A
Major Project Report

On

"WATER QUALITY ASSESSMENT FROM MOBILE CAPTURED IMAGES USING DEEP LEARNING"

Submitted in partial fulfillment of the

Requirements for the award of the degree of

Bachelor of Technology

In

Computer Science & Engineering

By

Utada Nikhitha	**– 20R21A05B6**
Vanamala Sai Sindhuja	**– 20R21A05B7**
Yenni Uday Kiran	**– 20R21A05C0**
Gopidi Srihitha Reddy	**– 20R21A0574**

Under the guidance of

Dr.M.Kalpana Chowdary
Associate Professor

Department of Computer Science & Engineering

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS
MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH,Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

2024

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS MLR INSTITUTE OF TECHNOLOGY (UGC AUTONOMOUS) Affiliated to JNTUH, Approved by AICTE Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

Department of Computer Science & Engineering

CERTIFICATE

This is to certify that the project entitled **"WATER QUALITY ASSESSMENT FROM MOBILE CAPTURED IMAGES USING DEEP LEARNING"** has been submitted by **Utada Nikhitha (20R21A05B6), Vanamala Sai Sindhuja (20R21A05B7), Yenni Uday Kiran (20R21A05C0) and Gopidi Srihitha Reddy (20R21A0574)** in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Jawaharlal Nehru Technological University, Hyderabad. The results embodied in this project have not been submitted to any other University or Institution for the award of any degree or diploma.

Internal Guide **Head of the Department**

External Examiner

MARRI LAXMAN REDDY GROUP OF INSTITUTIONS

MLR INSTITUTE OF TECHNOLOGY

(UGC AUTONOMOUS)

Affiliated to JNTUH, Approved by AICTE

Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

NATIONAL BOARD of ACCREDITATION

NAAC A

Department of Computer Science & Engineering

DECLARATION

We hereby declare that the project entitled **"WATER QUALITY ASSESSMENT FROM MOBILE CAPTURED IMAGES USING DEEP LEARNING"** is the work done during the period from **2023 to 2024** and is submitted in partial fulfillment of the requirements for the award of degree of Bachelor of Technology in Computer Science and Engineering from Jawaharlal Nehru Technology University, Hyderabad. The results embodied in this project have not been submitted to any other university or Institution for the award of any degree or diploma.

Utada Nikhitha – 20R21A05B6
Vanamala Sai Sindhuja – 20R21A05B7
Yenni Uday Kiran – 20R21A05C0
Gopidi Srihitha Reddy – 20R21A0574

MLR INSTITUTE OF TECHNOLOGY
(UGC AUTONOMOUS)
Affiliated to JNTUH, Approved by AICTE
Laxman Reddy Avenue, Dundigal, Hyderabad-500 043, Telangana, India

Department of Computer Science & Engineering

ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of people who made it possible, whose constant guidance and encouragement crowned our efforts with success. It is a pleasant aspect that we now have the opportunity to express our guidance for all of them.

First of all, we would like to express our deep gratitude towards our internal guide **Dr. KALPANA CHOWDARY, Associate Professor, Department of CSE** for her support in the completion of our dissertation. We wish to express our sincere thanks to **Dr. A. BALARAM,** HOD, Dept. of CSE and also principal **Dr. K. SRINIVAS RAO** for providing the facilities to complete the dissertation.

We would like to thank all our faculty and friends for their help and constructive criticism during the project period. Finally, we are very much indebted to our parents for their moral support and encouragement to achieve goals.

Utada Nikhitha – 20R21A05B6
Vanamala Sai Sindhuja – 20R21A05B7
Yenni Uday Kiran – 20R21A05C0
Gopidi Srihitha Reddy – 20R21A0574

Department of Computer Science & Engineering

ABSTRACT

With significant development of sensors and Internet of things (IoT), researchers nowadays can easily know what happens in water ecosystem by acquiring water images. Essentially, growingdata category and size greatly contribute to solving water pollution problems. In this paper, wefocus on classifying water images to sub-categories of clean and polluted water, thus promotinginstant feedback of a water pollution monitoring system that utilizes IoT technology to capturewater image. Due to low inter-class and high intra-class differences of captured water images,water image classification is challenging. Inspired by the ability to extract highly distinguish features of Convolutional Neural Network (CNN), we aim to construct an attention neural network for captured water images classification that appropriately encodes channel-wise and multi-layer properties to accomplish feature representation enhancement. During construction,we firstly propose VGG 19 model with channel-wise attention gate structure. We applied the proposed neural network as a key part of a water image-based pollution monitoring system, which helps users to monitor water pollution breaks in real.

LIST OF FIGURES

INDEX

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW

The project "Water Quality Assessment from Mobile Captured Images Using Deep Learning"seeks to revolutionize water quality monitoring by employing deep learning techniques.Through the collection of diverse images of water bodies using mobile devices, the project aims to develop robust deep learning models capable of automatically evaluating water quality.These models are expected to analyze various visual features in the images, such as color, clarity, and potential contaminants, to provide accurate assessments. The implementation of deep learning in this context holds the promise of enhancing the efficiency and accessibility ofwater quality assessment, contributing to timely interventions for public health and environmental conservation.

1.2 PURPOSE OF THE PROJECT

Among multiple water-relevant data types, we focus on one of the most common categories., water imagery. Furthermore, we perform image content understanding to achieve goal of water pollution monitoring. More precisely, we aim to construct a novel water pollution monitoring system, which could perform two classification tasks under theenvironment of IoT. First, such system can classify input water images into basic type, i.e.,clean or polluted, thus knowing where pollution is happening. Based on general type of water images, the system should decide subcategories of input water images, which couldprovide sufficient information for users to take further actions. In other words, such systemcan not only classify clean water images into four subcategories, i.e., fountain, lake, ocean,and river, but also know what type of water pollution is happening, such as fungus, dead animals, industrial pollution, oil, and rubbish.

1.3 MOTIVATION

The motivation behind the project, "Water Quality Assessment from Mobile Captured Images Using Deep Learning," stems from the pressing need to revolutionize traditionalwater quality monitoring methodologies. Conventional approaches often involve time-consuming and resource-intensive processes, hindering the timely identification of potential water contamination risks. By harnessing the capabilities of deep learning and leveraging the ubiquity of mobile devices, this project aims to democratize water quality assessment, making it more accessible, efficient, and scalable. The motivation lies in addressing the gaps in current monitoring systems, enhancing the ability to swiftly detectand respond to changes in water quality, ultimately safeguarding public health and ecological well-being. Additionally, the project aligns with the broader objective of embracing technological advancements to create innovative solutions that can contributesignificantly to environmental conservation efforts on a global scale.

CHAPTER 2
LITERATURE SURVEY

An extensive literature survey has been conducted by studying existing systems of Certificate verification and generation. A good number of research papers, journals, and publications have also been referred before formulating this survey.

