Division, KTH
Examiner: **Prof. Vladimir Vlassov** <vladv@kth.se>
Distributed Computing Group, SCS Division, KTH

Technical Requirements:

- Strong background in machine learning, deep learning, and healthcare informatics.
- Proficiency in programming languages like Python and experience with deep learning frameworks (e.g., PyTorch, TensorFlow).
- Familiarity with EHR data structures and healthcare data standards.
- Knowledge of transformer architectures and tabular data processing.
- Experience with data preprocessing techniques, feature engineering, and model evaluation.
- Strong analytical and problem-solving skills.
- Ethical considerations in healthcare data usage.

References:

- 1. Arik, Sercan Ö., and Tomas Pfister. "**Tabnet: Attentive interpretable tabular learning.**" Proceedings of the AAAI conference on artificial intelligence. Vol. 35. No. 8. 2021.
- 2. Hollmann, Noah, et al. "**Tabpfn: A transformer that solves small tabular classification problems in a second.**" arXiv preprint arXiv:2207.01848 (2022).
- 3. Carballo, Kimberly Villalobos, et al. "**TabText: A Flexible and Contextual Approach to Tabular Data Representation.**" arXiv preprint arXiv:2206.10381 (2022).