Water Level Identification Using LiDAR and Machine Learning

Summary Report

This project evaluated LiDAR and machine learning (ML) for predicting water levels in a simulated river environment under varying turbidity (water clarity). The goal was to determine if non-contact sensors like LiDAR could accurately measure water levels and if ML models could reliably predict those levels from sensor data.

Methods & Data

Sensor readings were collected under low, medium, and high turbidity using:

* LiDAR (distance & signal strength)
* Ultrasonic sensor (distance)
* Accelerometer/Gyroscope (motion/orientation)

These were paired with known water levels to create a supervised learning dataset. The task was framed as a regression problem.

Modeling Pipeline

* Data preprocessing & scaling
* Exploratory Data Analysis (EDA)
* Model training with Grid Search
* Evaluation using MAE, MSE, RMSE, and average residuals

Key Results

* LiDAR consistently showed the strongest correlation with water level.
* Tree-based models were top performers:
  + Gradient Boosting: Best across all metrics
  + Decision Tree: Strong with minimal tuning
  + Random Forest: Reliable across conditions
* Ultrasonic sensors were less effective in high turbidity.
* Accelerometer/Gyroscope data had minimal predictive value.
* XGBoost was excluded due to flat, non-responsive output.

Performance Ranking (Top Models)

1. Gradient Boosting - Best accuracy and lowest error
2. Decision Tree - 2nd in most metrics, lowest residual
3. Random Forest - Consistently strong, 3rd overall  
   Other models like Polynomial Regression and SVM performed poorly, with outliers or skewed errors.

Conclusion & Future Work

LiDAR combined with ML offers an accurate, scalable solution for water level monitoring-even in murky conditions-using affordable hardware. Future steps may include real-world river testing, neural network integration, IoT deployment, and time-series forecasting.

References:

1. <https://harcresearch.org/news/when-technology-and-nature-collide-lidar-and-the-importance-of-biodiversity/?gad_source=1&gad_campaignid=9558509408&gbraid=0AAAAACg5UaQcPomhU5IOhcKP991S_Zt33&gclid=Cj0KCQjw-ZHEBhCxARIsAGGN96JqAcTmzLz95JuFdmWgYF7HZ0INDpa1naJ_Ixy2ex9dtBtA26Crut0aAgWUEALw_wcB>

2. <https://www.kaggle.com/datasets/caetanoranieri/water-level-identification-with-lidar>

3. See Attached File

4. <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html>

5. <https://aws.amazon.com/what-is/hyperparameter-tuning/>

6. <https://medium.com/@abhishekjainindore24/svm-kernels-and-its-type-dfc3d5f2dcd8>

7. <https://duchesnay.github.io/pystatsml/machine_learning/ensemble_learning.html>

8. <https://medium.com/analytics-vidhya/what-is-gradient-boosting-how-is-it-different-from-ada-boost-2d5ff5767cb2>

9. <https://xgboost.readthedocs.io/en/stable/parameter.html>