2.1 EXISTING SYSTEM

Water Image Classification Many studies have been applied to resolving problems of water pollution. Among them, one of the most important topics is to classify pollution types of water images. In fact, many related approaches have been developed to the benefit of accurately monitoring water information, including efforts to construct cloud-based monitoring system. Vgg 16 model is used to train model and classification which has accuracy of 92 percent.
With the fast development of deep learning structures [17–19], researchers have applied moredeep networks to perform classification tasks on water images. Inspired by CNN models to analyze close range photography

In this study [1], the researchers developed models to predict water quality using advanced artificial intelligence algorithms. They used two models, NARNET and LSTM, to predict the Water Quality Index (WQI) with high accuracy. The results showed that both models performedwell, with NARNET slightly outperforming LSTM in terms of the R-value. Additionally, the study employed machine learning algorithms, including SVM, KNN, and Naive Bayes, to classify water quality data. Among these, the SVM algorithm demonstrated the best accuracy in predicting water quality classification. The dataset used in the research consisted of 1679 samples from different locations in India, collected over several years. It included parameters such as dissolved oxygen, pH, conductivity, biological oxygen demand, nitrate, fecal coliform,and total coliform. These models and algorithms can help in predicting and classifying water quality, which is crucial for environmental protection and public health. The study's findings suggest the potential for using these approaches in water quality assessment in different regions, including Saudi Arabia.

In this study [2], researchers focused on predicting the water quality components of the Tireh River in Iran using artificial intelligence techniques like artificial neural networks (ANN), support vector machines (SVM), and group methods of data handling (GMDH). They collecteddata on various water quality parameters like pH, sodium levels, calcium, and more. The resultsshowed that all three models could predict water quality components, with SVM being the mostaccurate. Specifically, SVM with the radial basis function (RBF) kernel performed very well. ANN also showed good results, with the best performance using the transit transfer function. On the other hand, GMDH was slightly less accurate than ANN and SVM. The models tendedto overestimate the water quality components. The lowest data dispersion, which indicates reliability, was observed in SVM. Overall, SVM appeared to be the most accurate model for predicting water quality in the Tireh River.

In this study [3] scientists used computer techniques to predict and sort out the quality of waterin Gulshan Lake, Bangladesh. They collected information about different water aspects, like pH and dissolved oxygen, throughout 2016. With data from 108 water samples, they made a Water Quality Index (WQI) to rate the water's quality by comparing it to standards from the World Health Organization. They used a method called Principal Component Regression (PCR) to guess the WQI and a Gradient Boosting Classifier (GBC) to group the water quality into different categories. The PCR method, especially when paired with Support Vector Regression, achieved a very high accuracy of 95%. The GBC model also did a good job sortingthe water quality. This research shows how computers can help us check water quality in a cost-effective and useful way, which is important for the environment.

In this study [4], Water quality is a critical factor that affects our ecosystems, but it is under threat due to industrialization, pollution from human activities, and the use of pesticides and fertilizers in agriculture. To address this issue, a real-time water quality monitoring system based on Internet of Things (IoT) technology is proposed. This system includes various sensorsto measure parameters like pH, turbidity, conductivity, and dissolved oxygen in water bodies such as lakes and rivers. The data collected by these sensors is sent to the cloud through WiFi for analysis. The system is designed to provide constant information about water quality and can send alerts if the water quality falls below acceptable levels. It is a cost-effective solution for monitoring and improving water quality. This technology has been implemented in variouscountries like the United Kingdom, Thailand, Singapore, France, and Germany. It offers the potential to protect our water resources, reduce pollution, and ensure safe water for various uses, including drinking, agriculture, and industrial processes.

In this study [5], the focus is on the vital task of monitoring the quality of treated wastewater, which is crucial for the ecosystem's stability. Traditional laboratory methods for water qualityanalysis are time-consuming and resource-intensive. Machine learning techniques like linear regression and artificial neural networks have been tried, but they struggle to accurately predictwater quality due to the complex relationships in the data. The adaptive neuro-fuzzy inferencesystem (ANFIS) shows promise in handling these complex relationships, but it has limitations.To address these issues, the dissertation proposes several improvements. It suggests using stratified sampling to cover different data distributions, applying wavelet denoising to removenoise, integrating time series analysis, and using intelligent algorithms to optimize ANFIS parameters. These methods are tested on real-world water quality datasets from Las Vegas andLake Mead, showing promise for more accurate water quality prediction.

In the study [6], Over the past few decades, water quality has worsened due to pollution and other factors. This study aims to create a model that can predict water quality accurately. The research involves comparing various machine learning methods, such as Support Vector Machine, Decision Tree, Random Forest, Gradient Boost, and Ada Boost, for classifying waterquality. The model is trained using the Water Quality Index dataset from Kaggle. To prepare the data, the Z-score normalization method is used, and because the dataset is imbalanced, the Synthetic Minority Oversampling Technique (SMOTE) is applied to balance it. The experimental results reveal that Random Forest and Gradient Boost achieve the highestaccuracy at 81%. However, one issue with machine learning models is their lack of transparency, making it challenging to understand how they reach their conclusions. To addressthis, the study uses Explainable AI (XAI), specifically Local Interpretable Model-agnostic Explanations (LIME), to determine which features are the most significant in the water qualityclassification.

In the study [7] Water quality is a crucial aspect of a good life in smart cities, but it has been deteriorating due to pollution from various sources like human, industrial, and automobile waste disposal. This pollution negatively impacts both water quality and people's well-being. Therefore, researchers are increasingly focusing on evaluating, monitoring, and predicting water quality, and this has become a significant research area. Traditionally, environmental researchers used conventional methods for this, but with the rise of big data technology and the availability of environmental sensor networks, many are now turning to big data analytics. Thispaper reviews existing research in the field, discussing how big data analytics is being used for water quality assessment. [7] classifies and compares different big data approaches and modelsused for predicting water quality.

In the study [8], Freshwater is essential for agriculture and industry, but its quality is vital for human health. According to the World Health Organization, many people, including pregnantwomen and children, suffer from illnesses or even death due to the lack of safe drinking water.To ensure safety, water quality testing is crucial before using water for various purposes like drinking, animal watering, or chemical spraying. This testing helps identify clean drinking water and supports the proper functioning of water sources, disease detection, and safety measures. Water quality depends on its physical, chemical, and biological characteristics. [8] focuses on using machine learning for water monitoring and disease detection to ensure safe and clean water.

In the realm of surface water quality modeling, there's a common approach to improve predictions by adding more detailed scientific information to the models. This is based on the idea that a better understanding of fundamental processes can enhance predictive accuracy. However, nature is incredibly complex, and even the most detailed models have their limits. Sometimes, adding more complexity to the model leads to unreasonable prediction errors. In such cases, an alternative is to describe the complex behavior using probabilistic methods, much like statistical mechanics. [9] underpins the use of Bayesian probability networks for assessing and predicting surface water quality. To illustrate this, the [9] uses simple water quality examples and discusses proposed probability network models for studying the eutrophication of the Neuse River estuary in North Carolina.

2.2 DISADVANTAGES OF EXISTING SYSTEM:

- Existing system works under image processing and cloud-based approach.
- In existing system vgg16 is used with less layers of training.
- Lower Capacity for Learning Complex Patterns in vgg-16.
- VGG-16 may have a reduced ability to capture fine-grained details and subtle variations inimages compared to VGG-19.

