CHRONIC KIDNEY DISEASE PATIENT CARE APPLICATION

J.P.M.L.Perera

IT20226596

B.Sc. (Hons) Degree in Information Technology
Specializing in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2023

CHRONIC KIDNEY DISEASE PATIENT CARE APPLICATION

J.P.M.L.Perera

IT20226596

Dissertation submitted in partial fulfillment of the requirements for the Bachelor of Science (Hons) in Information Technology Specializing in Data Science

Department of Information Technology

Sri Lanka Institute of Information Technology
Sri Lanka

September 2023

DECLARATION

Declaration of the Candidate

I declare that this is my own work, and this proposal does not incorporate without acknowledgement any material previously submitted for a degree or diploma in any other university or Institute of higher learning and to the best of our knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to Sri Lanka Institute of Information Technology, the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Name	Student ID	Signature
J.P.M.L Perera	IT20226596	

Declaration of the Supervisor

The above candidates are carrying out research for the undergraduate Dissertation under my supervision.

Name of the supervisor: Ms. Wishalya Tissera

…………………………… …………………
Signature of the supervisor Date

Name of the co-supervisor: Mr. Samadhi Rathnayake

…………………………… …………………
Signature of the co-supervisor Date

ABSTRACT

In Sri Lanka, chronic kidney disease (CKD) poses a significant health concern, affecting an estimated 20% of the population. One of the factors contributing to CKD is exposure to contaminated water, which is particularly common in rural areas where access to clean water is limited. Real-time water quality monitoring is paramount for detecting contamination early and preventing CKD-related complications. This research proposes an IoT-based water quality monitoring system for CKD patients in Sri Lanka, which uses machine learning (ML) to predict water quality and displays the results in a mobile application dashboard. The system being proposed comprises two main components: hardware and software.

The hardware component encompasses IoT devices for measuring various Water temperatures, pH, and turbidity. These devices are connected to a cloud-based platform, which collects and stores the data in real-time. The software component includes an ML algorithm for predicting water quality based on the data pre-collected by NIFS. The ML model is trained on pre-tested water sample data and uses various algorithms to identify changes in water quality and predict future trends. The analysis results are displayed on a mobile application dashboard, providing users with real-time water quality information. The proposed system has several benefits: It enables real-time water quality monitoring, crucial for detecting contamination early and preventing CKD-related complications. IoT devices and cloud-based platforms allow remote water quality monitoring, particularly in rural areas with limited access to clean water. The ML algorithm provides accurate water quality predictions, which can help identify potential issues before they become serious problems.

The mobile application dashboard provides an easy-to-use interface for users to access the water quality data and receive alerts for any issues. The proposed system will be evaluated through field trials and user feedback. The performance and effectiveness of the IoT and ML-based water quality monitoring system will be measured against traditional manual monitoring methods. The user feedback will improve the design and make it more user-friendly.

Keywords: Internet of things (IoT), Machine Learning (ML), chronic kidney disease (CKD), National Institute of Fundamental Studies (NIFS)

ACKNOWLEDGEMENT

We want to express our sincere gratitude to the Department of Information Technology and SLIIT University for granting permission and necessary support. For the completion of this project. We want to express our gratitude to our supervisor, Ms Wishalya Tissera, and our co-supervisor, Mr Samadhi Rathnayake, for proper guidance from the beginning to the end of the project and also for giving us their valuable time every working day, providing us with new and innovative ideas to continue our project.

TABLE OF CONTENTS

LIST OF FIGURES

\

LIST OF ABBREVIATION

AWS: Amazon Web Services

CKD: Chronic Kidney Disease

EC: Electrical Conductivity

GUI: Graphical User Interface

IoT: Internet of Things

IoTaaS: IoT as a Service

JSON: JavaScript Object Notation

LCD: Liquid Crystal Display

LED: Light Emitting Diode

MCU: Micro Controller Unit

ML: Machine Learning

MLR: Multiple Linear Regression

MQTT: Message Queuing Telemetry Transport

pH: Potential of Hydrogen

RF: Random Forest

R&D: Research and Development

SVM: Support Vector Machine

1 INTRODUCTION

1.1 Background and Literature Survey

Chronic kidney disease (CKD) is a significant health concern in Sri Lanka, affecting a considerable percentage of the population. Among the various risk factors that affect CKD, exposure to contaminated water has been identified as a major concern. The poor water quality prevalent in many parts of the country is assumed to play a major role in the development and progression of this disease [1].

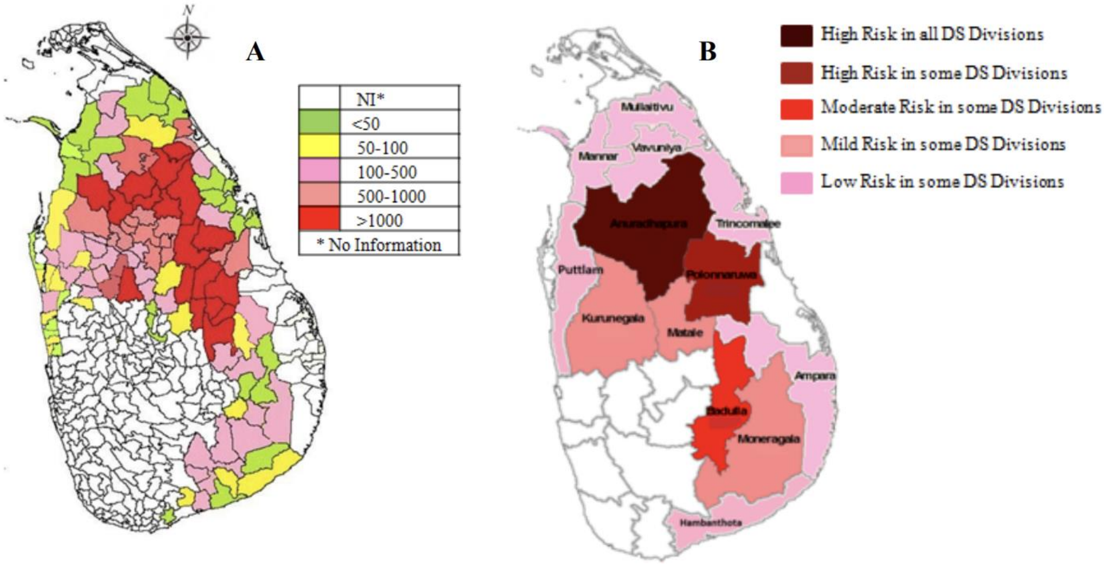


Figure 1: The reported CKD distribution in Sri Lanka [1]

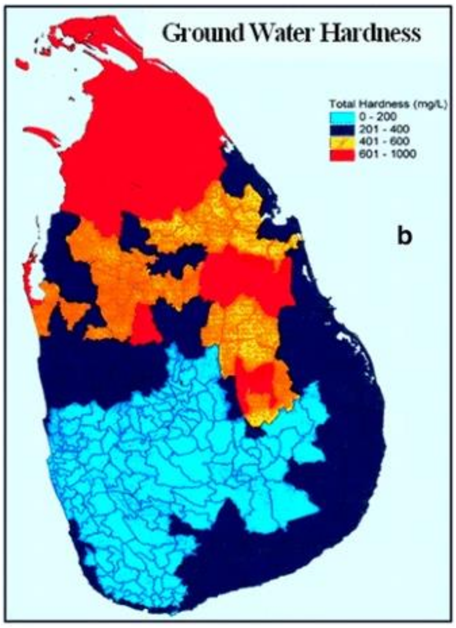


Figure 2: Ground water hardness in Sri Lanka [1]

Efforts have been made to improve water quality monitoring in Sri Lanka, but there still needs to be real-time, accurate, and reliable systems to address this issue effectively [1]. Fortunately, advancements on the Internet of Things (IoT) and machine learning (ML) technologies present promising opportunities for enhancing water quality monitoring practices. By leveraging IoT devices, real-time data on crucial water quality parameters like temperature, pH, and dissolved oxygen can be continuously collected. ML algorithms can then analyze this data, identifying changes in water quality patterns and predicting future trends. Although these technologies have proven successful in monitoring water quality in other countries, their implementation in Sri Lanka is still pending [2].

Temperature, pH, and turbidity are critical factors that can significantly impact kidney patients' well-being when it comes to drinking water. It is essential to maintain the temperature of drinking water within the range of 50-60°F (10-15°C) to prevent discomfort and avoid undue stress on the body. A pH level between 6.5 and 8.5 is crucial to maintain the body's acid-base balance, which is particularly important for individuals with impaired kidney function. Water clarity, indicated by turbidity, should be kept below 1 NTU to ensure that water is free from suspended particles and potential contaminants. These safe value ranges are effective in reducing the risk of exacerbating kidney-related health issues and ensuring that kidney patients can safely consume water without adding additional strain to their already compromised kidney function. Regular monitoring and water treatment are essential to meet these safe value ranges and safeguard the health of kidney patients [1].

To reduce this gap, a proposed IoT and ML-based water quality monitoring system aims to provide an efficient solution for kidney patients in Sri Lanka. The proposed system will include a network of IoT devices strategically placed to monitor essential water quality parameters continuously. These devices will gather data and transmit it to a central platform for analysis. ML algorithms, trained on the collected data, will be responsible for detecting any deviations in water quality and making predictions regarding future trends. One of the primary objectives of this system is to ensure the timely dissemination of critical information to kidney patients. By incorporating alerts and notifications, individuals can be promptly notified about water quality issues that may impact their health. This feature allows patients to take necessary precautions and adopt appropriate measures to manage their condition effectively [3].

Implementing the proposed IoT and ML-based water quality monitoring system in Sri Lanka holds numerous potential benefits. Real-time monitoring will enable instant water contamination detection, allowing immediate intervention and prevention of CKD-related complications [3]. Moreover, the system's remote monitoring capabilities are particularly beneficial for rural areas with limited access to clean water. Utilizing cloud-based platforms allows water quality data to be collected and analyzed from various locations, facilitating a comprehensive understanding of the situation across the country [4].

The machine learning (ML) algorithms used in this system are expected to produce accurate predictions regarding water quality, enabling early detection to address possible issues before they turn into major issues. Additionally, the user-friendly mobile application dashboard will offer those with kidney disease a simple interface to get real-time data on water quality. Patients will be kept informed about the quality of the water they drink due to this interface's simplicity of data retrieval.

Field experiments have been conducted to evaluate the effectiveness and performance of the IoT and ML-based water quality monitoring system, and user feedback has been used. This evaluation process will enable further improvements to be made to the system's design, enhancing its usability and effectiveness. Ultimately, the successful implementation of this system has the potential to revolutionize water quality monitoring practices in Sri Lanka, making a substantial positive impact on the lives of kidney patients and the overall population [24].

In a 2022 research publication, explored using machine learning techniques to identify alarming events in water quality. Their research focused on creating an edge device with sensors to measure water quality parameters. This device was designed to identify changes in water quality relative to baseline parameters and generate alert signals when water quality parameters exceeded predefined threshold values—additionally, the research aimed to classify various types of contamination [25]. The study employed three methods to calculate water quality: the Weighted Arithmetic Index, the N.S.F. Water Quality Index, and user feedback. Machine learning, specifically the Support Vector Machine (SVM), was employed to develop a lightweight model based on these water quality indexes. The results were promising, showcasing the potential of machine learning at the edge device level for the intelligent detection of alarming events in water quality [4].

In 2022, Amara Parangama and their research team conducted a study in Sri Lanka to investigate potential causative factors for CKD in the North Central Province (NCP). This region reported the country's highest incidence of CKD patients and fatality rates. There were suspicions that specific water quality measures in drinking water might be contributing to the issue. The study involved the analysis of water samples collected from shallow wells used for drinking water by CKD patients and non-patients in the NCP region. Various parameters were tested, including chemical species such as Cadmium, Sodium, Calcium, Fluorine, and Chlorine. Initial analysis revealed that most of the water quality parameters tested did not exceed the drinking water quality standards set by the World Health Organization (WHO) [27]. However, factor analysis techniques were employed to investigate critical water quality parameters that could be linked to CKD. The results indicated that water samples from CKD patients exhibited higher levels of Sodium (Na), Chlorine (Cl), Magnesium (Mg), Fluorine (F), and Calcium (Ca), which were grouped into one factor and identified as hydro-geologically originated. Another factor, Nitrogen (N) and Phosphorus (P), likely attributed to nutrients from fertilizers, was identified. Conversely, Cadmium (Cd) was classified as a distinct element. Interestingly, water quality measures in samples obtained from non-CKD patients did not fit into any specific group [27].

