UREAP APPLICATION FORM

THOMPSON RIVERS UNIVERSITY
Office of Research and Graduate Studies

First Name: Deepansh

Last Name: Sharma

Student ID: T00701291 Start Date of Project: 05/01/2025 (DD/MMM/YYYY)

Please complete all sections of this application form.

1. FACULTY MENTORS INFORMATION

1.1 Who is your Primary Faculty Mentor? Dr. Anthony Aighobahi

1.2 Who is your Secondary Faculty Mentor? Dr. Nisha Puthiyedth

NOTE: Your Primary and Secondary Faculty Mentors must each complete a Faculty Mentor Support Form. Forms can be found under the attachments tab within your TRU Romeo UREAP application and on the TRU UREAP webpage under information and Forms for Faculty Mentors..

2. PROJECT DESCRIPTION

2.1 Provide an abstract of your proposed research: (maximum 1500 characters)

Machine learning (ML) models can enhance healthcare by predicting in-hospital mortality, refining diagnostics, and informing treatment plans. However, these models often inherit biases from their training data, leading to inequitable predictions across race, gender, and socioeconomic groups. This research aims to identify and mitigate biases in ML models developed using two key North American healthcare datasets: MIMIC-III, MIMIC-IV (Medical Information Mart for Intensive Care III) and HCUP (Healthcare Cost Utilizaton Project). Our objective is to forecast in-hospital mortality while evaluating how biases affect predictive performance across demographic groups. We will preprocess the data, apply bias detection techniques, and implement mitigation strategies such as adversarial debiasing, re-weighting, and data augmentation. By integrating intersectional frameworks and counterfactual analysis, this research seeks to enhance fairness in healthcare ML models, ensuring equitable and reliable decision-making.

2.2 Provide a brief literature review for your proposed research: (maximum 3500 characters)

Machine learning (ML) models play a critical role in healthcare by predicting outcomes such as in-hospital mortality, aiding clinical decision-making. However, biases within datasets like MIMIC-III, MIMIC-IV and HCUP can result in disparities in model performance, disproportionately affecting certain patient groups (Luo et al., 2020). This review examines racial, gender, and socioeconomic biases, their impact on model accuracy, and strategies to mitigate them.

Racial Bias:

Racial bias in healthcare ML models arises from historically skewed data. Luo et al. (2020) found that predictive models underestimated mortality risk for Black patients due to reliance on outdated scoring methods. Moreover, pulse oximeters, which influence ML input data, have been shown to yield less accurate readings for individuals with darker skin tones, introducing further errors (Martins et al., 2024). These biases contribute to inequitable healthcare outcomes and necessitate targeted interventions to improve fairness.

Gender Bias:

ML models trained on patient records often reflect gender disparities in medical documentation. Minot et al. (2021) identified that gendered language in clinical notes can lead to skewed predictions, as models misinterpret context-specific terms. Likewise, Robinson et al. (2021) demonstrated that medical language models exhibit gender biases, though their impact on mortality prediction remains underexplored. Such biases can lead to inconsistent treatment recommendations, compromising patient care.

Socioeconomic Bias:

Socioeconomic status, including income and insurance type, significantly influences data quality in electronic health records. Röösli et al. (2022) found that ML models trained on MIMIC-III underperformed for patients with public insurance, who are often from lower-income backgrounds. Similarly, Meng et al. (2022) identified disparities in MIMIC-IV, highlighting systemic gaps in healthcare access and quality. These biases undermine model reliability for economically disadvantaged populations, reinforcing healthcare inequities.

Consequences of Bias:

Biases in ML models can lead to inaccurate predictions, resulting in inadequate or incorrect treatment recommendations. For example, Black and low-income patients may receive suboptimal care due to underestimations of their medical risks (Luo et al., 2020). Gender biases can further distort predictions, reducing trust in AI-driven decisions (Minot et al., 2021). Addressing these biases is critical to ensuring equitable and effective healthcare delivery (Meng et al., 2022).

Mitigation Strategies: Several approaches have been proposed to reduce bias in healthcare ML models,

1. Data Balancing: Increasing the representation of underrepresented groups in training datasets can enhance fairness, though it may slightly reduce overall predictive accuracy (Chaudhry et al., 2023).