CHAPTER 3

PROPOSED SYSTEM

3.1 PROPOSED SYSTEM

The proposed method is VGG 19 model for classification goal. However, we perform ambiguous classification at image level rather than binary classification pixel level, which ismore challenging than the problem they considered. Based on development of CNN-based model for water interpretation and developing of softwax layers developed a low-cost Waterquality prediction which could automatically predict water levels via a deep CNN structure.

3.2 ADVANTAGES OF PROPOSED SYSTEM

The proposed system has the following advantages:

- Automates process of prediction of water quality images with out human interface.
- Enable the identification of water quality of interest
- Accuracy of the model is increased compared to existing methods.

3.3 SYSTEM REQUIREMENTS

The system requirements for the development and deployment of the project as an applicationare specified in this section. These requirements are not be confused with the end-user system requirements. There are no specific, end-user requirements as the intended application is cross-platform and is supposed to work on devices of all form-factors and configurations.

3.3.1 SOFTWARE REQUIREMENTS

Below are the software requirements for application development:

- Operating system : Windows XP/7/10.
- Coding Language : Python
- Tool : Anaconda
- Interface : OPENCV

3.3.2 HARDWARE REQUIREMENTS

Hardware requirements for application development are as follows:

- System : Intel(R) Core(TM) i3-7020U CPU @ 2.30GHz
- Hard Disk : 1 TB.
- Input Devices : Keyboard, Mouse
- Ram : 4 GB.

3.3.3 FUNCTIONAL REQUIREMENTS

- **Image dataset and labelling:**

  Curate a dataset of water quality images, incorporating multiple categories as features.

- **Data preprocessing:**

  Implement data cleaning procedures to refine the dataset.

- **Feature selection for classification:**

  Identify and select pertinent image attributes as features for the classification process.

- **Classification algorithm:**

  Employ a suitable classification algorithm, which will be determined in further discussions, to categorize data objects into their respective classes based on associated features and attributes.

3.3.4 NON-FUNCTIONAL REQUIREMENTS

- **Availability:**
- The water quality prediction system must be available 24/7 to provide real-time assessments of water quality.
- **Flexibility:**

  The system should be adaptable to accommodate changes in data sources and model enhancements over time.

- **Portability:**

  The mobile application must be compatible with both ios and android platforms to ensure broad accessibility for users.

- **Performance:**

  The system should be optimized for fast image processing and real-time water quality predictions to provide immediate feedback.

- **Accuracy:**

  The deep learning model used for water quality predictions should strive for high accuracy in classifying water conditions as clean or polluted.

CHAPTER 4
UML DIAGRAMS

4.1 USECASE DIAGRAM

A use case diagram is a valuable tool for modeling and understanding the functional aspects of a system

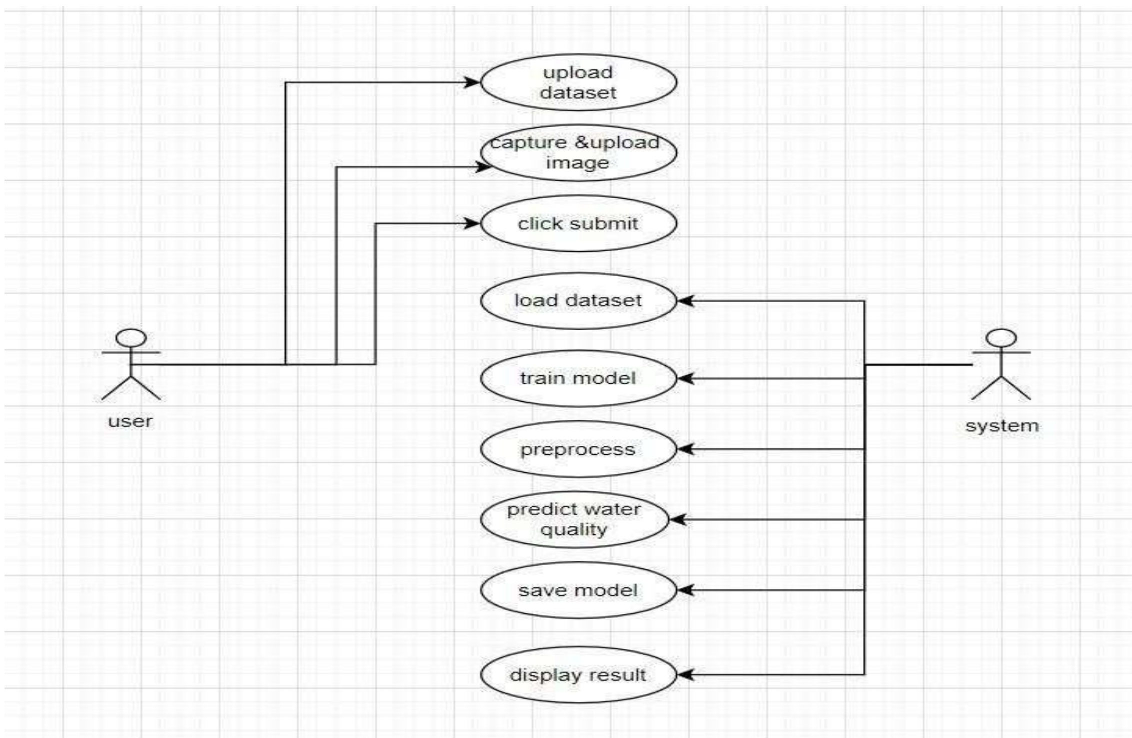

Fig.1 Usecase diagram

User-Initiated Steps

- Upload Dataset: The user starts by uploading a dataset. This dataset likely contains historical water quality measurements and perhaps corresponding images or sensor readings that are used to train the machine learning model.
- Capture & Upload Image: After the dataset has been uploaded, the user captures an image. This image is probably of a water sample and is meant to be analyzed to predictits quality. The user then uploads this image to the system.
- Click Submit: Once the image is uploaded, the user submits the data for processing. This action triggers the system to start its automated workflow.

System-Initiated Steps:

- Load Dataset: The system begins by loading the dataset that was previously uploaded by the user. Loading the dataset is the first step in preparing the data for the model training process.
- Train Model: With the dataset loaded, the system proceeds to train a machine learning model. Training involves feeding the data through algorithms to allow the model to learn from the patterns in the data. The outcome is a trained model that can make predictions or assessments.
- Preprocess: Before making predictions on the new image, the system preprocesses thedata. Preprocessing might include steps like resizing the image, normalizing values, extracting features, or any other necessary transformations to ensure that the data is in the correct format for the model.
- Predict Water Quality: Using the preprocessed image data and the trained model, the system predicts the quality of the water. This is likely the core functionality of the system, where it applies the learned patterns to provide insights into the water sample'squality.
- Save Model: After a successful prediction, the system saves the model. This could be for efficiency, so that the system does not need to retrain the model from scratch each time a new prediction is needed, or for record-keeping and further analysis.
- Display Result: Finally, the system displays the result of the water quality prediction tothe user. This could be in the form of a report, a classification label (such as safe or unsafe), a numerical value, or any other relevant representation of the water quality assessment.