In 2021, Dinithi Weerasinghe and their team presented a research paper on CKD in Sri Lanka. Over the past two decades, CKD has become a significant global health concern. Sri Lanka has been affected by the rapid rise of CKD of unknown etiology (CKDu) in agricultural regions. The paper introduced a model based on an Artificial Neural Network (ANN) that utilized soil's physical and chemical characteristics in agricultural zones to determine the type of CKD. The study compared the performance of the Multilayer Perceptron (M.L.P.) ANN model with Decision Tree and Support Vector Machine (SVM) models in terms of accuracy, precision, recall, Root Mean Squared Error (R.M.S.E.), and Mean Absolute Error (M.A.E.). The findings demonstrated that the ANN model outperformed the other models in classification and prediction. It is critical for the early detection and management of CKD and its etiologies in Sri Lanka [28].

In 2023, Andrew O.M.A.M.B.I.A. and their team proposed a system for monitoring water quality and detecting pilferage using IoT and machine learning technologies. Access to safe water is considered a fundamental human right; however, contamination, leakages, and pilferage often occur in water supply systems. The proposed system aims to address these challenges by continuously monitoring water quality and using machine learning algorithms to detect pilferage and wastage. Water is primarily sourced from pipes and springs around towns and consumed by consumers. The system will enable decision-making using machine learning algorithms [29].

In 2019, Sathira Hettiarachchi, Divan Proboshena, Hashan Rajapaksha, and Lakshan Stembo proposed a solution to address the pressing need for comprehensive research on sustainable water-quality management systems. As population growth and environmental pollution continue to escalate, the demand for such systems has become increasingly apparent [5]. Their proposed solution was an innovative water quality management system incorporating predictive capabilities [30]. This system allowed for the frequent monitoring of water quality measures at water treatment facilities through a user-friendly IoT device. Additionally, it identified water leakage points within the water distribution network using crowdsourcing and visualization techniques. Notably, the proposed system achieved a remarkable 99% accuracy rate in predicting upcoming changes in water quality and calculating the corresponding purification costs. It featured a digital dashboard that provided concise information on leaks, customer feedback, patterns of water quality, and purification-related costs in a summarized manner [30].

Existing research and literature suggest that IoT and ML-based water quality monitoring systems have been successfully implemented in various parts of the world, including neighboring countries like India. These systems have demonstrated their effectiveness in improving water quality monitoring and reducing the risk of waterborne diseases. However, there needs to be more research regarding implementing such systems in Sri Lanka, particularly for addressing the specific needs of kidney patients. Therefore, the proposed research seeks to contribute to the existing literature by exploring the feasibility and effectiveness of a water quality evaluation and monitoring system tailored to the unique requirements of kidney patients in Sri Lanka [4].

1.2 Research Gap

The current situation of water quality monitoring systems in Sri Lanka presents significant research gaps that demand critical attention and innovative solutions. These gaps collectively underline the need for a comprehensive overhaul of existing water quality assessment and management approaches [15].

One of the most considerable research gaps is the need for water quality monitoring systems designed for the specific health needs of chronic kidney disease (CKD) patients. The prevalence of CKD is increasing in Sri Lanka. However, the country's monitoring systems are mostly generalized and not designed to address the unique vulnerabilities of CKD patients [12]. This obvious oversight calls for developing specialized solutions that prioritize the well-being of this patient. A significant research gap lies in underutilizing cutting-edge technologies in water quality monitoring systems, particularly the Internet of Things (IoT) and machine learning (ML) [18]. Despite the considerable advancements in these domains, their integration into the scope of water quality monitoring in Sri Lanka still needs to be explored. IoT devices can collect real-time data on crucial water quality parameters, while ML algorithms can analyze this data to detect patterns anomalies and predict future trends. The collaboration of these technologies can revolutionize the accuracy, efficiency, and timeliness of water quality assessments, which is critical for maintaining public health.

Historically, research and resources have centered on monitoring water quality from an industrial or regulatory standpoint. While industrial applications are essential, this emphasis on non-health-centric monitoring has resulted in a research gap in addressing the health-related concerns of the general population and CKD patients. The prevalence of CKD is increasing in Sri Lanka, which calls for reevaluating priorities favoring comprehensive health-focused monitoring systems [11].

The accessibility and usability of water quality data constitute another critical research gap. It is not enough to merely collect water quality data; it must be presented in a manner that is easily comprehensible and accessible to the public and healthcare professionals. User-friendly interfaces or platforms for disseminating water quality information are necessary for informed decision-making regarding water consumption and potential health risks [20]. Bridging this gap involves collecting accurate data and making it readily available and understandable to all stakeholders. Timely detection of water quality issues is paramount, yet it remains a significant research gap. Current systems may not provide real-time data, essential for early detection and intervention in cases of contamination or deterioration in water quality [16]. This timeliness is particularly crucial for CKD patients, as their health is complicatedly linked to the water quality they consume. A comprehensive research gap pertains to the balance between embracing technological advancements and ensuring affordability, especially in resource-constrained settings like Sri Lanka. While advanced technologies offer substantial benefits, they must be implemented in ways that do not exacerbate economic disparities. Striking this delicate balance is essential to ensure equitable access to cutting-edge water quality monitoring solutions. An essential research gap revolves around developing strong data-driven decision support systems. While gathering data is essential, it is equally important to be able to analyze it well and provide insights that can be put into action. Existing systems often need more advanced analytics and decision support mechanisms to translate raw data into meaningful information for healthcare professionals and the general population [19].

Lots of existing water quality monitoring systems exhibit limited scalability and geographic accessibility. These challenges extend monitoring coverage to different regions, especially in remote or rural areas. Ensuring that water quality monitoring solutions are accessible and can be deployed across diverse geographic locations is a critical research gap to ensure equitable access to safe drinking water. A considerable research gap exists in fostering interdisciplinary collaboration among experts in water quality engineering, healthcare, and technology. Collaboration across these domains is essential to ensure that water quality monitoring systems collect data and interpret it in a medically relevant and actionable way [14]. A multidisciplinary approach is critical to developing holistic solutions catering to the complex health challenges of water quality issues. Finally, an important research gap pertains to the limited emphasis on community engagement and education regarding water quality. Even with advanced monitoring systems, their effectiveness can be compromised if communities are not adequately informed about the importance of water quality and the actions, they can take to safeguard their health. Bridging this gap involves proactive efforts to engage communities, raise awareness, and educate individuals about water quality issues and their implications for public health [21].

These connected research gaps highlight the necessity of a novel approach to Sri Lankan water quality monitoring. Addressing these significant research gaps requires a comprehensive strategy that includes tailoring solutions to CKD patients, utilizing IoT and ML technologies, prioritizing health-centric monitoring, improving data accessibility and suitability, balancing technology with affordability, improving data-driven decision support, ensuring scalability, fostering interdisciplinary collaboration, and encouraging community engagement. It is crucial to close these gaps to protect the public's health and ensure that everyone in Sri Lanka has fair access to clean water [23].

1.3 Research Problem

The research problem revolves around the critical absence of a reliable and efficient real-time water quality monitoring system in Sri Lanka. This issue is particularly pertinent to kidney patients who rely on continuous access to healthy water for their health and well-being. The existing water quality monitoring methods in Sri Lanka are manual, leading to several interconnected challenges and gaps that necessitate urgent attention and innovative solutions [17].

The foremost aspect of the research problem is the need for a real-time water quality monitoring system. There needs to be an efficient mechanism to provide kidney patients or the general population with timely information about the water quality they are consuming. This deficiency in real-time monitoring can have dire consequences for kidney patients, who require specific and up-to-date information to manage their water intake effectively. Access to real-time data on water quality is an obvious research problem that poses a significant risk to public health [21], [25].

The reliance on manual monitoring methods makes worse the research problem. While manual methods are currently employed for assessing water quality in Sri Lanka, they need to be more constrained in their capacity to detect sudden changes or anomalies in water quality. This limitation is particularly worrisome for kidney patients who depend on consistent water quality. The inability of manual methods to promptly identify shifts in water quality presents a significant research problem, as it threatens the well-being of individuals who rely on this critical information for their health [25].

Another part of the research problem remains the continued need for more reliable data management tools with human monitoring techniques. Proper data management systems are necessary to respond to water quality issues. Vital information regarding water quality may need to be adequately recorded, analyzed, or acted upon promptly. This deficiency in data management adds complexity to the research problem, as it inhibits the effective utilization of available data to address water quality concerns promptly [9].

Furthermore, the research problem encompasses the main challenge of limited access to water quality information, which profoundly affects kidney patients. Even when data is accessible, it is often presented in challenging formats to access and understand. This lack of accessibility to information is a formidable research problem, as it deprives kidney patients of the necessary insights required to make informed decisions regarding their water consumption, which is critical for their health management [7].

A critical dimension of the research problem is language barriers that hinder kidney patients' access to water quality information. Much of the existing resources and information are available exclusively in English, creating a significant hurdle for patients who need to be proficient in the language. This language barrier compounds the research problem, as it impairs the existing challenges of information inaccessibility and understanding, further demoting those who require essential water quality data [19].

The research problem is the need for a reliable and efficient real-time water quality monitoring system in Sri Lanka. This shortcoming, when combined with the drawbacks of manual monitoring techniques, data management issues, informational accessibility issues, and language obstacles, offers a complex problem with broad consequences for kidney patients and the general public health. Addressing this research problem is paramount to ensuring equitable access to safe and high-quality drinking water in Sri Lanka, particularly for those most vulnerable due to health concerns.

1.4 Research Objectives

Main Objective:

The overarching aim of this research endeavor is to conceive, develop, and implement a state-of-the-art Internet of Things (IoT) and Machine Learning (ML)-based water quality monitoring system tailored to the specific needs of kidney patients in Sri Lanka. This comprehensive system seeks to provide real-time, precise, and reliable information concerning water quality parameters, thereby substantially mitigating the risks associated with kidney patients' exposure to contaminated water sources. The central objective is to enhance the overall well-being of kidney patients by minimizing the potential complications stemming from the consumption of compromised water.

Specific Objectives

- IoT Hardware Design and Development: The foremost specific objective is to undertake the design and development of a cutting-edge IoT-based hardware system. This system is meticulously engineered to facilitate the real-time collection of critical water quality parameters. It encompasses components such as sensors for parameters like temperature, pH, and dissolved oxygen. The primary focus here is on creating a robust hardware infrastructure capable of seamless data acquisition.

- Machine Learning Algorithm Implementation: The second specific objective delves into the realm of machine learning, with the purpose of crafting a robust and adaptive algorithm. This algorithm will harness the power of historical water quality data derived from pre-tested water samples. Its role is to predict deviations and alterations in water quality based on the incoming data streams. The core aspiration is to develop an algorithm with a high degree of accuracy in identifying fluctuations in water quality.
- Cloud-Based Integration: The third specific objective pivots toward the integration phase, where the IoT hardware and ML algorithm converge into a cohesive whole. This integration culminates in the creation of a cloud-based platform, which serves as the nexus for data processing and analysis. The platform's functionality encompasses the aggregation, storage, and interpretation of water quality data in real-time.
- Mobile Application Development: The fourth specific objective entails the development of a user-friendly mobile application dashboard. This dashboard is engineered to provide kidney patients with immediate access to real-time water quality data and alerts. It is designed to be intuitive, informative, and responsive, ensuring that users can effortlessly comprehend and act upon the information presented.
- Performance Evaluation and Field Trials: The fifth specific objective hinges on the evaluation of the IoT and ML-based water quality monitoring system's performance and effectiveness. This is achieved through rigorous field trials and the collection of invaluable user feedback. The research aims to substantiate the system's reliability, accuracy, and real-world applicability under diverse conditions.
- Iterative Enhancement: The final specific objective underscores the commitment to continual improvement and refinement. It recognizes that technology and healthcare dynamics are dynamic and subject to evolution. As

such, mechanisms for ongoing enhancement of the IoT and ML-based system will be embedded, ensuring its longevity and relevance in the context of kidney patient care.

The main objective and specific objectives of this research initiative converge to form a holistic mission of paramount significance. The core mission revolves around the development of a transformative IoT and ML-based water quality monitoring system, meticulously designed to safeguard the well-being of kidney patients in Sri Lanka. Through these objectives, the research seeks to harness technology's potential to enhance the quality of life for kidney patients by affording them real-time, accurate insights into the safety of their water sources.