2. Fairness Training: Adjusting model learning processes to prioritize fairness can mitigate bias without significantly compromising performance (Sivarajkumar et al., 2023).

3. Bias Removal: Algorithmic techniques that eliminate unfair patterns from datasets can enhance equity but require careful tuning to avoid unintended consequences (Chaudhry et al., 2023).

4. Equity Constraints: Implementing fairness constraints during model training ensures balanced outcomes, though this may marginally impact predictive strength (Shi et al., 2024).

2.3 What is the hypothesis or research question for your proposed research? Include any specific objectives: (maximum 500 characters)

1. What are the prevalent biases related to race, gender, and socioeconomic status in the MIMIC-III and HCUP datasets concerning in-hospital mortality prediction?

2. How do these biases affect the accuracy and fairness of machine learning models trained to predict in-hospital mortality?

3. What are effective strategies to mitigate these biases in ML models, and how do they impact the model's performance and fairness?

2.4 Provide a description of the research methodology/methodologies and analysis that you intend to employ in completing this research: (maximum 1500 characters)

This research employs a structured approach to detect and mitigate bias in MIMIC-III and HCUP datasets using fairness metrics, counterfactual analysis, and intersectional frameworks.

We begin with data preprocessing, handling missing values through imputation and normalizing numerical variables. Categorical attributes like race and gender are encoded for consistency in ML models. Bias detection is conducted using fairness metrics such as demographic parity (ensuring equal prediction rates), equal opportunity (measuring the likelihood of correct predictions across groups), and disparate impact (comparing disparities between protected and non-protected groups). Counterfactual analysis helps determine whether altering demographic attributes affects predictions, while intersectional analysis examines compounded biases across multiple demographic factors.

ML models, including Logistic Regression, Random Forests, Neural Networks, and XGBoost, are trained to predict patient outcomes. Data is split into training, validation, and test sets to ensure unbiased evaluation. Hyperparameter tuning is performed to optimize both fairness and predictive performance.

Bias mitigation strategies include re-sampling underrepresented groups, applying weighted loss functions, and adjusting predictions post-processing to align with fairness metrics. Model performance is evaluated using precision, recall, and AUC-ROC, while fairness metrics are reassessed post-mitigation to validate improvements.

2.5 Provide a description of how your research will significantly impact your field of study:

(maximum 1500 characters)

This research advances both computer science and healthcare by optimizing bias detection and mitigation in ML models. Bias in healthcare AI can lead to disparities in diagnosis, treatment planning, and patient outcomes. By refining fairness-aware ML techniques, this study ensures equitable healthcare applications while contributing to broader advancements in algorithmic fairness, responsible AI deployment, and ethical machine learning.

From a computer science perspective, this research enhances ML model evaluation by integrating bias detection at multiple levels—data preprocessing, model training, and post-processing. It extends bias mitigation methodologies using fairness-aware optimization techniques applicable beyond healthcare, influencing AI development in finance, hiring, and public policy.

Additionally, this work contributes to software engineering best practices by integrating fairness audits into ML pipelines, ensuring bias-aware model deployment. The findings will inform regulatory frameworks for AI ethics and governance, promoting trust in AI-driven decision-making. By validating bias mitigation strategies in real-world datasets, this research bridges the gap between theoretical AI fairness models and their practical implementation in critical applications.

2.6 Describe your plans to disseminate your research findings: (maximum 500 characters)

Findings will be published in peer-reviewed journals like Journal of Medical AI and PLOS ONE. Presentations will be made at conferences such as TRU's Undergraduate Resesarch and Innovation Conference, BC Digital Health and ICAIMLH. Methodology and datasets will be shared via open-access repositories. Collaborations with local BC healthcare organizations will help implement bias-mitigated models in real-world clinical settings, ensuring practical impact and adoption.