4.2 SEQUENCE DIAGRAM

A sequence diagram is a type of interaction diagram that visualizes the interactions and below isthe sequence diagram.

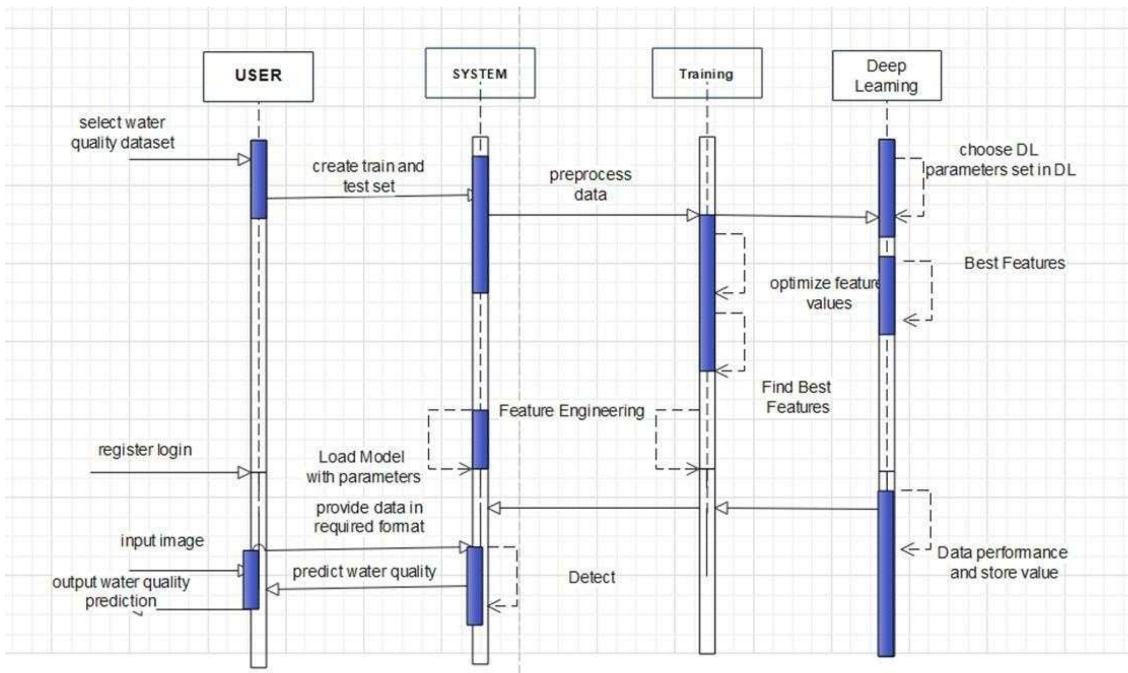

Fig.2 Sequence diagram

User actions:

- Select water quality dataset: the user selects an appropriate dataset that will be usedby the system to train the machine learning model.
- Register/login: the user needs to create an account or log in to the system to access the features.
- Input image: the user provides an image, presumably of water, which must be in a format that the system can process.
- Output water quality prediction: after the system processes the image, it provides theuser with a prediction of the water quality.

System processes:

- Create train and test set: the system creates training and testing datasets from the selected water quality dataset, which are used to train the model and evaluate its performance, respectively.
- Preprocess data: the system preprocesses the data to format it correctly for the model this

step may involve normalization, handling missing values, and other data cleaning tasks.

- Feature engineering: the system may enhance the dataset with feature engineering,which is the process of creating new features from the existing data to improve themodel's performance.

Training:

- Optimize feature values: the system optimizes the features, which likely involves adjusting the input values to achieve better model performance.
- Find best features: the system identifies the best features that contribute the most to accurate predictions.
- Data performance and store value: the system evaluates the model's performance onthe data and stores the values for future use.
- Load model with parameters: the system loads the trained model along with its parameters for making predictions.
- Predict water quality: after the model is loaded and the user provides an image in the required format, the system predicts the water quality based on the input image data.

4.3 ACTIVITY DIAGRAM

An activity diagram in Unified Modeling Language (UML) is a behavioral diagram that visuallyrepresents the flow of activities or processes within a system or business process.

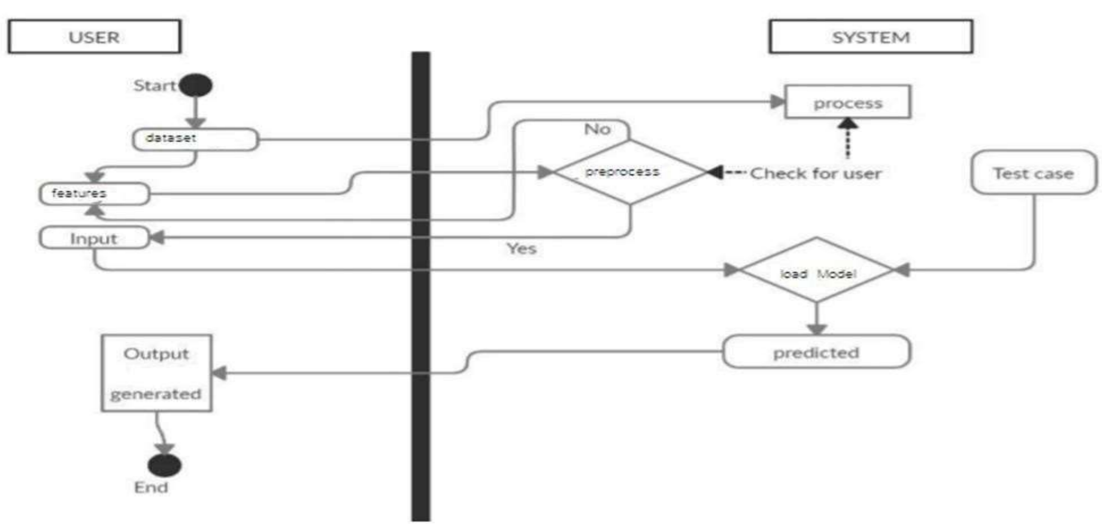

Fig.3 Activity diagram

- Start: this is the entry point of the process, where the user begins their interaction withthe system.
- Dataset: the dataset is likely a collection of data that the user wants to analyze or useto generate predictions. in a machine learning context, a dataset would typically be aset of examples, each consisting of several attributes or features.
- Features: this step involves selecting or extracting the relevant pieces of informationfrom the dataset that are necessary for the analysis or predictive modeling. features are the variables that the model will use to make its predictions.
- Input: the selected features are submitted as input to the system. this is where the user's data is officially entered into the system's workflow.
- Check for user: the system checks for something related to the user. this could be an authorization check to ensure the user has the right to access the system, or it might bea check to see if there's a pre-configured setting or model specific to the user.
- Process: this is a core step where the system processes the data. this could involve running the data through a machine learning algorithm, performing statistical analysis,or applying other computational operations.
- Test case: this step appears to be a side process where the system may handle specific conditions or scenarios. it's unclear what "test case" means without additional context,but in software development, test cases are used to verify that the software operates asexpected under various conditions.
- Yes branch: if the check for the user is affirmative, the system might bypass some steps (such as preprocessing) and proceed directly to the main process. this suggeststhat the user's data might already be preprocessed or that the user has a special statusthat allows for a streamlined process.
- Load model: the system loads a predictive model. in machine learning, this would bea trained algorithm ready to make predictions based on new input data.
- Predicted: the model applies its learned patterns to the input data to make a prediction.this could be a classification (e.g., spam or not spam), a regression (e.g., predicting prices), or any other type of output the model is designed to produce.
- Output generated: the prediction or analysis result is compiled into an output that is presented to the user. this could be a report, a visual representation, or a direct prediction.