2 METHODOLOGY

2.1 System Architecture

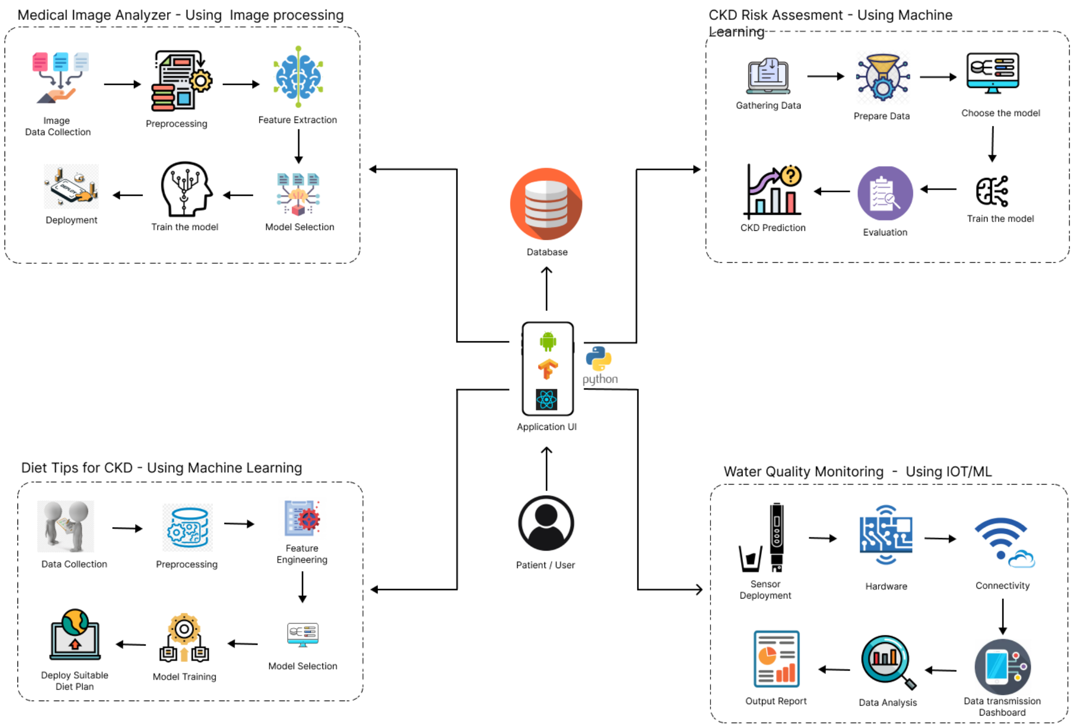


Figure 3: Overall system diagram

The overall system diagram for our research project is a testament to the synergy of cutting-edge technologies, seamlessly integrated to empower individuals with a holistic approach to health and well-being. At its core, our project leverages the power of Machine Learning (ML) and the Internet of Things (IoT) to offer a comprehensive health and lifestyle management solution accessible through a single, user-friendly mobile application.

The four central components of our project are as follows:

CKD Risk Assessment: This component employs advanced ML algorithms to assess an individual's risk of chronic kidney disease (CKD) based on a variety of health data inputs. By analyzing key health parameters, our system provides personalized risk assessments, enabling users to take proactive measures to safeguard their kidney health.

Diet Tips for CKD: Our system goes beyond risk assessment by offering personalized dietary recommendations. ML algorithms analyze user health data and dietary preferences to provide tailored diet tips aimed at promoting kidney health and overall well-being. These recommendations are dynamic and evolve with the user's health journey.

Medical Image Analyzer: This component utilizes DL to analyze medical images, aiding in the early detection and monitoring of health conditions. By integrating with medical imaging devices or apps, users can receive instant insights into their health, enhancing their ability to make informed decisions about their well-being.

Water Quality Monitoring: Ensuring access to clean and safe drinking water is paramount for good health. Our IoT-based water quality monitoring system continuously assesses the quality of water sources, providing real-time data and alerts to users. This feature is especially valuable for individuals concerned about the impact of water quality on their health.

All four components are seamlessly interconnected within a single, intuitive mobile application. Developed using React Native for the frontend and Flask for the backend, our application ensures a smooth and responsive user experience. The use of Amazon Web Services (AWS) for cloud services guarantees secure and reliable data storage and processing.

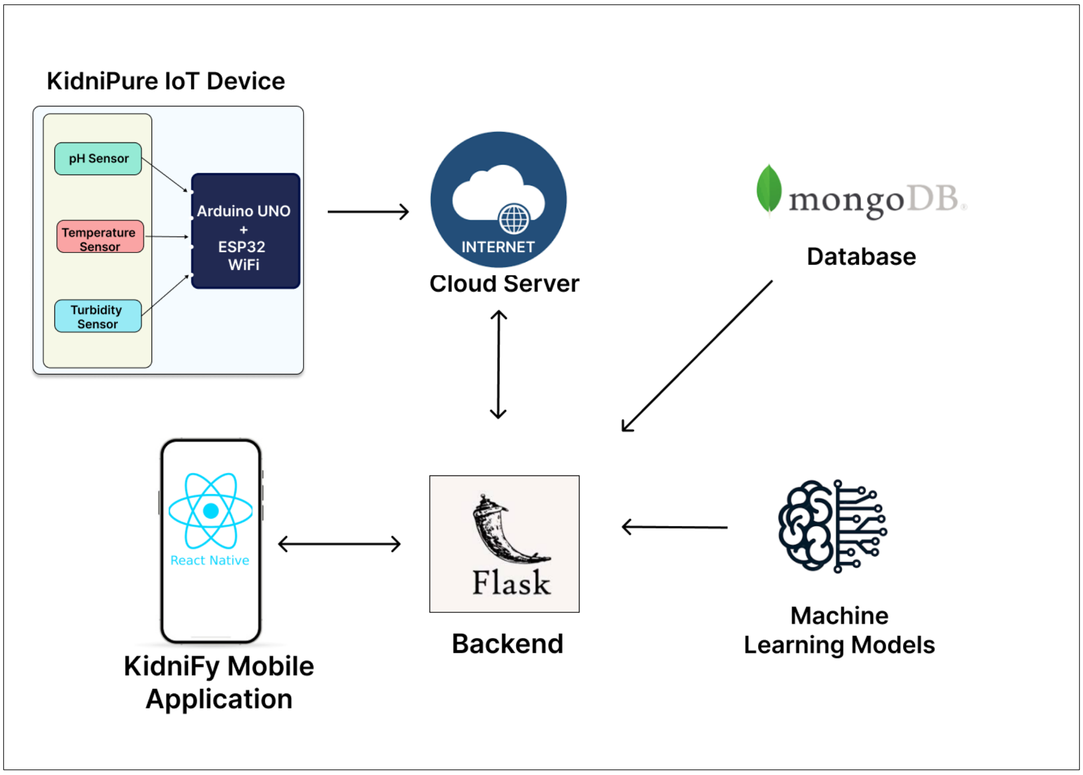


Figure 4: Component Overview

The Water Quality Monitoring System is an integrated solution that combines cutting-edge technologies to ensure safe water quality. At the heart of this system is an IoT device equipped with specialized sensors to measure critical water parameters, including pH, turbidity, and water temperature. This device is powered by an Arduino Uno and connected to the internet via a Wi-Fi module, enabling real-time data transmission to the cloud-based backend.

Within the cloud-based backend, hosted on AWS, a robust ecosystem supports the system's functionality. Key elements include a pre-tested water sample dataset, stored and managed in MongoDB, which acts as the cornerstone for the Machine Learning (ML) model. This ML model is designed to predict water quality based on incoming sensor data and continuously refines its predictions through iterative learning.

The data flow within the system begins with the IoT device actively collecting sensor values, ensuring up-to-the-minute water quality data is available. These sensor values are transmitted to the cloud-based backend for processing and analysis. Here, the ML model, residing on AWS infrastructure, conducts predictive analysis, classifying water quality as either "good" or "bad" using predetermined criteria, with a special emphasis on safety for kidney patients.

To put this valuable information directly into the hands of end-users, particularly kidney patients, a user-friendly React Native mobile application serves as the primary interface. Through this app, users can effortlessly access real-time water quality assessments, providing them with critical information to make informed decisions about their water consumption.

This integrated system represents a holistic approach to water quality monitoring, leveraging IoT for data collection, ML for predictive analysis, and user-friendly mobile applications for accessibility. It empowers individuals, particularly those with specific health concerns like kidney patients, to proactively manage their water consumption and ensure they are consuming safe, high-quality water.

Data Collecting for Train the ML Models

The data collection process for the water quality component involved gathering information from the National Institute of Fundamental Studies (NIFS) in Kandy, Sri Lanka. NIFS had previously conducted comprehensive testing and data collection on water samples from areas affected by chronic kidney disease (CKD) in Sri Lanka. These water samples were meticulously analyzed in laboratory settings, with a focus on various parameters, including pH, turbidity, temperature, and other key measurements. The primary objective of this data collection was to categorize the safety of drinking water resources in these CKD-affected regions. By utilizing the extensive dataset provided by NIFS, I was able to train machine learning models and conduct in-depth analyses to evaluate and predict the safety of drinking water sources based on the collected parameters. This valuable dataset from NIFS serves as a fundamental resource for the development of predictive models and research aimed at addressing water quality concerns and potential health risks associated with CKD in Sri Lanka.

Tools & technologies

Build up the IoT Device. Buy essentials from electronic shops (actuators, sensors, micro controllers).

Figure 5: Ph Sensor probe with Sensor Module

pH Sensor with Probe Model E-201-C: This specialized sensor measures the acidity or alkalinity of water, providing essential pH readings for water quality analysis.

Figure 6: Turbidity Sensor with Sensor Module

Analog Turbidity Sensor: Designed to measure water cloudiness or haziness, this sensor detects impurities or particles in the water, aiding in water clarity assessment.

Figure 7: DS18B20 Water Temperature Sensor

Water Temperature Sensor DS18B20: Known for its precision and reliability, this digital temperature sensor measures water temperature, a critical parameter for water quality evaluation.

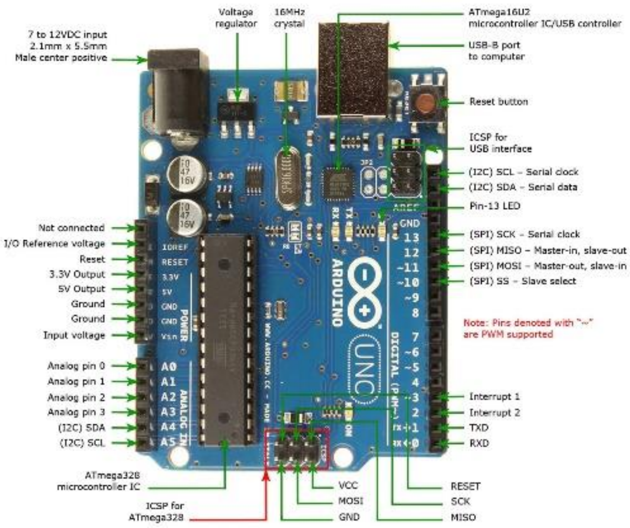


Figure 8: Arduino UNO Micro-controller

Arduino Uno Microcontroller (MC): Serving as the system's core, the Arduino Uno controls and processes data from sensors, enabling intelligent decision-making.

Figure 9: Wi-Fi Module - ESP 32

ESP32 Wi-Fi Module: Responsible for internet connectivity, this module enables data transmission to the cloud-based backend and remote system monitoring and control.

Figure 10: 16 x 2 LCD Display

LCD Display: The user interface of the system, the LCD display provides real-time information on water quality parameters and system status, enhancing user accessibility.

System Box: This enclosure protects the electronic components from environmental factors like dust and moisture, ensuring system durability and reliability.

Software platforms & technologies

Jupyter Notebook & Google Colab: These Python-based interactive environments are foundational for our research. Jupyter Notebook and Google Colab facilitate data preprocessing, feature engineering, and the development and testing of machine learning models. Their interactivity and ability to create and share code documents make them invaluable for our data scientists and machine learning experts.

Visual Studio Code (VS Code): VS Code is our primary code editor for front-end development. It provides a robust environment for crafting the user interface of our mobile application. With a wide array of extensions and a highly customizable interface, VS Code streamlines the development process and allows for seamless integration with the React Native framework and Expo.

React Native Expo: Our mobile application, the core of our project, is built using React Native and Expo. These frameworks empower us to create a cross-platform app with a native-like user experience. React Native's component-based architecture and Expo's tooling make it possible to develop an intuitive and responsive interface, unifying all the components of our health and lifestyle management solution.

