



Fr. Conceicao Rodrigues College of Engineering

Father Agnel Ashram, Bandstand, Bandra –west, Mumbai-50

Department of Computer Engineering

PROJECT PROPOSAL TEMPLATE

Group Information:

Roll No	Student Name	Div A/ B
9913 (Group Leader)	Mark Lopes	A
9914	Vivian Ludrick	A
9909	Pratyay Koley	A
9882	Aryan Daripkar	A

TITLE OF THE PROJECT:

AI-Based Disease Prediction in Rohu Fish Using Proteomics Data

**BROAD AREA (E.g. IoT, Machine Learning, Computer Vision, System Security etc.):
Machine Learning, Bioinformatics, Artificial Intelligence**

Category of the Project (Product based, Application based, or Research based):
Application Based

ABSTRACT (maximum 300 words):

Aquaculture faces major challenges due to disease outbreaks, which can severely impact fish stocks and cause substantial economic losses. Rohu (*Labeo rohita*), a key species in South Asian aquaculture, is particularly vulnerable to such diseases. Traditional detection methods are often slow, invasive, and may not offer timely results, increasing the risk of widespread losses. To address this, the project aims to develop a computational approach for rapid and accurate disease prediction in Rohu using mass spectrometry-based proteomics and artificial intelligence (AI).

The proposed solution involves building an AI-driven pipeline that analyzes protein expression patterns by leveraging high-quality proteomics data from healthy Rohu as a reference. By comparing healthy and diseased protein profiles, machine learning methods will be used to identify significant biomarkers and protein changes linked to disease. The process includes data cleaning, feature extraction, model training, and validation.

An interactive dashboard will be developed to present disease predictions and highlight key protein markers, supporting practical use by researchers and farm managers. This approach promises to improve early disease detection in Rohu, enabling faster response and smarter management, ultimately contributing to more resilient aquaculture and global food security.

MOTIVATION (maximum 100 word):

Disease outbreaks are a major cause of economic loss in aquaculture. Current detection methods can be slow or unreliable. By applying machine learning to Rohu proteomics data, we aim to enable rapid, accurate disease prediction, empowering farmers and researchers with actionable insights for timely intervention.



Fr. Conceicao Rodrigues College of Engineering

Father Agnel Ashram, Bandstand, Bandra –west, Mumbai-50

Department of Computer Engineering

PROBLEM STATEMENT:

How can artificial intelligence and proteomics data be harnessed to predict disease states in Rohu fish, enabling early detection and intervention in aquaculture environments?

SDGs Mapped:

SDG 2: Zero Hunger

SDG 3: Good Health and Well-being

SDG 9: Industry, Innovation, and Infrastructure

OBJECTIVES:

- Collect and preprocess proteomics data from healthy Rohu tissues (baseline) and, where possible, from diseased conditions.
- Design and implement data pipelines for feature extraction (protein/peptide profiles).
- Develop and evaluate AI/ML models to classify or predict disease states based on proteomic signatures.
- Identify and prioritize candidate protein biomarkers indicative of disease.
- Build an interactive web tool for data visualization, model predictions, and biomarker exploration.

METHODOLOGY (TENTATIVE IF ANY):

Data Collection: Obtain mass spectrometry-based proteomics data from both healthy and diseased Rohu tissues covering various organs.

Data Preprocessing: Clean, normalize, and transform protein/peptide data, addressing missing values as needed.

Feature Engineering: Build data matrices at the protein/peptide level and extract disease- or organ-specific features.

Machine Learning Pipeline: Conduct exploratory analysis and use machine learning methods to classify samples and identify biomarkers.

Visualization: Create an interactive dashboard for intuitive exploration of model results and biomarker insights.



Fr. Conceicao Rodrigues College of Engineering

Father Agnel Ashram, Bandstand, Bandra –west, Mumbai-50

Department of Computer Engineering

HARDWARE / SOFTWARE REQUIREMENTS:

Python (Pandas, Scikit-learn, NumPy, SciPy, Matplotlib, Seaborn)

Google Collab

React/ Next (for UI)

INNOVATIVENESS:

AI and machine learning enable species-specific disease prediction from proteomics data in aquaculture by integrating advanced data mining, statistical modeling, and interactive visualization, with flexible potential to expand such solutions to other aquatic species and multi-omics applications.

SOCIETAL RELEVANCE? (e.g. Health, Agriculture, Environment, Smart Solution Etc...)

Health: Supports the timely detection and management of diseases, improving fish welfare and consumer safety.

Agriculture & Food Security: Minimizes losses, contributing to sustainable aquaculture and global food supply.

Smart Solution: Provides AI-powered tools for non-specialist end-users in the aquaculture sector.