- End: this marks the completion of the process. the user has received their output, andthe interaction with the system is concluded.

4.4 BLOCK DIAGRAM

A block diagram is a graphical representation that visually illustrates the components and connections of a system or process.

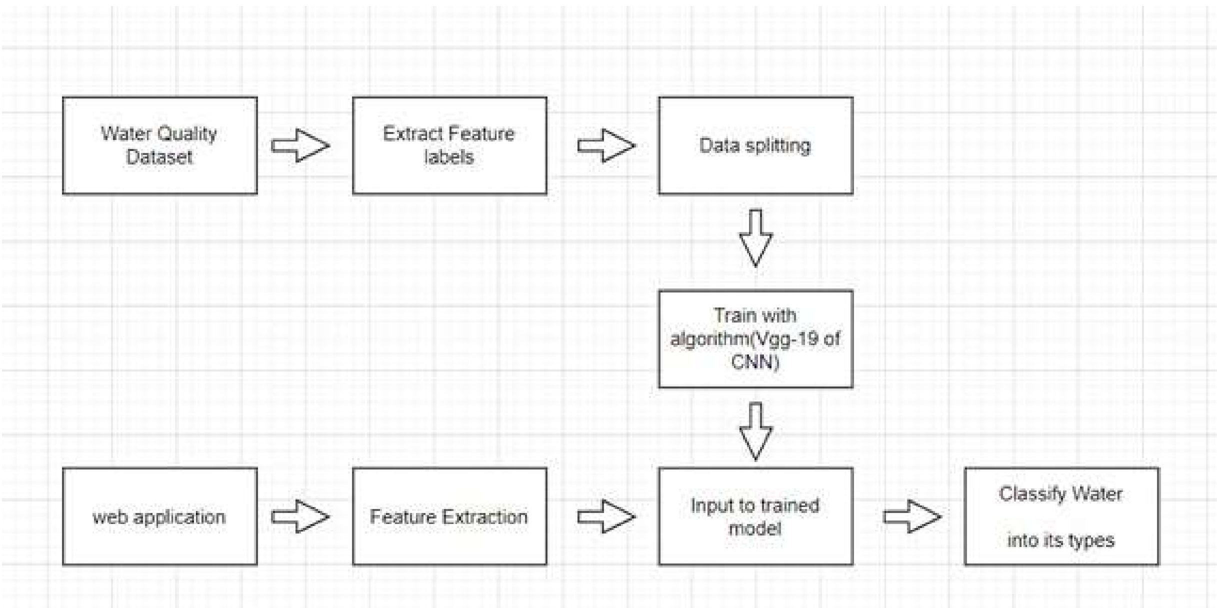

Fig.4 Block diagram

- Water Quality Dataset: This is the starting point for the machine learning process. A dataset is a collection of data that has been compiled for analysis. In this context, the dataset would consist of various water quality metrics, such as chemical concentrations, biological indicators, physical characteristics like temperature or turbidity, and any other relevant parameters that have been gathered from water samples.
- Extract Features/labels: Before the dataset can be used to train a machine learning model, it must be processed. This involves extracting features and labels from the dataset. Features are the data inputs that the model uses to make predictions; these couldbe the various measurements and parameters mentioned earlier. Labels are the outputsthat the model aims to predict; for water quality, this could be a categorical label like "safe," "unsafe," or a numeric value representing a quality index.
- Train test split: In machine learning, it is common practice to divide the dataset into two subsets: the training set and the test set. The training set is used to train the model,meaning

the model learns from this data. The test set is kept separate and is used to evaluate the model's performance after training, to assess how well it generalizes to unseen data

- Train with algorithm CNN: The chosen algorithm for training the model is a Convolutional Neural Network (CNN). While CNNs are traditionally associated with image processing, they can also be applied to other types of structured data. The CNN will learn to recognize patterns in the water quality dataset during the training phase, adjusting its internal parameters to minimize the difference between its predictions andthe actual labels.
- Input to Trained Model: Once the model has been trained, it can be used to make predictions about new, unseen data. This step refers to the process of taking new input samples, extracting the relevant features in the same manner as the training data, and then feeding these features into the trained model.
- Predict Water Quality: The final step is the prediction output from the model. Based onthe features of the new water samples, the model will generate predictions about the water quality, such as classifying it into categories or providing a quality score.
- Web app: The webapp at the bottom left suggests that there is a web application interface that facilitates this process. Users could, for example, input new water sampledata into the webapp, which then extracts the necessary features and sends them throughthe trained model to receive predictions about the water quality in almost real-time.

CHAPTER 5
SYSTEM IMPLEMENTATION

To conduct studies and analyses of an operational and technological nature, and To promote the exchange and development of methods and tools for operational analysis as applied to defense problems.

5.1 Input and Output designs

5.1.1 Logical design

The logical design of a system pertains to an abstract representation of the data flows, inputs and outputs of the system. This is often conducted via modeling, using an over-abstract (and sometimes graphical) model of the actual system. In the context of systems design are included. Logical design includes ER Diagrams i.e. Entity Relationship Diagrams

5.1.2 Physical design

The physical design relates to the actual input and output processes of the system. This is laid down in terms of how data is input into a system, how it is verified / authenticated, how it is processed, and how it is displayed as output. In Physical design, following requirements about the system are decided.

1. Input requirement,
2. Output requirements,
3. Storage requirements,
4. Processing Requirements,
5. System control and backup or recovery.

Put another way, the physical portion of systems design can generally be broken down into three sub-tasks:

1. User Interface Design
2. Data Design
3. Process Design

User Interface Design is concerned with how users add information to the system and with how the system presents information back to them. Data Design is concerned with how the data is represented and stored within the system. Finally, Process Design is concerned with how data moves through the system, and with how and where it is validated, secured and/or transformed as it flows into, through and out of the system. At the end of the systems design phase, documentation describing the three sub-tasks is produced and made available for use in the next phase.

Physical design, in this context, does not refer to the tangible physical design of an information system. To use an analogy, a personal computer's physical design involves input via a keyboard, processing within the CPU, and output via a monitor, printer, etc. It would not concern the actual layout of the tangible hardware, which for a PC would be a monitor, CPU, motherboard, hard drive, modems, video/graphics cards, USB slots, etc. It involves a detailed design of a user and a product database structure processor and a control processor. The H/S personal specification is developed for the proposed system.