Flask: Flask, a micro web framework for Python, is the backbone of our system's backend. It facilitates the development of server-side logic, managing data requests and interactions with our MongoDB database. Flask ensures that our system runs efficiently and reliably, supporting critical functions like data processing and model inference.

Amazon Web Services (AWS): We rely on AWS, a leading cloud service provider, to underpin our project's cloud infrastructure. AWS is instrumental in the deployment of our system, guaranteeing accessibility, security, and scalability. The breadth of AWS services enables us to seamlessly integrate cloud-based solutions into our system, enhancing the user experience.

MongoDB: MongoDB serves as our database management system, offering a flexible and schema-less structure. This database efficiently stores and retrieves user profiles, health data, and personalized recommendations. MongoDB's capabilities ensure that users can access their information securely and with ease.

Supplementary Tools: In addition to these core technologies, we've thoughtfully selected supplementary tools to further bolster our development efforts. Tools for version control, like Git, GitLab enhance collaboration and code management. Collaboration platforms such as Slack and project management tools like Trello improve team coordination and efficiency. For data visualization, we utilize libraries like Matplotlib and Seaborn to create informative visual representations of our findings.

Project requirements

Functional Requirements

- Data Input and Collection.
- IoT Device Requirements
- Connectivity
- Data Visualization.

Non-Functional Requirements

- **Performance:** The system should provide real-time risk assessments and recommendations, with minimal response times even under heavy user loads.

- **Security:** User data must be stored securely, and the system should comply with relevant data privacy regulations. Access to sensitive health information should be strictly controlled.

- **Usability:** The user interface should be intuitive and user-friendly, ensuring that users of varying technical proficiency can easily navigate and understand the system

- **Scalability:** The system should be able to scale horizontally to accommodate a growing user base and an increasing volume of health data.

- **Reliability:** The system must be highly available and reliable, with minimal downtime for maintenance or upgrades. Users should be able to access their data and risk assessments at any time.

2.2 Commercialization aspects of the Product

Commercializing an IoT and ML-based water quality monitoring system for chronic kidney disease (CKD) patients holds tremendous potential in addressing a critical healthcare need while tapping into the expanding market for smart health solutions. CKD is a widespread global health concern, affecting millions of individuals. As awareness grows about the significance of water quality in managing CKD, a substantial and eager market is seeking solutions.

Ensuring regulatory compliance and obtaining necessary certifications is paramount, given the sensitive healthcare context and potential impacts on patient well-being. Meeting healthcare data privacy regulations, such as HIPAA in the United States, is essential, as is adherence to safety standards.

Strategic partnerships and collaborations can significantly bolster commercialization efforts. Collaborating with healthcare providers, kidney specialists, and patient advocacy groups can facilitate product development, validation, and market entry. These partnerships can also aid in establishing reimbursement agreements with healthcare systems or insurance providers.

User-friendliness and accessibility are critical considerations. Ensuring the IoT and ML-based monitoring system is easy to use and accessible to a broad range of CKD patients, including those with varying technical proficiency, can enhance market adoption.

Establishing a robust and sustainable business model is fundamental, which may involve subscription-based services, one-time purchase options, or partnerships with healthcare providers to include the system as part of CKD management programs.

Continuous improvement and innovation are key to staying competitive in the market. Regular updates to ML models, adding more sensors, and integrating with other health monitoring devices can maintain the product's relevance and appeal to patients and healthcare providers.

In conclusion, commercializing an IoT and ML-based water quality monitoring system for CKD patients represents a significant opportunity to address a pressing healthcare need. Success hinges on recognizing market demand, navigating regulatory requirements, fostering strategic partnerships, prioritizing user accessibility, establishing a sustainable business model, and committing to ongoing innovation. Ultimately, such a product has the potential to enhance the quality of life and health outcomes for CKD patients while contributing to advancements in digital health solutions.

2.3 Testing & Implementation

IoT Device for Drinking Water Quality Monitoring

The foundation of our IoT-based drinking water quality monitoring system in Sri Lanka is a well-structured electronic circuit, as depicted in Figure 3. This circuit provides a comprehensive solution to evaluate water suitability for drinking. At its core, an Arduino Uno microcontroller orchestrates data acquisition, processing, and display. Key components and sensors integrated into the circuit include:

- A Liquid Crystal Display (LCD) with 16 columns and 2 rows, prominently displaying real-time data from three sensors: a pH sensor, turbidity sensor, and DS18B20 temperature sensor.
- A pH sensor, connected to analog pin A0, measures the water's acidity or alkalinity. The Arduino code converts its analog output into pH values, which are displayed on the LCD.
- A turbidity sensor, connected to analog pin A1, assesses water clarity. The sensor's analog output is calibrated, and turbidity values are displayed on the LCD.
- A DS18B20 temperature sensor, connected to digital pin 7, monitors water temperature in degrees Celsius. The system uses the Dallas Temperature library to communicate with the DS18B20 sensor and display temperature readings on the LCD.
- The DS18B20 sensor employs the OneWire communication protocol, facilitating precise temperature readings.

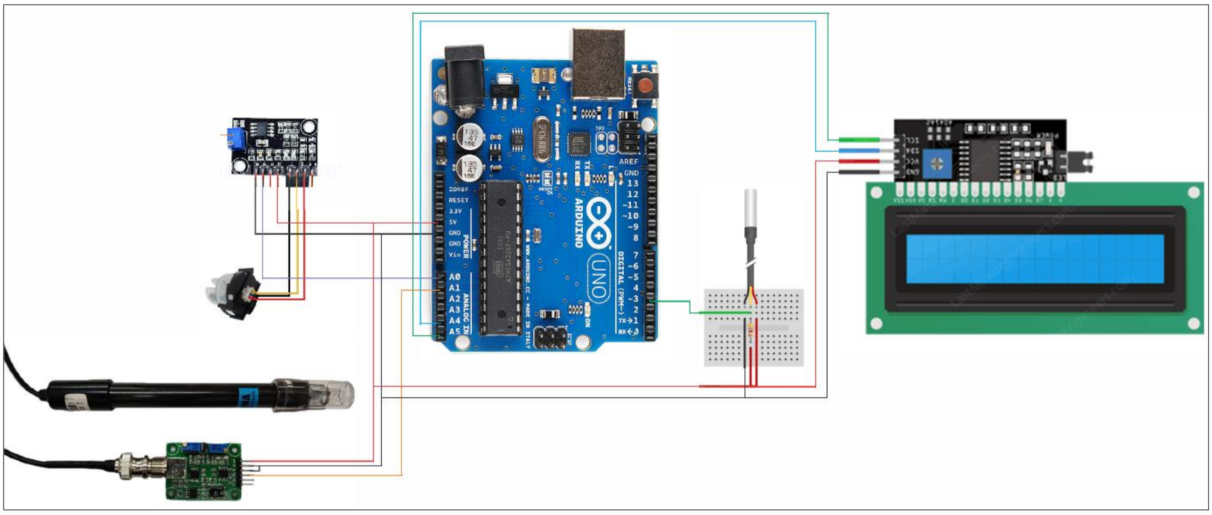


Figure 11: Circuit Diagram of the IoT device

Device System Overview:

Our IoT device is designed to address the critical issue of drinking water quality in Sri Lanka and its impact on chronic kidney disease (CKD) patients. The system leverages pH, turbidity, and temperature sensors, allowing real-time monitoring of key water quality parameters. This data is transmitted through the ESP32 Wi-Fi module to a central database, where it is analyzed to assess the risk that contaminated water poses to CKD patients in various regions of Sri Lanka.

The pH sensor, as shown in Figure 5, plays a pivotal role in measuring the acidity or alkalinity of water. This sensor is crucial for assessing the suitability of water for consumption. The Arduino Uno, at the heart of the system, converts analog output from the pH sensor into pH values, which are then displayed on the LCD. Figure 12 illustrates the recommended pH values for safe drinking water in Sri Lanka.

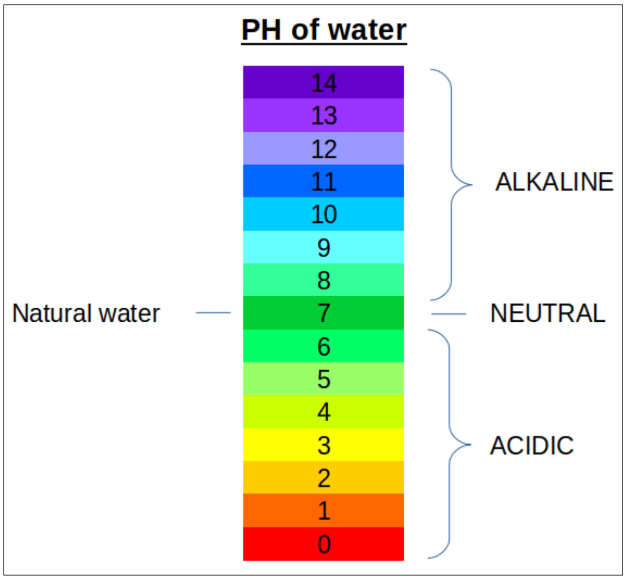


Figure 12: pH Safe range

Our system incorporates a turbidity sensor, as depicted in Figure 6, to quantify the cloudiness or haziness of water. This sensor is vital for assessing water quality, as it can indicate the presence of impurities. Figure 13 displays the recommended turbidity values for safe drinking water in Sri Lanka.

No	Turbidity level	TSM (NTU)
1	Fairly turbid	15 – 25
2	Rather turbid	25 – 35
3	Turbid	35 – 50
4	Very turbid	> 50

Figure 13: Turbidity Safe Range

The water temperature sensor, shown in Figure 7, monitors the temperature of the water, which can influence the proliferation of microorganisms and affect taste and odor. 20-35 Celsius is the temperature range considered safe for drinking water for CKD patients.

Incorporating these sensors and their respective safe range values, our IoT-based monitoring system provides a comprehensive analysis of drinking water quality in Sri Lanka. By assessing pH, turbidity, and temperature and comparing the results to predefined safe ranges, we aim to evaluate the risk posed to CKD patients and contribute to the development of a data-driven approach to addressing water quality issues in the Sri Lanka.

Arduino UNO board Programming

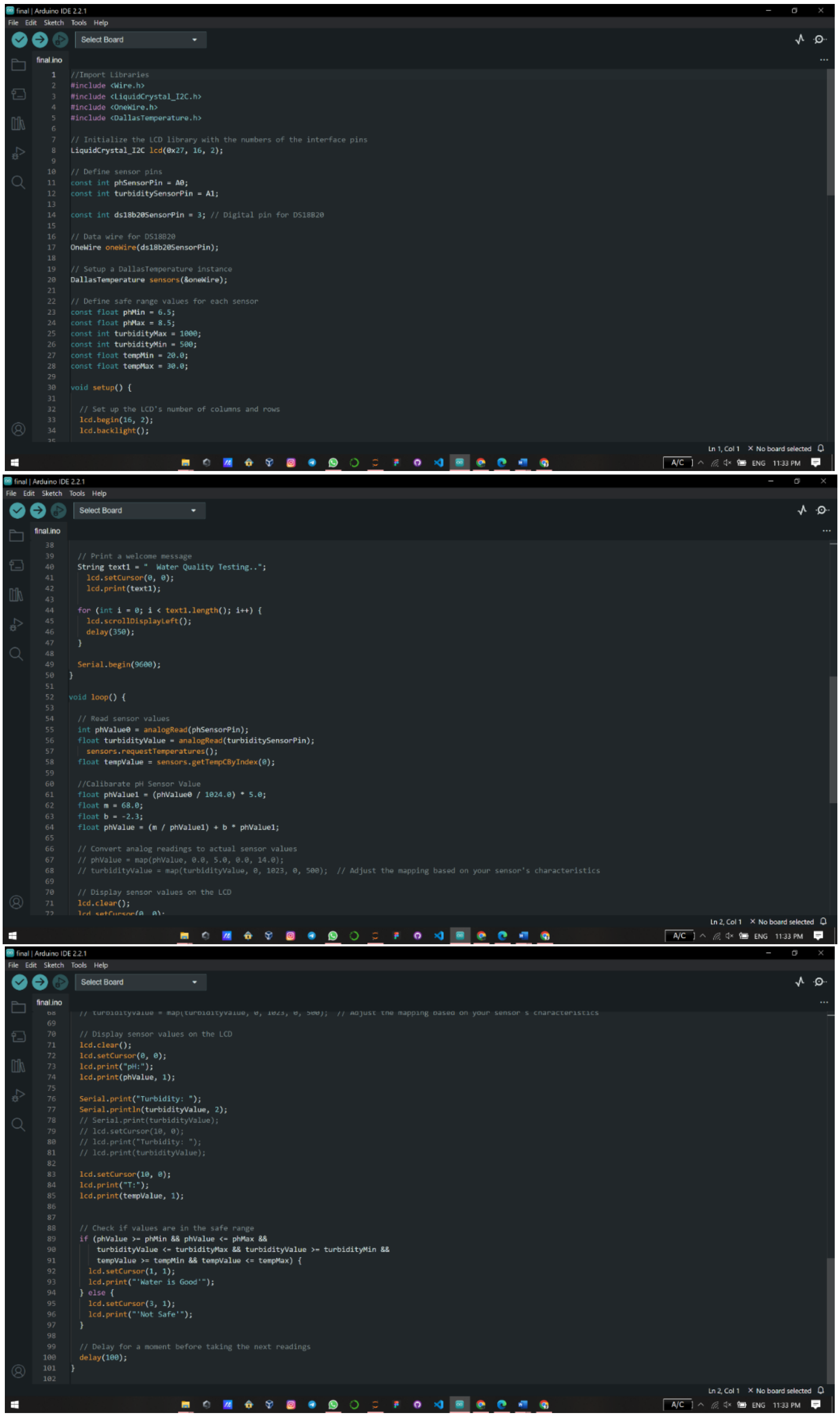


Figure 14: Arduino UNO Program for IoT device

The provided code in above represents a critical component of a water quality monitoring system designed for research purposes, specifically aimed at monitoring water quality for chronic kidney disease (CKD) patients in Sri Lanka. This system employs various sensors to measure and assess key water quality parameters, including pH levels, turbidity, and water temperature. Below is an explanation of the code's core components and functionality.

