Introduction

Introduction

Aquaculture relies on the cultivation of aquatic organisms, necessitating effective environmental monitoring for sustainable practices. The increasing deployment of sensor technologies provides vast amounts of data, but extracting meaningful insights and managing associated complexities require advanced analytical approaches. This review explores how machine learning techniques can be leveraged to analyze historical aquaculture data (including farm management records, yield information, and diverse environmental sensor readings) to uncover significant associations between various factors and improve decision-making processes.

Methodology

The application of machine learning in aquaculture involves several distinct but related approaches. For predicting the closure of shellfish farms due to environmental stressors or poor conditions, researchers utilize data gathered by extensive **sensor networks** (Figure 1). This methodology employs feature ranking algorithms to pinpoint the most influential variables contributing to closure decisions and utilizes time series analysis techniques like Principal Component Analysis (PCA) and Auto Correlation Function (ACF) for predictive modeling. Another focus is on predicting harmful algae blooms, achieved through extracting logical rules from sensor data using ensemble methods to identify critical environmental drivers associated with bloom formation and propagation along waterways.

To address the common issue of missing or incomplete sensor readings, an approach based on multiple classifiers predicts relevant events directly from available data rather than relying solely on imputation. This method demonstrates superior performance compared to traditional techniques (Figure 5). Furthermore, machine learning facilitates model relocation – adapting algorithms developed for one geographical location to predict conditions accurately in a new, similar area. This is accomplished by comparing environmental characteristics and validating the transferred models' effectiveness visually against ground truth data (Figure 4).

In terms of habitat mapping, particularly benthic habitats on the seafloor, machine learning integrates with image processing techniques. By analyzing visual data from cameras deployed underwater, these algorithms automatically generate detailed habitat maps, showcasing high accuracy in classification tasks (Figure 2). Finally, ensuring reliable sensor data is crucial for informed decisions; an ensemble classifier approach guided by clustering principles provides a robust method for assessing the quality of incoming sensor readings, effectively distinguishing between valid measurements and erroneous ones based on patterns learned from multi-classifier outputs (Figure 3).

Results

The implementation of machine learning techniques yielded significant advancements across multiple aquaculture domains. For instance, predictive models for shellfish farm closures demonstrated high accuracy in anticipating events by analyzing data streams from sensors monitoring water quality parameters like temperature, salinity, dissolved oxygen, and nutrient levels (Figure 1). Feature ranking analyses consistently identified specific environmental factors as primary triggers for closure decisions.

In the context of algae bloom prediction, ensemble methods successfully extracted complex rules governing algal growth dynamics. These models effectively correlated various environmental variables with bloom intensity and spatial propagation patterns along waterways (Figure 6), providing valuable lead times for preventative measures. The multiple classifier system for handling missing sensor data proved highly effective in predicting operational events without the need for potentially inaccurate imputation, achieving performance levels comparable to ground truth analysis.

The model relocation technique showed considerable promise by effectively transferring predictive capabilities between geographically distinct sites based on environmental similarity metrics (Figure 4). This significantly mitigates issues arising from sparse or unavailable local calibration data. For benthic habitat mapping, the integration of image processing and machine learning resulted in automated classification systems achieving accuracy levels often exceeding manual interpretation standards (Figure 2).

The sensor data quality assessment method achieved high precision and recall rates by employing multi-classifier ensembles combined with clustering-based under-sampling strategies to handle imbalanced datasets effectively. The confusion matrices generated clearly indicated the system's proficiency in identifying valid versus invalid sensor readings (Figure 3).