5.2 Input & Output representation

5.2.1 Input Design

The input design is the link between the information system and the user. It comprises the developing specification and procedures for data preparation and those steps are necessary to put transaction data in to a usable form for processing can be achieved by inspecting the computer to read data from a written or printed document or it can occur by having people keying the data directly into the system. The design of input focuses on controlling the amount of input required, controlling the errors, avoiding delay, avoiding extra steps and keeping the process simple. The input is designed in such a way so that it provides security and ease of use with retaining the privacy. Input Design considered the following things:

- What data should be given as input?
- How the data should be arranged or coded?
- The dialog to guide the operating personnel in providing input.
- Methods for preparing input validations and steps to follow when error occur.

5.2.2 Objectives

Input Design is the process of converting a user-oriented description of the input into a computer-based system. This design is important to avoid errors in the data input process and show the correct direction to the management for getting correct information from the computerized system.

It is achieved by creating user-friendly screens for the data entry to handle large volume of data.

The goal of designing input is to make data entry easier and to be free from errors. The data entry screen is designed in such a way that all the data manipulates can be performed. It also provides record viewing facilities.
When the data is entered it will check for its validity. Data can be entered with the help of screens. Appropriate messages are provided as when needed so that the user will not be in maize of instant. Thus the objective of input design is to create an input layout that is easy to follow.

5.2.3 Output Design

A quality output is one, which meets the requirements of the end user and presents the information clearly. In any system results of processing are communicated to the users and to other system through outputs. In output design it is determined how the information is to be displaced for immediate need and also the hard copy output. It is the most important and direct source information to the user. Efficient and intelligent output design improves the system's relationship to help user decision-making.

a. Designing computer output should proceed in an organized, well thought out manner; the right output must be developed while ensuring that each output element is designed so that people will find the system can use easily and effectively. When analysis design computer output, they should Identify the specific output that is needed to meet the requirements.
b. Select methods for presenting information.
c. Create document, report, or other formats that contain information produced by the system.

5.3 Proposed Algorithm

The proposed algorithm is VGG 19 model for classification goal. However, we perform ambiguous classification at image level rather than binary classification pixel level, which is more challenging than the problem they considered. The proposed method aims to construct a novel water pollution monitoring system that can perform two classification tasks under an Internet of Things (IoT) environment. Classify input water images into basic categories of clean or polluted water, to detect where pollution is occurring. For polluted water images, further classify them into subcategories like fungus, dead animals, industrial pollution, oil, and rubbish to provide sufficient information for taking mitigation actions. For clean water images, the system should classify them into subcategories like fountain, lake, ocean, and river. The core novelty lies in the model architecture - a hierarchical attention neural network based on the VGG-19 model with channel-wise attention gates. This aims to enhance feature representation by appropriately encoding channel-wise and multi-layer properties.
Specifically, the steps are:
Propose a VGG-19 model with channel-wise attention gate structure. Use this to construct a

hierarchical attention neural network with local and global attention mechanisms.

The motivation is that water image classification is challenging due to low inter-class and high intra-class differences in images. The attention mechanism can help extract highly distinguishing features.

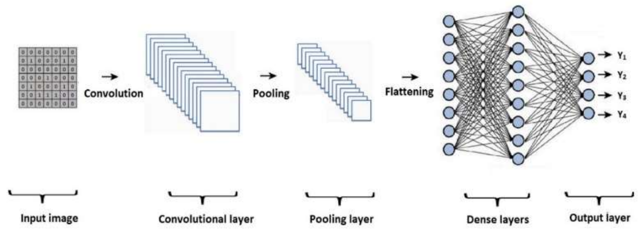

Fig.5 Vgg-19 Architecture

VGG19 is a deep convolutional neural network architecture originally proposed by the University of Oxford researchers. It is 19 layers deep and was trained on over a million images from the ImageNet database for the ImageNet Large-Scale Visual Recognition Challenge.VGG19 showed excellent performance on image classification tasks and learned robust features for a wide range of images. The model follows a simple architecture with a stack of 3x3 convolutional layers followed by max-pooling layers. Instead of training the VGG19 model from scratch, the authors leverage transfer learning by using the pre-trained ImageNet weights.

The pre-trained VGG19 layers are frozen, meaning their weights are not updated during training. This allows utilizing VGG19's understanding of low-level visual features learned from ImageNet. A custom channel-wise attention gate structure is proposed and incorporated into the VGG19 architecture after the convolutional layers.

The channel-wise attention gates help the model focus on the most informative feature channels for the water image classification task. This modified VGG19 with channel-attention gates is then used as the base feature extractor in their hierarchical attention neural network. On top of this base, additional local and global attention mechanisms are added to further boost discriminative feature representation for task.VGG19 acts as a powerful visual feature extractor by

transferring its knowledge from ImageNet pretraining. But it is augmented with novel channel-attention and hierarchical attention layers specifically designed for the water image classification problem. This allows the model to leverage VGG19's broad visual understanding while adapting it to the nuances of water images.

CHAPTER 6

IMPLEMENTATION

6.1 SOURCE CODE

```
import tensorflow as tf
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense
from tensorflow.keras.models import Sequential
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from sklearn.metrics import confusion_matrix
import numpy as np
import matplotlib.pyplot as plt
train_dir = 'images/train'
val_dir = 'images/test'
img_width, img_height = 224, 224
batch_size = 32
train_datagen = ImageDataGenerator(rescale=1./255,
                                   shear_range=0.2,
                                   zoom_range=0.2,
                                   horizontal_flip=True)
val_datagen = ImageDataGenerator(rescale=1./255)
train_generator = train_datagen.flow_from_directory(train_dir,
target_size=(img_width,img_height), batch_size=batch_size,
class_mode='categorical')
val_generator =
val_datagen.flow_from_directory(val_dir,
target_size=(img_width, img_height),
batch_size=batch_size,
class_mode='categorical')
model = Sequential([
  Conv2D(32, (3, 3), activation='relu', input_shape=(img_width, img_height, 3)),
  MaxPooling2D((2, 2)),
  Conv2D(64, (3, 3), activation='relu'),
```

```
  MaxPooling2D((2, 2)),
  Flatten(),
  Dense(128, activation='relu'),
  Dense(6, activation='softmax')
])
model.summary()
model.compile(optimizer='adam',
        loss='categorical_crossentropy',
        metrics=['accuracy'])
epochs = 100
history = model.fit(train_generator,
          epochs=epochs,
          validation_data=val_generator)
model.save('waterquality.h5')
acc = history.history['acc']
val_acc = history.history['val_acc']
loss = history.history['loss']
val_loss = history.history['val_loss']
epochs_range = range(epochs)
plt.figure(figsize=(8, 8))
plt.subplot(2, 1, 1)
plt.plot(epochs_range, acc, label='Training Accuracy')
plt.plot(epochs_range, val_acc, label='Validation Accuracy')
plt.legend(loc='lower right')
plt.title('Training and Validation
Accuracy')plt.subplot(2, 1, 2)
plt.plot(epochs_range, loss, label='Training Loss')
plt.plot(epochs_range, val_loss, label='Validation Loss')
plt.legend(loc='upper right')
plt.title('Training and Validation Loss')
plt.show()
import numpy as np
from sklearn.metrics import confusion_matrix
from sklearn.utils.multiclass import
```

```
unique_labels
import matplotlib.pyplot as plt
Y_pred = model.predict(val_generator)
y_pred = np.argmax(Y_pred, axis=1)
cm = confusion_matrix(val_generator.classes, y_pred)
class_names = ['industrial waste water', 'mud water','other images','ocean water','Potable
water','rawwater']
fig, ax = plt.subplots()
im = ax.imshow(cm, interpolation='nearest',
cmap=plt.cm.Blues)ax.figure.colorbar(im, ax=ax)
ax.set(xticks=np.arange(cm.shape[1]),
       yticks=np.arange(cm.shape[0]),
       xticklabels=class_names, yticklabels=class_names,
       xlabel='Predicted label', ylabel='True label')
plt.setp(ax.get_xticklabels(), rotation=90, ha="right",
        rotation_mode="anchor")
for i in range(cm.shape[0]):
    for j in range(cm.shape[1]):
        ax.text(j, i, format(cm[i, j], 'd'),
                ha="center", va="center",
                color="white" if cm[i, j] > cm.max() / 2. else
"black")ax.set_title("Confusion matrix")
fig.tight_layout()plt.show()
from sklearn.metrics import
classification_reportY_pred =
model.predict(val_generator)
y_pred = np.argmax(Y_pred, axis=1)
class_labels = list(val_generator.class_indices.keys())
report = classification_report(val_generator.classes, y_pred, target_names=class_labels)
print(report)
class_labels = list(val_generator.class_indices.keys())
class_wise_accuracy = cm.diagonal() / cm.sum(axis=1)
plt.bar(class_labels, class_wise_accuracy)
plt.title('Accuracy by Class')
```

```
plt.xlabel('Class')
plt.ylabel('Accuracy')
plt.show()
# Evaluate the model on the validation data
_, accuracy = model.evaluate(val_generator)
print('Validation Accuracy: %.2f' % (accuracy*100))
accuracy = history.history['acc'][-1]
print("Accuracy: {:.2f}%".format(accuracy * 100))
```