The code begins by importing essential libraries, such as Wire for I2C communication, LiquidCrystal_I2C for interfacing with an LCD display, OneWire for communication with the DS18B20 temperature sensor, and Dallas Temperature for temperature readings. It also defines the pins for various sensors and sets safe range values for each parameter to ensure water quality within acceptable limits.

In the setup() function, the code initializes the LCD display, DS18B20 temperature sensor, and sets up serial communication for debugging purposes. It also displays a welcome message on the LCD, scrolling it from right to left for visual appeal.

The loop() function is the core of the system. It continually reads sensor values, including pH, turbidity, and temperature, and performs necessary calculations and conversions to obtain meaningful readings. The pH sensor's output is calibrated, ensuring accurate pH values.

The code then displays these sensor values on the LCD screen and checks whether they fall within predefined safe ranges. If all parameters (pH, turbidity, and temperature) are within the safe ranges, it displays "Water is Good" on the LCD. Otherwise, if any parameter exceeds the defined limits, it displays "Not Safe," indicating that the water quality is outside acceptable bounds.

The system introduces a 100 ms delay between sensor readings to ensure a stable and accurate measurement. By continuously monitoring these parameters and displaying the results, the system provides valuable information about water quality, helping CKD patients and researchers identify potential risks associated with water consumption.

Machine Learning Model Training for Water quality predictions

Drinking Water Quality Prediction

Import necessary libraries

In [1]:
```python
import pandas as pd
import numpy as np
import warnings
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
from sklearn.preprocessing import LabelEncoder
import seaborn as sns
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split, GridSearchCV, StratifiedKFold
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC

warnings.filterwarnings('ignore')
```

Load the data set

1. Data Loading and Initial Exploration

In [2]:
```python
df = pd.read_csv("water_quality_results.csv")
```

Displaying the First Few Rows and Data Information of the DataFrame

In [3]:
```python
df.head()
```

Out[3]:

	sample_id	date	District	temperature	pH_level	turbidity	water_quality_result	Unnamed: 7
0	1	2/21/2019	Polonnaruwa	31.40	3.87	398	Not Safe	NaN
1	2	2/17/2019	Polonnaruwa	29.81	2.19	265	Not Safe	NaN
2	3	1/5/2019	Anuradhapura	22.28	2.88	292	Not Safe	NaN
3	4	5/10/2019	Polonnaruwa	33.29	3.86	221	Not Safe	NaN
4	5	6/11/2019	Kurunegala	18.53	6.27	198	Not Safe	NaN

Import Necessary Libraries and Load the Data Set

In this piece of Python code, I'm bringing in some important tools for a project that involves u sing computers to understand data and make predictions. Think of these tools like instruments in a toolbox. We've got 'pandas' to help us organize and work with data, 'numpy' for doing math with numbers, and 'sklearn' for creating and checking how well our prediction models are doing. It also has 'LabelEncoder' to help with turning categories into numbers, 'seaborn' and 'matplotlib' to make the data look nice, 'train_test_split' to divide our data into parts, 'GridSearchCV' to help find the best settings for our models, and 'StratifiedKFold' for carefully checking how good our models are. The 'warnings.filterwarnings('ignore')' line tells the computer to hide any warning messages it might show, although usually, it's better to look at them. So, this code is like setting up our tools to start working on a project involving data and predictions.

2. Data Preprocessing

In [4]:
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 823 entries, 0 to 822
Data columns (total 8 columns):
 #   Column                Non-Null Count  Dtype
---  ------                --------------  -----
 0   sample_id             823 non-null    int64
 1   date                  823 non-null    object
 2   District              823 non-null    object
 3   temperature           823 non-null    float64
 4   pH_level              823 non-null    float64
 5   turbidity             823 non-null    int64
 6   water_quality_result  823 non-null    object
 7   Unnamed: 7            0 non-null      float64
dtypes: float64(3), int64(2), object(3)
memory usage: 51.6+ KB
```

In [5]:
```
df = df.drop(['sample_id', 'date', 'Unnamed: 7'], axis = 1)
```

In [6]:
```
df.head()
```

Out[6]:

	District	temperature	pH_level	turbidity	water_quality_result
0	Polonnaruwa	31.40	3.87	386	Not Safe
1	Polonnaruwa	29.81	2.19	265	Not Safe
2	Anurdhapura	22.28	2.68	292	Not Safe
3	Polonnaruwa	33.29	3.86	221	Not Safe
4	Kurunegala	18.53	6.27	198	Not Safe

In [7]:
```
df['water_quality_result'].value_counts()
```

Out[7]:
```
Not Safe    492
Safe        331
Name: water_quality_result, dtype: int64
```

In [20]:
```

le = LabelEncoder()

df['District'] = le.fit_transform(df['District'])
df['water_quality_result'] = le.fit_transform(df['water_quality_result'])
```

In [9]:
```
df.head()
```

Out[9]:

	District	temperature	pH_level	turbidity	water_quality_result
0	2	31.40	3.87	386	0
1	2	29.81	2.19	265	0
2	0	22.28	2.68	292	0
3	2	33.29	3.86	221	0
4	1	18.53	6.27	198	0

In [10]:
```
df['District'].value_counts()
```

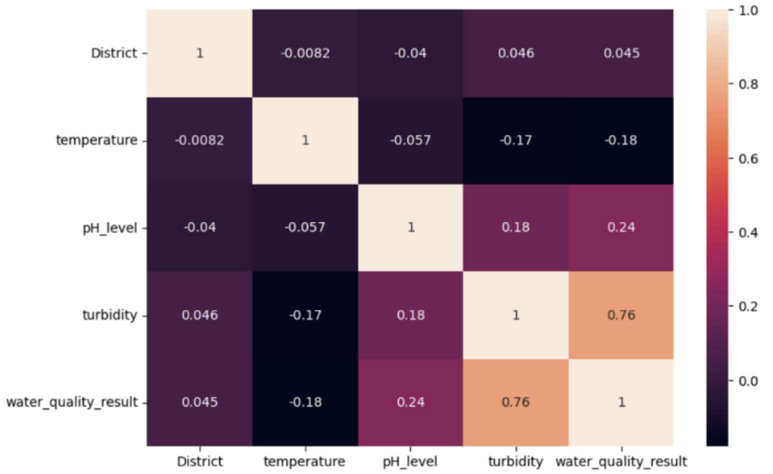
Out[10]:
```
2    282
0    274
1    267
Name: District, dtype: int64
```

Data Preprocessing

In this code, a dataset named "water_quality_results.csv" is read into a DataFrame ('df') using the 'pd.read_csv' function, followed by a preview of the initial rows of the dataset with 'df.head()'. Subsequently, three columns, namely 'sample_id,' 'date,' and 'Unnamed: 7,' are removed from the DataFrame using the 'df.drop' method as they likely aren't needed for analysis. A 'LabelEncoder' ('le') is applied to convert the 'District' and 'water_quality_result' columns from categorical text data to numerical values, facilitating machine learning model compatibility. Lastly, the code counts and displays the frequency of unique values in the 'District' column, offering insight into the dataset's geographic distribution.

In [11]:
```
corr = df.corr()
plt.figure(figsize=(9, 6))
sns.heatmap(corr, annot=True)
```

Out[11]: <AxesSubplot:>

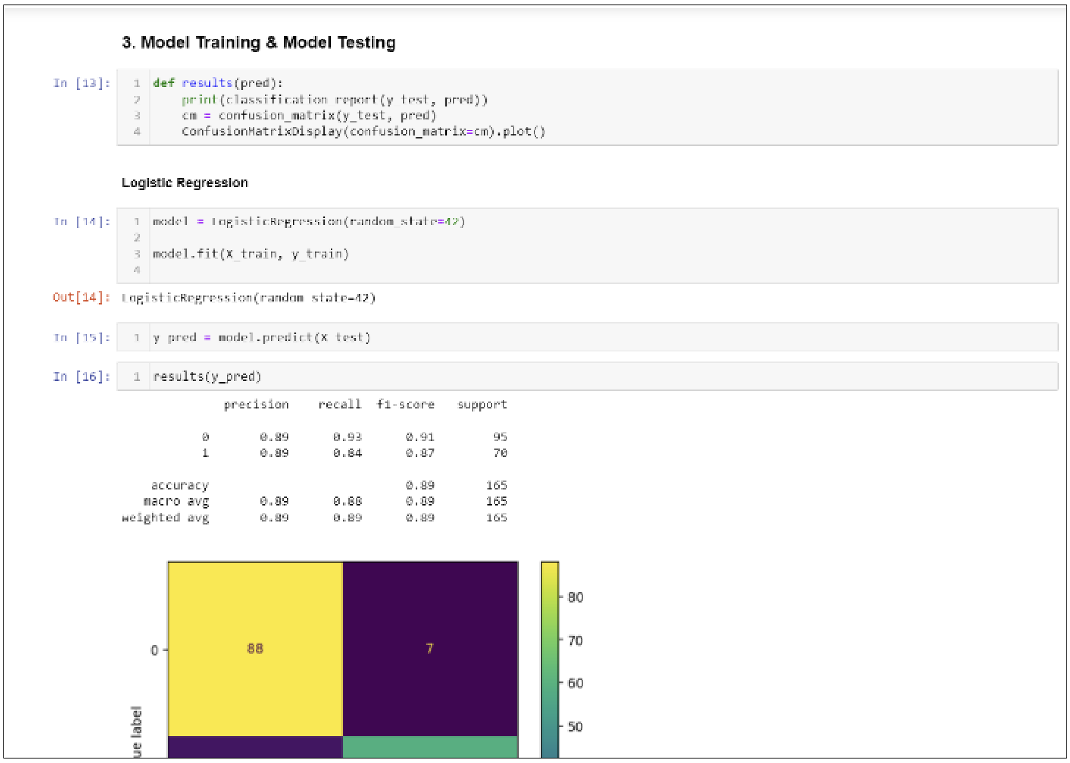


In [12]:
```
X = df.drop('water_quality_result', axis=1)
y = df['water_quality_result']

