VIETNAM NATIONAL UNIVERSITY, HO CHI MINH CITY
UNIVERSITY OF INFORMATION TECHNOLOGY

SOCIALIST REPUBLIC OF VIETNAM
Independence – Freedom - Happiness

RESEARCH PROPOSAL

VIETNAMESE TITLE: ĐO PH KHÔNG XÂM LẤN TRONG NƯỚC SỬ DỤNG QUANG PHỔ CẬN HỒNG NGOẠI VÀ HỌC MÁY

ENGLISH TITLE: NON-INVASIVE PH MEASUREMENT IN WATER USING NEAR-INFRARED SPECTROSCOPY AND MACHINE LEARNING

Supervisor(s) (Học vị, họ và tên, nơi công tác cán bộ hướng dẫn và cán bộ hướng dẫn phụ nếu có)**:**

- PhD, Phạm Quốc Hùng, Faculty of Computer Engineering, UIT, VNUHCMC
- Master's Student, Nguyễn Thành Nhân, Faculty of Computer Engineering, UIT, VNUHCMC

Implementation Time: Từ ngày……..đến ngày……

Students:

Nguyễn Duy Khanh - 21520979

Hà Việt Hoàng - 21520240

2. Project Content:

2.1 Overview of the project *(sinh viên cần nêu rõ các đề tài, sản phẩm liên quan đã được nghiên cứu hoặc đã có trên thị trường trước thời điểm hiện tại và nêu thực trạng của chúng, để từ đó nêu bật được lý do thực hiện nghiên cứu trong KLTN này)*:

Water pH is an important indicator in various fields such as environment, industry, agriculture, and healthcare. However, traditional pH measurement methods such as electrodes or litmus paper have many drawbacks, including the need for direct contact with the sample, contamination of the sample, sensor degradation over time, and difficulty in continuous measurement.

The non-invasive pH measurement solution using Near-Infrared Spectroscopy (NIR) combined with machine learning helps eliminate these limitations. NIR utilizes changes in the absorption spectrum of water according to pH to predict the pH value without physical contact with the sample, minimizing noise and significantly extending the lifetime of the measurement system.

2.2 Objective of the project (Nêu cụ thể mục tiêu của KLTN, đặc biệt phải nêu được mục tiêu cải tiến sẽ là gì so với thực trạng nêu trong phần tổng quan đề tài):

General Objective:

Develop a non-invasive pH measurement system for water that is accurate, easy to use, and applicable in various fields such as environment, industry, and agriculture.

Specific Objectives:

- Develop a non-invasive pH measurement system using NIR sensor AS7265x, halogen light source, and Raspberry Pi 4B.
- Collect and construct an experimental dataset consisting of NIR spectra and pH values of 500 water samples for machine learning model training.
- Research and apply machine learning algorithms such as Gaussian Process Regression (GPR) and Random Forest (RF) to accurately predict pH based on NIR spectra.
- Test and optimize the system to ensure high accuracy, real-world applicability, and system stability.

Improvements Over Existing Solutions: This system presents several advancements over traditional pH measurement methods, such as glass electrodes and litmus paper:

- Non-contact measurement: This feature prevents contamination and sensor degradation, leading to more accurate and long-lasting results.
- Rapid, continuous, and automated measurement: Unlike traditional electrodes that necessitate frequent calibration, this system provides real-time pH measurements with minimal maintenance.
- Reduced long-term costs: It minimizes expenses related to sensor replacement and ongoing calibration.
- Versatile application: This system is suitable for diverse fields, including environmental monitoring, agriculture, and wastewater treatment, eliminating the need for manual sampling.

2.3 Methodology (Nêu tổng quan phương pháp thực hiện):

2.3.1 Hardware System Design

- NIR Sensor: Measures the water spectrum at 18 wavelengths from 410-940 nm.
- Halogen Light Source: Provides a broad-spectrum light to excite the sample.
- Raspberry Pi 4B: Controls the system and processes the data.
- OLED Display (SSD1306): Displays the measurement results.
- Protective Enclosure: Made of black mica to reduce external light interference.

