MCRAGE: Synthetic Healthcare Data for Fairness

Keira Behal *, Jiayi Chen *, Caleb Fikes †, and Sophia Xiao *

Project advisor: Yuanzhe Xi

Abstract. In the field of healthcare, electronic health records (EHR) serve as crucial training data for developing machine learning models for diagnosis, treatment, and the management of healthcare resources. However, medical datasets are often imbalanced in terms of sensitive attributes such as race/ethnicity, gender, and age. Machine learning models trained on class-imbalanced EHR datasets perform significantly worse in deployment for individuals of the minority classes compared to those from majority classes, which may lead to inequitable healthcare outcomes for minority groups. To address this challenge, we propose Minority Class Rebalancing through Augmentation by Generative modeling (MCRAGE), a novel approach to augment imbalanced datasets using samples generated by a deep generative model. The MCRAGE process involves training a Conditional Denoising Diffusion Probabilistic Model (CDDPM) capable of generating high-quality synthetic EHR samples from underrepresented classes. We use this synthetic data to augment the existing imbalanced dataset, resulting in a more balanced distribution across all classes, which can be used to train less biased downstream models. We measure the performance of MCRAGE versus alternative approaches using Accuracy, F1 score and AUROC of these downstream models. We provide theoretical justification for our method in terms of recent convergence results for DDPMs.

Key words. synthetic electronic health records, conditional denoising diffusion probabilistic model, healthcare AI, tabular data, fairness, synthetic data

1. Introduction.

In recent years, reliance on machine learning algorithms to facilitate decision-making processes across various industries has grown. In healthcare, clinicians may use machine learning models to predict disease progression, improve diagnosis accuracy, and optimize treatment plans [25]. However, machine learning approaches may perpetuate existing societal biases, leading to inequitable treatment for minority groups, because machine learning models trained on imbalanced datasets may replicate and thus amplify these biases [5].

These issues are of utmost concern in healthcare applications where fair and equitable treatment is of critical importance. Ideally, a well-engineered machine learning model should be fair, optimizing health outcomes to provide high-quality, individualized care to all patients, regardless of their demographic characteristics [23]. Unfortunately, healthcare datasets are often imbalanced across several dimensions, including race, socioeconomic status, age, and gender [15, 8]. As a result, models trained on these datasets struggle to generalize effectively to individuals who are not well represented in the data [28].

EHRs are a valuable data source in healthcare, providing a comprehensive snapshot of a patient's health history, including diagnoses, treatments, and demographic information [26]. Certain demographic groups, such as specific racial or ethnic minorities, are often underrepresented in the EHR datasets [33]. This imbalance might lead to inequitable health outcomes,

*Department of Mathematics, Emory University, Atlanta, GA 30322 (keira.behal@emory.edu,jiayi.chen@emory.edu,sxxiao@emory.edu)

†Department of Mathematics, Rice University, Houston, TX 77005 & Department of Mathematics, Emory University, Atlanta, GA 30322 (cfikes@emory.edu)