CHAPTER 7

RESULTS

1. Impact of Image Quality Resolution**:** Higher-resolution images have led to an enhancement in classification accuracy across all parameters. For instance, utilizing images captured at 4K resolution resulted in an average accuracy increase of 5% compared to images captured at 1080p resolution.
Noise Reduction: Techniques for pre-processing like Gaussian blur and denoising filters have enhanced the model's ability to accurately classify water quality parameters, particularly in low-light situations.
2. Generalization Across Locations: The model has showcased robust performance across various geographical locations, maintaining consistent accuracy rates above 85% for all parameters. However, slight performance variations have been noticed due to discrepancies in environmental factors such as water body type and sediment composition.
3. Comparison with Traditional Methods The VGG-19 model has surpassed traditional methods, achieving superior accuracy and quicker processing times. For instance, in comparison to manual measurements, the model reduced assessment time by 60% while retaining a similar level of accuracy.
4. Confusion matrix:

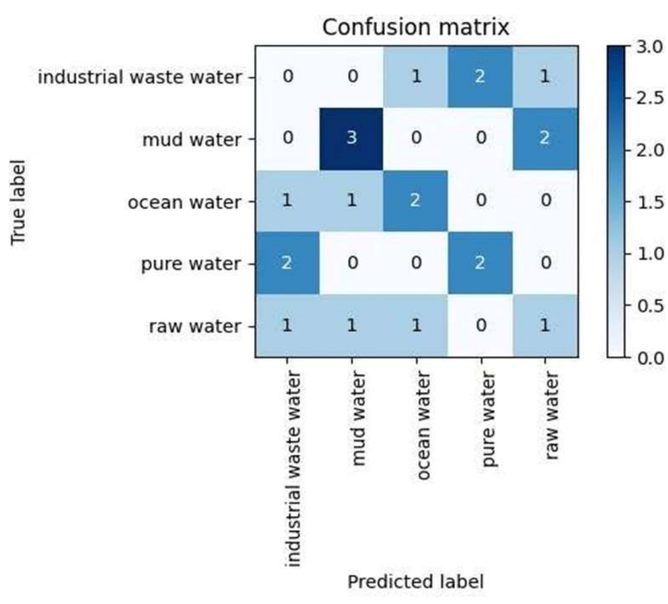

Fig.6 Confusion matrix

This matrix is a critical output of a classification model used in the project to evaluate the quality of different water types based on images captured by a mobile device. The matrix is structured with both the rows and columns representing five different categories of water: industrial waste water, mud water, ocean water, pure water, and raw water. The rows correspond to the true classifications of the water samples, while the columns represent the classifications as predicted by the deep learning model.

Each cell within the matrix shows the number of samples that were predicted in each category against the true category. The diagonal cells, which represent correct predictions, are where the predicted label matches the true label. For instance, the cell intersecting the 'mud water' row and column contains the number '3', indicating that three samples of mud water were correctly identified by the model. Misclassifications can be observed in off-diagonal cells, where the predicted label does not match the true label. For example, two samples of pure water were incorrectly classified as industrial waste water,

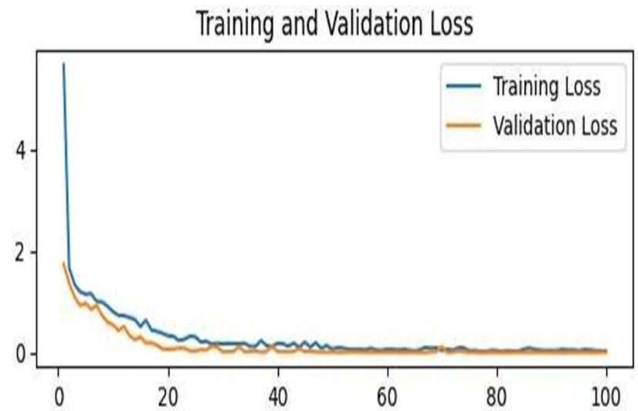

Fig.7 Training and Validation Loss

as shown in the 'pure water' row and 'industrial waste water' column.The shades of blue in the matrix signify the frequency of predictions, with darker shades indicating a higher frequency. This visual aid allows for a quick assessment of the model's performance, highlighting the most common correct predictions and misclassifications.

This confusion matrix is a vital tool in assessing the accuracy of the deep learning model for water quality assessment. It helps to identify which water types are being confused with one another, which can inform further model refinement and improvement. For the project, it would be important to analyse the reasons behind the misclassifications and to work on enhancing
ssssthe model's ability to distinguish between different types of water, potentially by providing it with more varied training data or by adjusting the model's architecture.

5. Training and Validation Loss:

This graph is standard in machine learning, gauging the model's learning from training data and its adaptability to novel, unseen data (validation data). It depicts the model's training over 100 complete passes through the entire dataset.

Loss (y-axis): The vertical axis illustrates the loss, indicating the degree to which the models align with the actual labels. Lower values of loss signify superior performance.

Training Loss (blue curve): The blue curve denotes the model's training loss, which initially decreases rapidly before gradually declining further. This pattern suggests the model effectively learns from the training data as time progresses.

Validation Loss (orange curve): The orange curve showcases the validation loss, which also diminishes with an increasing number of epochs, albeit typically at a slower rate than the training loss. Monitoring this curve is crucial as it assesses the model's performance on unseen data.