2.3.2 Data Collection and Processing

- Dataset Acquisition: Prepare 500 water samples with pH ranging from 3.08 to 9.10.
- Reference pH Measurement: Use a precise pH meter with ±0.05 accuracy.

- NIR Spectra Recording: Use the AS7265x sensor to capture the spectra for each water sample.

2.3.3 Data Preprocessing

- Standard Normal Variate (SNV): Reduces noise and normalizes the data.
- Savitzky-Golay Filtering: Smooths the data while preserving essential features.

2.3.4 Machine Learning Model Training

Gaussian Process Regression (GPR): A non-parametric model capable of estimating uncertainty.

Random Forest (RF): A decision tree-based model effective with non-linear data.

Model Evaluation: Based on the coefficient of determination (R^2) and Root Mean Squared Error (RMSE).

3. Key Contents and Limitations of the Proposal:

3.1 Key Contents: This project focuses on developing a non-invasive pH measurement system utilizing the AS7265x NIR sensor, a halogen light source, and a Raspberry Pi 4B. The main components include:

3.1.1 Hardware System Design and Construction: This entails integrating the NIR sensor, light source, Raspberry Pi, and OLED display into a unified system.

3.1.2 Data Collection and Processing: An experimental dataset of 500 water samples will be generated, encompassing NIR spectra and corresponding reference pH values. The data will be preprocessed using Standard Normal Variate (SNV) and Savitzky-Golay methods to reduce noise and improve signal smoothness.

3.1.3 Machine Learning Model Training and Evaluation: Implement Gaussian Process Regression (GPR) and Random Forest (RF) algorithms to create a predictive model for pH level determination from NIR spectra. Model performance will be evaluated using R^2 and Root Mean Square Error (RMSE) metrics.

3.1.4 System Testing and Optimization: The accuracy, stability, and real-world applicability of the system will be assessed.

3.2 Limitations:

- pH Range: The study is limited to a pH range of 3.08 to 9.10 due to experimental equipment constraints.
- Water Sample Types: The analysis is restricted to laboratory water samples, excluding complex water types such as industrial wastewater with diverse chemical compositions.

- Machine Learning Algorithms: This project is limited to GPR and RF algorithms; exploring alternative algorithms that may yield better results is outside the scope of this study.
- Testing Environment: Evaluations are primarily conducted in a laboratory setting, without accounting for real-world environmental variables, such as ambient light fluctuations.

4. Implementation plan:

Month	Main Tasks	Student in Charge	Detailed Tasks
1	Research and System Design	Nguyễn Duy Khanh	- Conduct research on the applications of NIR spectroscopy and pH measurement. - Acquire knowledge about the required components. - Develop a thorough design for both hardware and software systems. - Formulate a data collection strategy and outline the experimental procedures.
2	Hardware Setup	Hà Việt Hoàng	- Purchase and assembly of hardware components. - Setting up Raspberry Pi development environment and installing necessary software.
3	Data Collection	Nguyễn Duy Khanh	- Conduct NIR spectra data collection for 500 water samples with varying pH values. - Reference pH measurement using a precise pH meter.
4	Machine Learning Model Training and Evaluation	Nguyễn Duy Khanh	- Preprocessing NIR spectra data. - Develop and train GPR and RF models to predict pH from NIR spectra. - Evaluate model performance. - Fine-tune and optimize machine learning models.
5	System Integration and Testing	Both	- Integrate trained machine learning model into the Raspberry Pi system. - Develop a simple user interface (OLED display) - System testing for accuracy, stability, and applicability.
6	Optimization and Report	Both	- System optimization based on testing results. - Write the final undergraduate thesis report. - Prepare presentation of results.

Supervisor's Confirmation (Signature and ful name)	**Ho Chi Minh City, 17th February 2025** **Student** (Signature and full name)