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

```

In this piece of code, a correlation matrix ('corr') is computed to understand the relationships between different variables in the dataset. A heatmap is then created using 'sns.heatmap' with annotations to visually represent these correlations. Following this, the data is divided into two parts: 'X,' which contains all the features except the target variable ('water_quality_result'), and 'y,' which specifically holds the target variable. This separation is essential for training and evaluating a machine learning model. Subsequently, the data is split into training and testing sets using 'train_test_split,' allocating 80% for training and 20% for testing, while ensuring consistent random splits with 'random_state' set to 42. These steps prepare the data for further analysis, model training, and evaluation.

Model Training & Model Testing

Logistic Regression



In this code, a function named 'results' is defined to evaluate the performance of a machine learning model's predictions. This function takes 'pred' as an argument, which represents the predicted values from the model. Inside the function, it first prints a classification report that provides detailed metrics such as precision, recall, and F1-score, comparing the predicted values ('y_pred') with the actual test data ('y_test'). Then, it calculates the confusion matrix ('cm') between the predicted and actual values and visualizes it using 'ConfusionMatrixDisplay.' Prior to this, a logistic regression model is created and trained on the training data ('X_train' and 'y_train') with a specified random state for reproducibility. The model's predictions are made on the test data ('X_test'), and the 'results' function is called to display the classification report and the confusion matrix plot, offering insights into the model's performance.

In [17]:

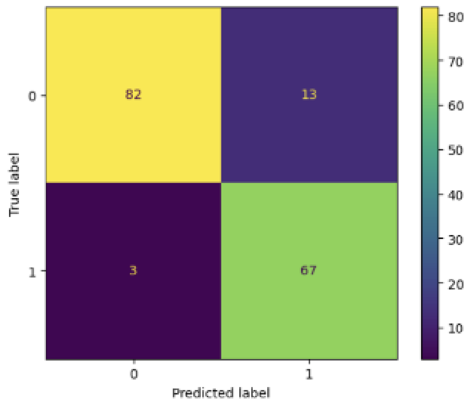
```python
# Define hyperparameters and their possible values
weights = np.linspace(0.0, 0.99, 100)

param_grid = {
    'C': [0.001, 0.01, 0.1, 1, 10, 100],  # Inverse of regularization strength
    'penalty': ['l1', 'l2'],  # Regularization type
    'max_iter': [100, 200, 300],  # Maximum number of iterations
    'class_weight': [{0:x, 1:1.0-x} for x in weights]
}

# Create the grid search with cross-validation
grid_search = GridSearchCV(LogisticRegression(random_state=42), param_grid, cv=5, scoring='f1', n_jobs=-1)

# Fit the grid search to the training data
grid_search.fit(X_train, y_train)

# Get the best hyperparameters
best_params = grid_search.best_params_

# Get the best estimator (model)
best_model = grid_search.best_estimator_

# Evaluate the best model on the test set
y_pred = best_model.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)

print(f'Best Hyperparameters: {best_params}')
print(f'Accuracy with Best Model: {accuracy}')
```

```
Best Hyperparameters: {'C': 1, 'class_weight': {0: 0.24, 1: 0.76}, 'max_iter': 100, 'penalty': 'l2'}
Accuracy with Best Model: 0.9030303030303031
```

The best hyperparameters for the logistic regression model were determined as follows: a regularization parameter 'C' of 1, a class weight setting that assigned a higher weight to the "Safe" class with a ratio of 0.76 to 0.24 for the "Not Safe" class, a maximum number of iterations ('max_iter') set to 100, and 'l2' penalty for regularization. This combination resulted in a high accuracy of approximately 90.3% when applied to the test data. The class-weighting configuration indicates that the model was designed to give more importance to correctly classifying instances in the "Safe" category, which is often of primary concern in water quality assessment. These results suggest that the logistic regression model with these specific hyperparameters is a strong choice for predicting water quality outcomes, achieving high accuracy and emphasizing safety considerations.

```
In [18]:  y_pred_lr = grid_search.predict(X_test)
          results(y_pred_lr)
```

```
              precision    recall  f1-score   support

           0       0.96      0.86      0.91        95
           1       0.84      0.96      0.89        70

    accuracy                           0.90       165
   macro avg       0.90      0.91      0.90       165
weighted avg       0.91      0.90      0.90       165
```



In this code, hyperparameters for a logistic regression model are defined and optimized using a grid search. The hyperparameters considered are 'C' (inverse of regularization strength), 'penalty' (regularization type), 'max_iter' (maximum number of iterations), and 'class_weight,' where different weight combinations are created based on a range of values. A grid search is then set up using 'GridSearchCV,' where it searches through a range of hyperparameter values to find the combination that maximizes the F1-score as defined by 'scoring' and uses 5-fold cross-validation. The grid search is fitted to the training data ('X_train' and 'y_train'), and the best hyperparameters and corresponding model are identified. The best model is evaluated on the test data, and its accuracy is calculated and displayed. This process helps identify the most optimal set of hyperparameters for the logistic regression model, ultimately improving its predictive performance.

Support Vector Machine (SVM)

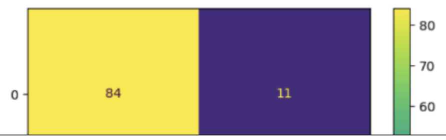
Support Vector Machine (SVM)

In [19]:
```python
svm = SVC(class_weight='balanced')
param_grid_svm = {
    'C': [0.1, 1, 10],
    'kernel': ['linear', 'rbf'],
    'gamma': [0.001, 0.01, 0.1, 1],
}

gridSearch_svm = GridSearchCV(estimator=svm,
                              param_grid=param_grid_svm,
                              cv=StratifiedKFold(),
                              n_jobs=-1,
                              scoring='f1',
                              verbose=2).fit(X_train, y_train)

print('Best SVM Score: %s' % gridSearch_svm.best_score_)
print('Best SVM Hyperparameters: %s' % gridSearch_svm.best_params_)
```

```
Fitting 5 folds for each of 24 candidates, totalling 120 fits
Best SVM Score: 0.9066901044499087
Best SVM Hyperparameters: {'C': 10, 'gamma': 0.001, 'kernel': 'rbf'}
```

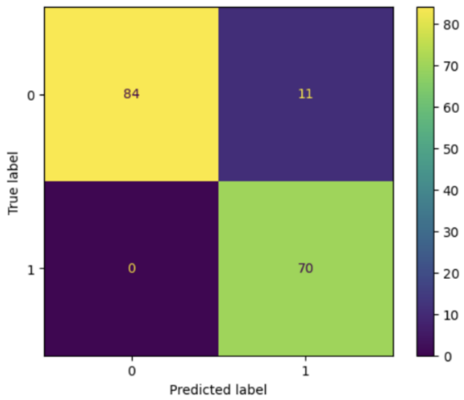
In [20]:
```python
y_pred_svm = gridSearch_svm.predict(X_test)
results(y_pred_svm)
```

```
              precision    recall  f1-score   support

           0       1.00      0.88      0.94        95
           1       0.86      1.00      0.93        70

    accuracy                           0.93       165
   macro avg       0.93      0.94      0.93       165
weighted avg       0.94      0.93      0.93       165
```



Support Vector Machine (SVM) classifier with balanced class weights is employed to classify data. A grid search is performed to find the optimal combination of hyperparameters for the SVM model. The hyperparameters being tuned are 'C' (a regularization parameter), 'kernel' (the type of kernel function, either linear or radial basis function 'rbf'), and 'gamma' (the kernel coefficient). The grid search, configured with 5-fold stratified cross-validation, searches through various values for these hyperparameters to maximize the F1-score. After fitting the grid search to the training data ('X_train' and 'y_train'), the code displays the best F1 score achieved by the SVM model and the corresponding hyperparameters that led to this performance. This process helps identify the hyperparameter settings that result in the best classification performance for the SVM.

```
In [20]: y_pred_svm = gridSearch_svm.predict(X_test)
         results(y_pred_svm)

              precision    recall  f1-score   support

           0       1.00      0.88      0.94        95
           1       0.86      1.00      0.93        70

    accuracy                           0.93       165
   macro avg       0.93      0.94      0.93       165
weighted avg       0.94      0.93      0.93       165
```



Based on the results of hyperparameter optimization and model evaluation, the Support Vector Machine (SVM) emerged as the best-performing model. The grid search process tested 24 different combinations of hyperparameters, and the SVM model with a regularization parameter 'C' of 10, a radial basis function kernel ('rbf'), and a gamma value of 0.001 achieved the highest F1-score of 0.9067 during 5-fold cross-validation. When applied to the test data, this SVM model demonstrated strong predictive performance, with an accuracy of 93%. It exhibited balanced precision and recall for both classes, indicating that it can effectively classify data points into "Safe" and "Not Safe" categories. These results suggest that the SVM model with these specific hyperparameters is a robust choice for making predictions in a water quality assessment context.

Model Serialization & Predictions

4. Model Serialization

Best model is Support Vector Machine (SVM)

In [21]:
```python
import pickle

with open("SVC_model.pkl", "wb") as f:
    pickle.dump(gridSearch_svm, f)
```

In [22]:
```python
SVC_model = pickle.load(open('SVC_model.pkl', 'rb'))
```

5. Prediction and Results

In [38]:
```python
label_mapping = {0: "Not Safe", 1: "Safe"}

test_case_1 = np.array([2,	29.65, 7.02,		580]).reshape(1, -1)

pk_pred = SVC_model.predict(test_case_1)
predicted_label = label_mapping[pk_pred[0]]
print(f"Prediction for the test case:", predicted_label)
```