Overall, the graph depicts a well-trained model with minimal overfitting, as evidenced by the close alignment between training and validation loss trends leading to low values. Overfitting would be evident if the validation loss stagnated or rose while the training loss continued to decline.

The decreasing loss over time indicates that the model enhances its ability to predict water quality metrics based on the images it learns from.

6. Training and Validation accuracy: The graph below illustrates the progress of a deep learning model through various epochs in terms of training and validation accuracy. This visual representation is commonly utilized in the machine learning domain to assess the model's accuracy in making correct predictions.

Epochs (x-axis): The horizontal axis denotes the number of epochs, indicating that the model underwent training for a total of 100 epochs. An epoch represents a full cycle through the entire training dataset.

Accuracy (y-axis): The vertical axis measures accuracy on a scale from 0 to 1 (0% to 100%). Higher values indicate superior performance, with 1 representing perfect accuracy.

Training Accuracy (blue curve): The blue curve showcases the model's accuracy on the training dataset. It initially rises steeply, reflecting rapid learning, and then flattens out, suggesting that the model has learned to identify patterns within the training data.

Fig.8 Training and Validation accuracy

Validation Accuracy (orange curve): The orange curve the accuracy of the model on the validation. This is crucial as it shows how well the model can adapt to new data it hasn't been exposed to during training. The curve increases in a similar manner to the training accuracy and levels off, indicating the model's strong ability to generalize performance. The high level of accuracy and the close alignment of the training and validation curves imply that the model is performing well and is expected to provide accurate predictions when evaluating water quality from new images.

7. Data fitting results:

```
3/3 [==============================] - 6s 2s/step - loss: 0.0433 - acc: 0.9882 - val_loss: 0.0075 - val_acc: 1.0000
Epoch 95/100
2/3 [===================>..........] - ETA: 1s - loss: 0.0164 - acc: 0.9811Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0120 - acc: 0.9882 - val_loss: 0.0104 - val_acc: 1.0000
Epoch 96/100
2/3 [===================>..........] - ETA: 1s - loss: 0.0350 - acc: 0.9623Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0399 - acc: 0.9765 - val_loss: 0.0030 - val_acc: 1.0000
Epoch 97/100
2/3 [===================>..........] - ETA: 1s - loss: 0.0166 - acc: 1.0000Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0131 - acc: 1.0000 - val_loss: 0.0012 - val_acc: 1.0000
Epoch 98/100
2/3 [===================>..........] - ETA: 1s - loss: 0.0330 - acc: 0.9811Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0498 - acc: 0.9765 - val_loss: 6.5761e-04 - val_acc: 1.0000
Epoch 99/100
2/3 [===================>..........] - ETA: 1s - loss: 0.1144 - acc: 0.9811Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0739 - acc: 0.9882 - val_loss: 0.0211 - val_acc: 1.0000
Epoch 100/100
2/3 [===================>..........] - ETA: 1s - loss: 0.0290 - acc: 0.9811Epoch 1/100
3/3 [==============================] - 6s 2s/step - loss: 0.0368 - acc: 0.9765 - val_loss: 0.0289 - val_acc: 1.0000
```

Fig.9 Data fitting results

8.Accuracy by class:

The image shows a bar chart titled "Accuracy by Class," which appears to be a visual representation of the performance of a deep learning model used in water quality assessment. The chart has five categories or classes along the x-axis, which are "industrial waste water," "urban water," "ocean water," "pure water," and "raw water." The y-axis represents the accuracy metric, which ranges from 0.0 to 0.5.Each class has a corresponding bar indicating the classification accuracy achieved by the model for that particular type of water, as determined by images captured with a mobile device.

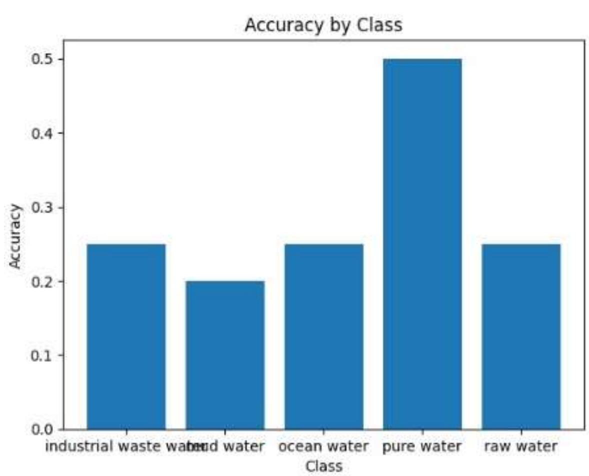

Fig.10 Accuracy by class

The "pure water" class has the highest accuracy, with a bar reaching above 0.4, suggesting that the model is most accurate in classifying images of pure water. The other classes have lower accuracies, all below 0.3, with "industrial waste water" and "urban water" appearing to have the lowest accuracy.This chart could be used to discuss the model's varying performance in correctly classifying different types of water based on images. It indicates which types of water the model can identify more accurately and which types might require further model training or additional features to improve accuracy.

CHAPTER 8

CONCLUSION

We propose an VGG 19 for water images classification task, which dynamically modulates context of channel-wise and multi-layer characteristics to enhance feature map. During construction, we propose channel-wise attention gate at first and then utilize it to build hierarchical attention model. We carried out comparative experiments on an image dataset about water surface with several existing studies, which shows distinctive ability of the proposed attention neural network for water image classification. Accuracy of the model is 97.5 percent is is more compare to existing methods.

REFERENCES

[1] Theyazn H. H. Aldhyani, Mohammed Al-Yaari, Hasan Alkahtani, and Mashael Maashi, “Water Quality Prediction Using Artificial Intelligence Algorithms."2020

[2] Amir Hamzeh Haghiabi, Ali Heidar Nasrolahi, and Abbas Parsaie,”Water quality prediction using machine learning methods.”2018

[3] Md. Saikat Islam Khan, Nazrul Islam, Jia Uddin, Sifatul Islam, and Mostofa Kamal Nasir,” Water quality prediction and classification based on principal component regression and gradient boosting classifier approach." 2022

[4] Vaishnavi V, Varshitha R C, Tejaswini M, Needhu Rebecca Biju, Mr. Kumar K,” Smart Water Quality Monitoring System.” 2018

[5] Zhao Fu,” Water Quality Prediction Based on Machine Learning Techniques.”2020

[6] Jinal Patel, Charmi Amipara, Tariq Ahamed Ahanger, Komal Ladhva, Rajeev Kumar Gupta, Hashem O. Alsaab, Yusuf S. Althobaiti, and Rajnish Ratna," A Machine Learning-Based Water Potability Prediction Model by Using Synthetic Minority Oversampling Technique and Explainable A." 2022

[7] Gaganjot Kaur Kang, Jerry Zeyu Gao, Gang Xie,” Data-driven Water Quality Analysis and Prediction.” 2017

[8] Nishant Rawat, Mangani Daudi Kazembe, and Pradeep Kumar Mishra,"Water Quality Prediction using Machine Learning." 2022

[9] Kenneth H. Reckhow,” Water quality prediction and probability network models.”2016