# 📊 Machine Learning Internship - ImagoAI

# 🚀 Short Report

## 1. Preprocessing Steps & Rationale

- **Missing Values:** Filled missing values with the **median** to ensure robustness.
- **Feature Scaling:** Used **MinMaxScaler** to normalize spectral data between 0 and 1 for stable training.
- **Dimensionality Reduction:** Applied **PCA (n_components=50)** to retain **95% variance** and reduce noise.

## 2. Insights from Dimensionality Reduction (PCA)

- PCA reduced **original feature dimensions** from **450** to **50** components.
- Retained **95% of variance**, improving computational efficiency while keeping essential information.
- **Scatter plot analysis** showed clear patterns between PCA components and DON concentration.

## 3. Model Selection, Training & Evaluation

- **Selected Model:** Convolutional Neural Network (**CNN-Conv1D**) for spectral data feature extraction.
- **Loss Function: Mean Squared Error (MSE)** for stable regression performance.
- **Hyperparameters Optimized:** Filters (**128, 64, 32**), Kernel Size (**5, 3, 3**), Dropout (**0.3, 0.2**).
- **Training Setup:** 80% training, 20% testing, **batch size = 16, epochs = 80**.

## 4. Key Findings & Suggestions for Improvement

✅ **Performance Metrics:**

- **Mean Absolute Error (MAE): 0.0338**
- **Root Mean Squared Error (RMSE): 0.0798**
- **$R^2$ Score: 0.6092**

🔹 **Limitations & Improvements**

- **PCA Feature Loss:** Some spectral details might be lost; **alternative methods like Autoencoders** could be explored.
- **Alternative Models: LSTM** or **Transformer-based models** could capture sequential dependencies better.

- **More Data Needed:** The dataset size is limited, **data augmentation or more samples** could improve generalization.
- **Hyperparameter Tuning:** Using **Grid Search or Bayesian Optimization** could fine-tune the CNN architecture.