

# Advancing Rare Disease Research by Unifying Both Supervised and Unsupervised Methods

Name: YAO Lean Franzl Lim

Laboratory's name: Social Computing Laboratory

Supervisor's name: Professor Eiji Aramaki

## Abstract

Modern natural language processing (NLP) tools revolutionize language understanding with improved semantics and context awareness. However, their adoption raises concerns for non-experts. Traditional machine learning maintains interpretability. In rare disease research, accessibility is crucial. Researchers must creatively select models.

In the field of rare disease research, equipping researchers with knowledge of available tools and their practical applications is essential for promoting inclusivity and technological accessibility. Researchers now face the challenge of selecting suitable models and approaches for diverse contexts while harnessing their creativity to integrate different methodologies for meaningful outcomes.

This dissertation clarifies effective approaches, emphasizing traditional machine learning's relevance. It contributes to rare disease research with two studies:

1. The development of a supervised NLP-based tool for quantifying and monitoring the trends of medical research prioritization. This tool can help relevant stakeholders realign their resources and attention to important disease categories.
2. Exploration of COVID-19's impact on rare disease communities via social media data and models. This highlights social media as a source of insights.

This dissertation explores NLP and traditional machine learning in rare disease research, offering guidance for advancing medical knowledge and patient well-being.