2.7 List the references that you have citied throughout your research proposal observing the appropriate citation style for your discipline: (maximum 3500 characters)

1. Luo, Y., Wunderink, R. G., & Lloyd-Jones, D. (2020). A racially unbiased, machine learning approach to prediction of mortality. JMIR Public Health and Surveillance, 6(4), e23570. https://doi.org/10.2196/23570

2. Martins, I., Shen, S., Zemskova, M., & Arnold, T. (2024). Evaluating the impact of pulse oximetry bias in machine learning. arXiv preprint arXiv:2408.04396. https://arxiv.org/abs/2408.04396

3. Minot, J. R., Cheney, N., Maier, M., et al. (2021). Interpretable bias mitigation for textual data. arXiv preprint arXiv:2103.05841. https://arxiv.org/abs/2103.05841

4. Robinson, R., Suresh, V., Williams, M. Y., & Ghassemi, M. (2021). Assessing gender bias in medical language models. arXiv preprint arXiv:2111.08088. https://arxiv.org/abs/2111.08088

5. Röösli, B., Oakden-Rayner, L., De Fauw, J., & Ronneberger, O. (2022). Fairness and generalizability of a MIMIC-III benchmarking model. Scientific Data, 9(1), 24. https://doi.org/10.1038/s41597-021-01110-7

6. Meng, C., Trinh, L., Xu, N., et al. (2022). Interpretability and fairness evaluation of deep learning models. Scientific Reports, 12(1), 7165. https://doi.org/10.1038/s41598-022-11012-2

7. Sivarajkumar, S., Wang, Y., & Ko, H. (2023). Fair patient model: Mitigating bias in EHRs. Journal of Biomedical Informatics, 146, 104488. https://doi.org/10.1016/j.jbi.2023.104488

8. Chaudhry, F., Klein, E., & Tomson, P. (2023). Survey on machine learning biases and mitigation. Digital, 4(1), 1. https://doi.org/10.3390/digital4010001

3. PROJECT TIMELINE WITH BENCHMARKS

3.1 Provide a timeline for your project that includes key benchmarks: (maximum 1000 characters)

May 1 – May 14: Finalize literature review, complete dataset preprocessing, and refine research objectives.

May 15 – May 28: Conduct EDA on MIMIC-III and HCUP datasets, assess biases using fairness metrics, and document findings.

May 29 – June 11: Implement bias detection techniques, including demographic parity, counterfactual analysis, and intersectional evaluation.

June 12 – June 25: Develop and test ML models (Logistic Regression, Random Forests, Neural Networks, XGBoost) for mortality prediction.

June 26 – July 9: Apply bias mitigation strategies like re-weighting, adversarial debiasing, and data augmentation.

July 10 – July 23: Validate mitigation strategies, compare pre- and post-mitigation results, and assess fairness and accuracy.

July 24 – August 7: Fine-tune models, conduct interpretability analysis with SHAP, and finalize results.

August 8 – August 14: Complete final report, prepare presentations, and disseminate findings.

NOTE: Please refer to the UREAP Help Guide for a project timeline example. Students must demonstrate a willingness to engage in 12 weeks or equivalent of sustained research per the Terms of Reference.

4. OPERATING GRANT BUDGET PROPOSAL

4.1 The UREAP award offers up to $1000 toward direct research expenses. These expenses must be preapproved by the UREAP committee in the adjudication phase. Use the provided template under the Attachments tab in the TRU Romeo UREAP application to complete your budget proposal. Copy amount from the TOTAL AMOUNT line of the budget here. Total Amount: $ 1,000.00

4.2 Additional budget information: (maximum 500 characters)

The amount includes the conference registration fee and various tech subscriptions required to synthesize the research project.

5. CONTRIBUTION TO ACADEMIC/PROFESSIONAL GOALS

5.1 Describe how this project will contribute to your academic and/or professional goals:

(maximum 1000 characters)

This research serves a vital step in my academic and professional trajectory, augmenting both my present studies and future aspirations. As a computer science student with a fervour for AI and machine learning, I am afforded the opportunity to investigate a critical real-world application of AI: the discovery and mitigation of bias in healthcare. I have actively influenced the project's trajectory, curated datasets, and pinpointed critical obstacles from idea to execution.

This project enhances my proficiency in machine learning, setting me apart from colleagues by concentrating on practical datasets such as MIMIC-III and HCUP. It underscores the development of models that identify and alleviate bias while ensuring robust clinical performance, hence enhancing healthcare outcomes.

This experience corresponds with my objective of obtaining a master's degree in AI and machine learning, showcasing my capacity to recognize deficiencies, formulate solutions, and implement them effectively

NOTE: Include your role in conceiving of the project, your role in the implementation of the project, and your overall academic objectives – explaining how this project will help to advance those objectives.