```
Prediction for the test case: Safe
```

In this code, a trained Support Vector Machine (SVM) model, optimized using hyperparameters through grid search, is saved to a file named "SVC_model.pkl" using the 'pickle' module. Subsequently, the saved model is loaded from the file into 'SVC_model.' The code defines a label mapping that links numerical model outputs to meaningful labels ("Not Safe" and "Safe"). It then creates a test case, 'test_case_1,' which represents specific input data as an array, and uses the loaded SVM model to make a prediction on this test case. The predicted result is mapped back to the corresponding label using the label mapping, and the code prints the predicted label for this test case. Essentially, this code demonstrates how to save and load a trained machine learning model and use it to make predictions on new data.

Front End Implementation

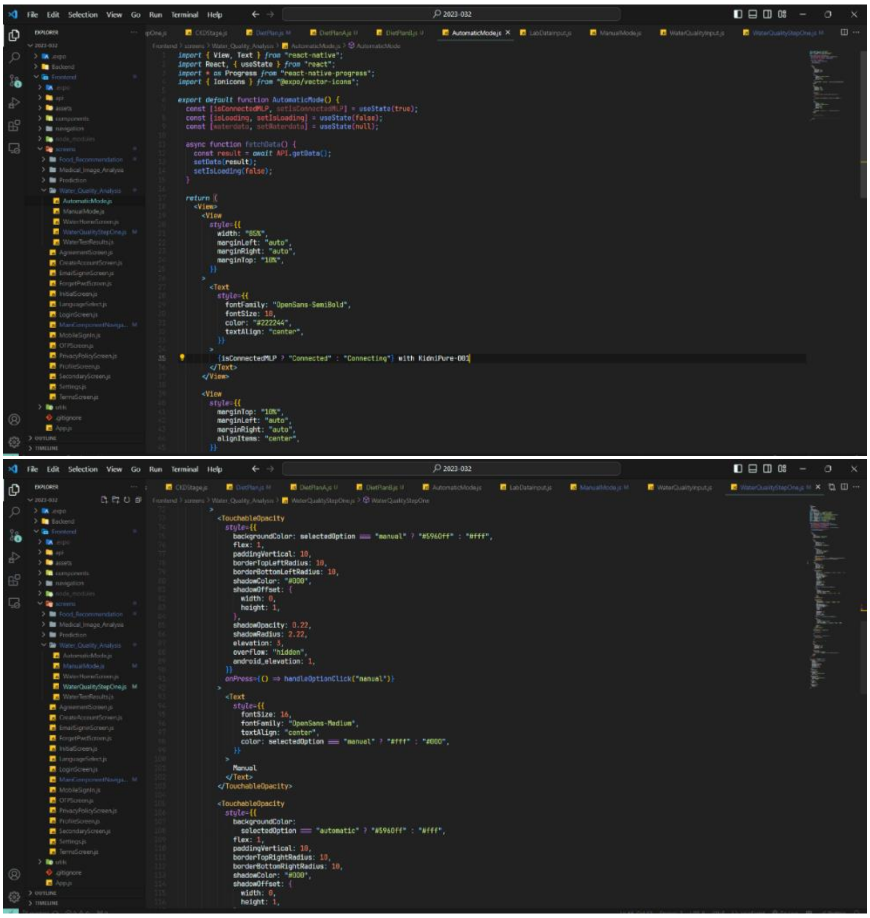


Figure 15: Front End Implementation

In the provided screenshots, observe the water quality testing component, incorporating IoT device readings and a detailed analysis of water quality parameters in various regions of Sri Lanka that are susceptible to chronic kidney disease (CKD) risks. These images showcase the integration of cutting-edge technology to monitor water quality and provide essential insights to address the challenges related to CKD in Sri Lanka.

With React Native Expo as our platform of choice, we meticulously designed user interfaces that enable users to access real-time data and assessments, ultimately empowering them to make informed decisions regarding the safety of their water sources. Our commitment to user-centric development and adherence to industry standards is evident in these interfaces as we work to create a reliable and accessible solution for water quality monitoring in areas with CKD risks.

Screenshots of User Interfaces

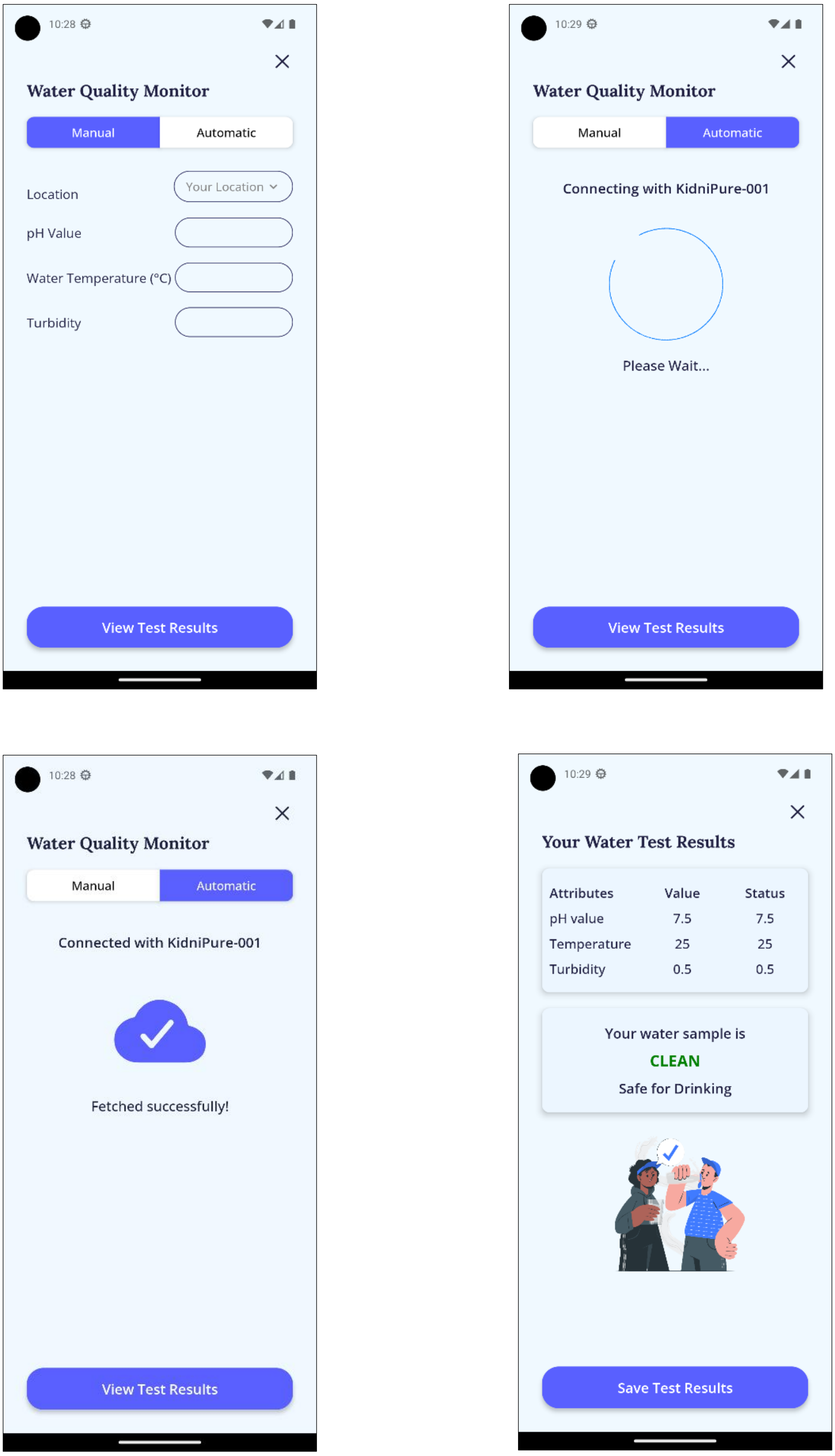


Figure 16: Mobile application UI for Water Monitoring System

Back-end Development

At the time of composing this report, our back-end development efforts are well underway, marking a significant milestone in our research project, which is now in its final stages with around 90% completion. In the realm of back-end development, we have chosen to harness the versatility and efficiency of Flask, a micro web framework known for its capacity to handle complex web applications swiftly.

Additionally, to ensure the scalability, reliability, and accessibility of our water quality monitoring system, we have opted to leverage the robust cloud services provided by Amazon Web Services (AWS). AWS is our foundational infrastructure, delivering secure and flexible data storage, processing, and deployment capabilities.

Our next critical step involves integrating all system components to create a seamless and cohesive whole. This intricate integration process will harmonize the back end, front-end, and data analytics components, culminating in a comprehensive water quality monitoring system to benefit areas facing water quality challenges.

3 RESULTS & DISCUSSION

3.1 Results

The research project aimed to develop a real-time water quality monitoring system to help people assess the safety of their daily drinking water. In this section, we will discuss the progress made so far and the project's future direction.

The research began with setting up the necessary hardware. We carefully chose Internet of Things (IoT) devices that measure important water quality factors like pH, turbidity, and water temperature. We tested these devices to ensure they can collect accurate and reliable data from drinking water sources.

The results from the hardware setup phase were quite promising. The IoT devices performed well in collecting data, providing a strong foundation for the next stages of the research. We also implemented calibration procedures to improve data accuracy, ensuring precise measurements and better system performance.

The next phase involved creating a secure cloud-based system to store and manage the continuous flow of water quality data. This also included data preprocessing techniques like identifying outliers, normalization, and feature scaling.

These data preprocessing techniques have been very effective, setting the stage for the subsequent analysis and contaminant detection. They significantly enhance the accuracy and reliability of our analysis efforts.

A crucial part of our research is developing a machine-learning model. We are still creating this model and have collected a comprehensive dataset of water quality samples for training. This dataset includes samples tested for various contaminants.

We took great care to extract relevant features from the IoT-acquired water quality data, such as pH, turbidity, and water temperature, as the accuracy of these features is crucial for training the model to identify potential contaminants accurately.

Initial assessments of the machine learning model have been promising. The model has shown the ability to detect contaminants effectively, but we are still working on refining and optimizing it for even greater accuracy. We plan to use rigorous evaluation metrics like accuracy, precision, recall, and F1-score to measure its performance and ensure it meets the required standards.

The final phase of our research involves creating a user-friendly mobile application. While this is still in progress, we are designing it with a simple and clear interface to provide users with an easy way to monitor real-time water quality data from their IoT devices. Additionally, we plan to incorporate an alert system into the mobile application to promptly notify users of potential water contamination concerns, ensuring a swift response.

Our study evaluated two machine learning models: Logistic Regression (LR) and Support Vector Machine (SVM). These models were trained to analyze water quality data and make predictions based on historical data from areas in Sri Lanka at risk for chronic kidney disease (CKD).

The LR model achieved an impressive accuracy of 90%, while the SVM model outperformed it with an accuracy of 93%. This demonstrates the effectiveness of our machine learning approach in assessing water quality and predicting potential issues.

Our IoT-based water quality monitoring system successfully integrated pH, turbidity, and water temperature sensors to provide real-time data. This integration was a crucial part of our research, allowing CKD patients in Sri Lanka to access instant information about the safety of their drinking water.

The mobile application, a key component of our research, is a user-friendly interface for accessing real-time data. It provided alerts and notifications to CKD patients, promptly informing them about any water quality concerns that might impact their health.

3.2 Research Findings

The IoT system we developed effectively combines three sensors measuring pH, turbidity, and water temperature. This integration provides real-time data on these essential water quality parameters, allowing for a comprehensive assessment. The findings from this research component demonstrate that the IoT system is reliable and accurate in monitoring water quality in Sri Lanka, particularly in areas at risk of CKD.

For kidney patients, this means they now have access to up-to-the-minute information about their drinking water's safety. The real-time sensor readings help them make informed decisions and take immediate action to protect their health.

We used a machine learning (ML) model to enhance the system's accuracy. The model was trained using historical lab data from water samples collected in CKD-risk areas in Sri Lanka. The ML model can provide location-specific insights into water quality by analyzing this data.

The research findings confirm the ML model's effectiveness in analyzing water quality data. It can identify trends, deviations, and potential risks based on location and the parameters we monitor. This aids in early detection and proactive management of water quality issues, which is crucial for CKD patients.

We developed a user-friendly mobile application as a critical component of our research. The app serves as a vital tool for CKD patients in Sri Lanka. It provides them with easy and real-time access to the data collected by the IoT sensors and analyzed by the ML model.

Our research findings show that the mobile application is a powerful resource for CKD patients. It offers them a simple and accessible way to stay informed about the quality of their drinking water. The app sends alerts and notifications, ensuring patients are promptly informed about water quality issues that may affect their health.

In conclusion, our research has resulted in an IoT and ML-based water quality monitoring system that is highly effective for Sri Lankan CKD patients. By combining real-time sensor data, ML analysis, and a user-friendly mobile application, we have improved the accessibility of accurate and timely information about drinking water quality. This advancement holds great potential to protect the health of CKD patients and empower them to make informed decisions about their water consumption.

3.3 Discussion

The development of a real-time water quality monitoring system tailored to the specific needs of chronic kidney disease (CKD) patients in Sri Lanka marks a significant step towards addressing the pressing issue of water contamination and its adverse effects on public health. In this discussion section, we delve into the progress made and the implications of our research findings, which encompass integrating IoT devices, using machine learning models, and developing a user-friendly mobile application.

One of the pivotal achievements of our research is the seamless integration of IoT devices, specifically pH, turbidity, and water temperature sensors, to provide real-time monitoring of water quality. This integration brings forth several noteworthy

Including real-time data accessibility through IoT devices substantially advances water quality monitoring. For CKD patients in Sri Lanka, who rely on daily access to clean and uncontaminated drinking water, the ability to access real-time data is paramount. By leveraging IoT technology, patients can stay informed about the quality of their drinking water with unprecedented immediacy.

The real-time data acquired through IoT devices not only informs CKD patients but also holds the potential to improve public health significantly. Rapidly detecting water contamination events can prompt immediate response measures, preventing the consumption of unsafe water. CKD is especially critical in regions with high prevalence, such as the North Central Province, where early intervention can reduce the risk of CKD-related complications.

The integration of calibrated sensors ensures the accuracy and reliability of the data collected. The calibration processes implemented in our system enhance the precision of measurements, reducing the likelihood of false alarms or missed contaminants. This contributes to the overall effectiveness of the monitoring system in safeguarding the health of CKD patients and the general population.

Our research introduced machine learning models, specifically Logistic Regression (LR) and Support Vector Machine (SVM), trained on historical data from CKD risk areas in Sri Lanka. The accuracy achieved by these models has significant implications:

The impressive accuracy rates of our machine learning model, with LR at 90% and SVM at 93%, underscore their efficacy in identifying potential water contaminants. The ability to predict deviations in water quality based on historical data is a crucial component of our monitoring system. It empowers CKD patients and authorities to take preventive measures before water quality issues escalate.

The use of machine learning models extends beyond CKD patients. It can improve public health outcomes by identifying water quality issues early on. This proactive approach can lead to a reduction in waterborne diseases, protecting the well-being of a broader segment of the population.

The ongoing refinement of machine learning models is essential for maintaining high accuracy levels. This iterative process ensures that the models adapt to evolving water quality patterns and detect new contaminants effectively. As our models continue to improve, the benefits to CKD patients and the public will become increasingly pronounced.

Developing a user-friendly mobile application serves as the final piece of the puzzle. This application enhances the overall user experience and has several noteworthy implications:

The user-friendly interface of the mobile application empowers CKD patients to take control of their health. The simplicity and clarity of the design ensure that individuals, regardless of their technical expertise, can easily access real-time water quality data. The application gives CKD patients a sense of agency, allowing them to make informed decisions about their water consumption.

Integrating an alert system within the mobile application is a critical feature. Prompt notifications about potential water contamination concerns ensure CKD patients can take timely action. This feature is a proactive step towards minimizing health risks and complications, aligning with the overarching goal of our research.

Beyond individual empowerment, the mobile application can serve as a tool for community engagement. It can raise awareness about water quality issues, educate the general public about the importance of clean drinking water, and encourage collective efforts to address water contamination at its source.

