Name: Tumusiime Andrew Gahwera

RegNo: 2019/HD05/25307

Dissertation Title: **Machine Learning Models for Short-Term Rainfall Prediction Using Uganda's Lake Victoria Basin Weather Dataset**

Abstract:

As climate change intensifies, accurate short-term rainfall forecasting has become an urgent research priority. Numerical weather prediction models often struggle with precipitation due to high computational demands and large error margins. This dissertation addresses these challenges by introducing a curated multi-station dataset for the Lake Victoria Basin (LVB) and systematically evaluating machine learning and deep learning approaches for rainfall forecasting in a data-scarce African context.

Regression experiments benchmarked random forest, support vector regression, neural networks, least absolute shrinkage and selection operator, gradient boosting, and extreme gradient boosting (XGBoost). Among these, XGBoost consistently achieved the lowest error, with mean absolute error (MAE) values as low as 0.006, 0.018, and 0.005 $\mathrm{mm}^{\mathrm{h}^{-1}}$ for Uganda, Kenya, and Tanzania, respectively. To complement these continuous forecasts, rainfall classification was implemented using multi-layer perceptrons, convolutional neural networks, and long short-term memory (LSTM) networks. LSTM outperformed alternative architectures, achieving weighted F1-scores above 90% at multiple stations.

To overcome data scarcity, a transfer learning strategy was developed by fine-tuning pre-trained LSTMs from data-rich stations and applying them to the data-limited station of Kisumu, yielding performance improvements of up to 3%. An ensemble of these transfer-learned models using exponential weighting further enhanced robustness, delivering up to 5% gains in F1. Overall, the dissertation demonstrates that an ensemble-based transfer learning framework, grounded in a regionally curated dataset, can substantially improve rainfall forecasting in East Africa. The integration of regression, deep learning classification, and transfer learning offers both methodological advancement and operational potential, contributing to more reliable weather services in data-scarce regions such as the LVB.