# **Parallel & Distributed Computing Project**

***Enhancing XGBoost Efficiency: A Comparative Study of Parallel and Non-Parallel Processing***

## Task 1

### Group Members

1. Haseeb Ahmed , ERP: 26077
2. Abdullah Iqbal , ERP: 26904
3. Azizullah Khan , ERP: 24931

Code

The attached code uses XGBoost model, focused on predicting the presence of a medical condition from a given set of features. The model is designed to maximize the AUC-ROC for the positive class. It is trained on patient data with hyperparameters tuned to enhance predictive performance. The dataset is taken from a Challenge in Kaggle.

## Problem Statement

However, as observed in the code, XGBoost takes approximately 5 minutes for evaluation, which is a significant concern given the increasing number of patients. With a growing patient load, this prolonged processing time adds up, making it crucial to optimize computation without compromising accuracy. XGBoost is renowned for its predictive accuracy, it often faces limitations in processing speed. In XGBoost, computation time typically follows a **non-linear** pattern as data size increases making parallel computing crucial for efficiency. By introducing parallelism into the data processing pipeline, this project seeks to enhance efficiency, thereby reducing both computation time and resource utilization. This project endeavors to address these speed constraints by implementing parallel computing techniques, optimizing XGBoost's performance to better handle the increasing volume of healthcare data.

[Base Paper Link](http://ijcs.net/ijcs/index.php/ijcs/article/view/3798/482)

[Google Drive](https://drive.google.com/drive/folders/1jxaXPUdVEANpEkE041rnchiwASOaSUfR?usp=sharing)