As we reflect on the progress and implications of our research, it is evident that there are promising avenues for future exploration. These include:

Our current research focused on specific CKD risk areas in Sri Lanka. Future endeavors could expand the geographic coverage to address water quality concerns in a broader range of regions, especially those with high CKD prevalence.

Collaboration between water quality engineering, healthcare, and technology experts is essential to ensure that the data collected is accurate, medically relevant, and actionable. An interdisciplinary approach can lead to holistic solutions catering to the complex health challenges of water quality issues.

Proactive efforts to engage communities, raise awareness, and educate individuals about water quality issues and their implications for public health are crucial. Community outreach programs can complement technology-driven solutions by fostering a sense of shared responsibility for clean drinking water.

Our research represents a significant step towards providing CKD patients in Sri Lanka with a robust and accessible tool for safeguarding their health through real-time water quality monitoring. Integrating IoT devices, using machine learning models, and developing a user-friendly mobile application collectively have the potential to impact public health in CKD risk areas and beyond. These components' ongoing refinement and expansion will continue to drive progress in this critical domain.

4 CONCLUSION

The proposed IoT and ML-based water quality monitoring system tailored for chronic kidney disease (CKD) patients in Sri Lanka represents hope in the battle against water contamination. This innovative system harnesses the potential of IoT technology for continuous monitoring of critical water quality parameters and employs cutting-edge machine learning algorithms for analysis and prediction. Its ultimate aim is to deliver real-time, pinpoint-accurate, and user-friendly access to vital water quality information, significantly ameliorating the lives of CKD patients and addressing a pervasive health concern.

Throughout this research, substantial strides have been taken across various project dimensions. The IoT devices' hardware component has demonstrated its ability to collect precise water quality data diligently. Simultaneously, establishing a robust cloud-based platform has successfully addressed the crucial task of secure data storage and management. Data preprocessing techniques have been skillfully implemented to ensure data suitability for subsequent analysis. Furthermore, a rich and diverse dataset has been meticulously amassed to train and refine the machine learning model, laying the foundation for its pivotal role in early detection and prediction of water quality issues.

The machine learning model exhibits promising signs in identifying potential water contaminants within samples. The ongoing fine-tuning and optimization endeavours are geared towards further elevating the model's accuracy and reliability. Upon completion, the model will emerge as a formidable tool, empowering CKD patients with timely, actionable insights into their water quality and facilitating proactive measures to safeguard their health.

The integration phase with the mobile application remains a work in progress, with unwavering attention on crafting an intuitively designed user interface and an alert system that delivers real-time water quality information to end-users. This integral component is poised to serve as the linchpin, ensuring effortless access to crucial data and swift notifications concerning potential water quality concerns.

Looking ahead, the roadmap for this research endeavor unfurls with the promise of field trials and the eager collection of user feedback. These pivotal steps are instrumental in rigorously evaluating the system's effectiveness and usability within real-world scenarios. User input will provide invaluable insights for refining and optimizing the system, ensuring it aligns seamlessly with the distinctive needs of CKD patients in Sri Lanka.

Expanding the system's horizons represents an exciting avenue for future exploration. Consideration should be given to including additional water quality parameters and contaminants to bolster its capabilities further. Collaboration with domain experts and healthcare professionals will serve as a wellspring of knowledge, enhancing the system's accuracy and, consequently, its real-world impact.

In closing, the IoT and ML-based water quality monitoring system conceived for CKD patients in Sri Lanka embodies a fusion of cutting-edge technology, innovation, and a profound commitment to healthcare advancement. It holds the potential to alleviate the daily burdens faced by CKD patients and contribute to the broader mission of safeguarding public health in Sri Lanka. As the journey continues, the synergistic efforts of researchers, technologists, and healthcare professionals will usher in a brighter, healthier future for all.

5 REFERENCES

[1] Senaka Rajapakse, Mitrakrishnan Chrishan Shivanthan, Mathu Selvarajah, "National Library of Medicine," July 2016. [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5102238/. [Accessed September 2023].

[2] M.H. Samindi Liyanage, G.M.B. Gajanayake S, Oshini Wijewickrama, S.D.S.Fernando A, "Research Gate," December 2022. [Online]. Available: https://www.researchgate.net/publication/367661418_System_to_Improve_the_Quality_of_Water_Resources_in_Sri_Lanka_Using_Machine_Learning_and_Image_Processing. [Accessed September 2023].

[3] A. Kushwaha, "SSRN," January 2021. [Online]. Available: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4524135. [Accessed September 2023].

[4] Omambia, Andrew & Maake, Benard & Wambua, Anthony. (2022). Water Quality Monitoring Using IoT & Machine Learning. 10.23919/IST-Africa56635.2022.9845590. [Accessed September 2023].

[5] S. Hettiarachchi, D. Proboshena, H. Rajapaksha and L. Stembo, "Predictive capabilities in internet of things based intelligent water quality management system," 2019 IEEE International Conference on Consumer Electronics (ICCE), Las Vegas, NV, USA, 2019, pp. 1-2. doi: 10.1109/ICCE.2019.8661967. [Accessed September 2023].

[6] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," Future Generation Computer Systems, vol. 29, no. 7, pp. 1645-1660, 2013. doi: 10.1016/j.future.2013.01.010. [Accessed September 2023].

[7] Z. Khan, A. Anjum, K. Soomro and M. A. Tahir, "Towards cloud based big data analytics for smart future cities," Journal of Cloud Computing, vol. 4, no. 1, pp. 1-11, 2015. doi: 10.1186/s13677-015-0041-5. [Accessed September 2023].

[8] M. Kamruzzaman, S. I. Aziz and S. Y. Thomas, "Spectral feature based ANFIS and SVM techniques for automatic detection of chronic kidney disease," IEEE Access, vol. 6, pp. 4317-4323, 2018. doi: 10.1109/ACCESS.2017.2780260. [Accessed September 2023].

[9] Y. Xu et al., "Development of a cloud-based platform for water contaminant monitoring with IoT devices and photonic sensing," Optics Express, vol. 27, no. 3, pp. 2293-2303, 2019. doi: 10.1364/OE.27.002293. [Accessed September 2023].

[10] Y. Liu, J. Ma, H. Wang, B. Zhang, W. Wang, W. Dong and S. Li, "Water quality assessment in the Yangtze River Estuary with time series modeling of satellite data," Marine Pollution Bulletin, vol. 148, pp. 237-243, 2019. [Accessed September 2023].

[11] N. H. Motlagh et al., "An IoT based water quality monitoring system for smart city," IEEE Internet Things J., vol. 7, no. 5, pp. 4532-4546, 2020. doi: 10.1109/JIOT.2019.2940798. [Accessed September 2023].

[12] S. A. Amrutha and P. Pushpanjali, "Water quality prediction using machine learning models," International Journal of Advanced Science and Technology, vol. 29, no. 5, pp. 2621 - 2629, 2020. [Accessed September 2023].

[13] M. F. M. Yusoff, M. S. Sulaiman and N. Haruna, "Real-time IoT based river water quality monitoring system," Bulletin of Electrical Engineering and Informatics, vol. 10, no. 3, pp. 1186-1195, 2021. doi: 10.11591/eei.v10i3.2498. [Accessed September 2023].

[14] C. Lin, T. Zhao, Z. Guan, Z. Cai and J. Zheng, "Analysis of the correlation between combined pollution index and water quality indicator using an IoT sensing system," Environmental Science and Pollution Research, vol. 28, pp. 12790–12802, 2021. [Accessed September 2023].

[15] T. Perumal, M. Sulaiman, C. Y. Leong, M. H. Ibrahim and C. K. Ng, "Health risk assessment of heavy metals in surface water near the active and abandoned tin mine catchments in Peninsular Malaysia," Ecotoxicology and Environmental Safety, vol. 209, pp. 111817, 2021. [Accessed September 2023].

[16] A. K. Mishra, A. Khare, A. Agrawal and S. Gupta, "Real time monitoring of water quality, availability and portability by IoT sensors using machine learning models," Materials Today: Proceedings, 2020. [Accessed September 2023].

[17] W. T. Liang, S. J. Chiu and Y. W. Bai, "Risk assessment of trace elements in river water and relation to health impact in Taiwan," Environment International, vol. 126, pp. 495-503, 2019. [Accessed September 2023].

[18] G. E. Obafemi, J. Coetzee and E. I. Iwuoha, "IoT sensor networks for water quality monitoring and control: Design challenges and opportunities," Talanta, vol. 225, 2021. [Accessed September 2023].

[19] M. T. Amin, M. S. Islam, T. Z. Bappy, and A. K. M. Azad, "Water quality monitoring and contamination detection by using IoT and machine learning techniques," SN Applied Sciences, vol. 2, no. 979, 2020. [Accessed September 2023].

[20] A. Khare and A. K. Mishra, "Real time water quality monitoring system using IoT and machine learning," Materials Today: Proceedings, vol. 33, pp. 3459-3463, 2020. [Accessed September 2023].

[21] L. Nunes, S. Deb and K. Berket, "A survey on human feedback for machine learning with applications in water management," Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no. 17, pp. 15152-15160, May 2021. [Accessed September 2023].

[22] O. Foster, C. Cater and D. Kaur, "Machine learning for monitoring and improving drinking water quality," npj Clean Water, vol. 4, no. 20, 2021. [Accessed September 2023].

[23] M. S. Idrees, M. U. Abbasi, A. Umer, Z. A. Jhatial and H. Bibi, "Water quality prediction using machine learning techniques: A systematic literature review," Applied Sciences, vol. 10, no. 18, p. 6321, September 2020. [Accessed September 2023].

[24] F. Y. Jau, W. M. Sani, N. B. Muhamad and S. Yusoff, "Development of IoT-based river water quality monitoring system," IETE Journal of Research, pp. 1-10, 2021. [Accessed September 2023].

[25] M. F. M. Firdhous, S. Karunarathne and D. Kasthurirathne, "A convolutional neural network model for real time water quality monitoring using IoT," 2021 International Research Conference on Smart Computing and Systems Engineering (SCSE), Colombo, Sri Lanka, 2021, pp. 83-88. doi: 10.1109/SCSE53038.2021.9582166. [Accessed September 2023].

[26] S. K. U. Yogendra Kumar, "Machine learning model for IoT-Edge device based," IEEE, 2022. [Accessed September 2023].

[27] N. J. M. A. B. Amara Paranagama, "WATER QUALITY PARAMETERS IN RELATION TO CHRONIC KIDNEY DISEASE IN SRI LANKA," IEEE, 2021. [Accessed September 2023].

[28] B. K. B. K. S. G. Dinithi Weerasinghe, "Identifying the Type of Chronic Kidney Disease," IEEE, 2021. [Accessed September 2023].

[29] B. M. A. W. Andrew OMAMBIA, "Water Quality Monitoring Using IoT & Machine Learning," IEEE, 2019. [Accessed September 2023].

[30] D. P. L. S. H. R. Sathira Hettiarachchi, "An Integrated Platform of Water Quality," IEEE, 2019. [Accessed September 2023].

[31] Abdelaziz, Ahmed & Salama, Ahmed & Riad, Alaa el-din & Mahmoud, Alia. (2019). A Machine Learning Model for Predicting of Chronic Kidney Disease Based Internet of Things and Cloud Computing in Smart Cities. 10.1007/978-3-030-01560-2_5. [Accessed September 2023].

6 GLOSSARY

- CKD - chronic kidney disease
- IoT - Internet of Things
- ML - Machine Learning
- WHO - World Health Organization
- pH - Potential of hydrogen, a measure of acidity or alkalinity of water
- Turbidity - Cloudiness or haziness of water caused by suspended particles, used as an indicator of water quality and contamination
- Arduino - An open-source electronics platform used for building electronic devices and interactive objects with sensors and controllers.
- Microcontroller (MCU) - An integrated circuit chip that contains a processor core, memory, and programmable input/output peripherals.
- Wi-Fi Module - A device that provides wireless internet connectivity to electronics boards and systems, allowing them to connect to the internet and exchange data. Popular modules include ESP8266 and ESP32.
- LCD Display - Liquid crystal display, a flat panel display that uses liquid crystals and backlighting to display text, images, and graphics. It is used to display sensor readings, notifications, etc.
- MQTT - Message Queuing Telemetry Transport, a lightweight publish-subscribe messaging protocol used in IoT systems to communicate sensor data between devices.
- API - Application programming interface, a set of protocols, routines, functions for building software applications to access features of an operating system, software, or web service.
- JSON - JavaScript Object Notation, an open standard file and data format that uses human-readable text to transmit data between servers and web applications.

7 APPENDICES