Spatiotemporal fusion for improved water prediction: A hybrid model for the Burnett river Australia

By

Mehreen Waqar

Fall-2023-MS-IT SEECS

Supervisor

Dr Safdar Abbas Khan

Department of Computing

A thesis submitted in partial fulfillment of the requirements for the degree of Masters ofScience in Information Technology (MS IT)

In

School of Electrical Engineering & Computer Science

(SEECS) ,National University of Sciences and Technology

(NUST), Islamabad, Pakistan.

(August 2023)

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS/MPhil thesis entitled "Spatiotemporal fusion for improvedwater prediction: A hybrid model for the Burnett River, Australia" written by MehreenWaqar, (Registration No 00000326953), of SEECS has been vetted by the undersigned, found complete in all respects as per NUST Statutes/Regulations, is free of plagiarism,errors and mistakes and is accepted as partial fulfillment for award of MS/M Phil degree. Itis further certified that necessary amendments as pointed out by GEC members of thescholar have also been incorporated in the said thesis.

Signature: ____________ ____

Name of Advisor: Dr. Safdar Abbas Khan

Date: 23-Nov-2023

HoD/Associate Dean:_

Date: _23-Nov-2023___

Signature (Dean/Principal): ____________________

Date: ______________________________________

Shahzad

Date: ____________________________

Approval

It is certified that the contents and form of the thesis entitled "Spatiotemporal fusion for improved water prediction: A hybrid model for the Burnett River, Australia" submitted by Mehreen Waqar have been found satisfactory for the requirement of the degree

Advisor : Dr. Safdar Abbas Khan

Signature: ____________

Date: 23-Nov-2023

Signature: ____________

23-Nov-2023

Committee Member 2: Dr. Muhammad Khuram

Signature: ____________

Date: 23-Nov-2023

Signature: ____________________________

Date: ____________________________

Dedication

This thesis is dedicated to all the deserving children who do not have access to quality educationespecially young girls.

Certificate of Originality

I hereby declare that this submission titled "Spatiotemporal fusion for improved water prediction: A hybrid model for the Burnett River, Australia" is my own work. To the best of my knowledge it contains no materials previously published or written by another person, nor material which to a substantial extent has been accepted for the award of any degree or diploma at NUST SEECS or at any other educational institute, except where due acknowledgement has been made in the thesis. Any contribution made to the research by others, with whom I have worked at NUST SEECS or elsewhere, is explicitly acknowledged in the thesis. I also declare that the intellectual content of this thesis is the product of my own work, except for the assistance from others in the project's design and conception or in style, presentation and linguistics, which has been acknowledged. I also verified the originality of contents through plagiarism software.

Student Name:Mehreen Waqar

Student Signature: Mehreen _

Acknowledgments

Glory be to Allah (S.W.A), the Creator, the Sustainer of the Universe. Who only has the power to honour whom He please, and to abase whom He please. Verily no one can do anything without His will. From the day, I came to NUST till the day of my departure, He was the only one Who blessed me and opened ways for me, and showed me the path of success. Their is nothing which can payback for His bounties throughout my research period to complete it successfully.

Mehreen Waqar

Abstract

This research is motivated by the imperative to develop robust computational methods for the control of pollution in water and the enhancement of quality in water of the Burnett River, Aus- tralia. Recognizing the intricate interplay of spatial and temporal dynamics in water quality, we propose hybrid model, denoted as "CNN-LSTM." The amalgamation of (CNN) Convolutional Neural Networks and (LSTM) Long Short-Term Memory networks addresses the unique chal- lenges posed by this complex system. Dissolved Oxygen is identified as a pivotal parameter for prediction, and meticulous feature engineering techniques are employed to refine its role within the model.

The empirical results of our investigation unveil a notable enhancement in prediction perfor- mance when employing the "CNN-LSTM" hybrid model in comparison to the AT-LSTM model.This improvement underscores the efficiency of combining CNN for spatial data and LSTM for temporal data, aligning with the inherent characteristics of water quality time series. The hybrid model demonstrates its capability to capture the intricate relationships between various environ- mental factors, leading to more accurate predictions.

Additionally, the study highlights the importance of spatial and temporal considerations when predicting future impacts of dissolved oxygen in the Burnett River. The combined methodof CNN and LSTM not only uses the distribution of negative water but also collects the ex- pected time to better understand the performance of the system. In addition to providing a high- performance and effective method for predicting water in water bodies, this research contributes to the expansion of environmental management by providing good practices in decision-making and management of fixed assets.

Contents

Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

Abbreviations

TSA Time Series Analysis

AT-LSTM Attention-based Long Short-Term Memory

CNN-LSTM Long Short Term Memory with Convolutional Neural Network

LSTM Long Short Term Memory

CNN Convolutional Neural

Network**GRU** Gated Recurrent Unit

CHAPTER 1

Introduction

1. Background

In terms of environmental health, water quality is an important factor that directly affects ecosys- tems and population. The Burnett river located in Australia plays an important role in the local environment, being an important source of water for agricultural, industrial and domestic pur- poses but with high levels of human activity and natural resources a it threatens permanent freshwater with river.

This study focuses on the use of hybrid models for time series-based data of the Burnett River. Hybrid models combine the strengths of different forecasting methods, improving forecast ac- curacy and reliability. The combination of machine learning and statistical methods enables detailed analysis of the complex interactions involved in water quality development

Several studies have shown the effectiveness of the machine learning and deep learning models in environmental forecasting tasks. In last decade, many machine learning and deep learning models applied on water contaminants datasets.As noted by [24], both ANN and SVM models have good performance in predicting the water quality of the Tireh River in Western Iran. When evaluating the accuracy of the model used, the most accurate model according to the error metric is considered to be SVM. We perform a periodic analysis comparing various machine learning and deep learning models as described in the work of Chawla et al. [**WQM-10**]. Their study evaluated regression and machine learning models such as linear regression, random forest, support vector machine (SVM), and long-term memory (LSTM) to predict salinity levels and progeny in the Salton Sea. Studies were conducted on the water quality of the Salton Sea and focused on estimating chemical or physical properties such as salinity, pH, and dissolved oxygen. It also investigates the relationship between hydrological parameters and environmental parameters around the Salton Sea. Findings from this research have the potential to influence policymakers, provide an understanding of water salinity, and support policies to reduce salinity for freshwater management. From observational analysis, it is clear that the machine learning process is easier and more accurate than statistical models and helps improve results.Wang et al. [3] focuses on LSTM neural network deep learning model for water quality prediction in Taihu Lake. As mentioned in [28], continuing deep learning, Long Term Memory Recurrent Neural Network (LSTM-RNN) was introduced using monthly data from five different sites in Damavand district. Dataset covering the years 2009 to 2021 to estimate water pollution in Iran. Neural network architectures are evaluated using the “f-score” metric to determine their performance.

Through the utilization of machine learning and deep learning models, as discussed by Hu, Yankun et al. in [33], a novel hybrid improved temporal convolution network model is intro- duced for the time series prediction of the Burnett River and Liao River. The authors con- duct a comparative analysis of the model’s performance against several other models, including the Support Vector Regressor (SVR) model, the temporal convolutional network (TCN) model,the long short-term memory (LSTM) model, and the autoregressive integrated moving average (ARIMA) model. The outcomes of the comparison reveal that the proposed model surpasses the other models in terms of prediction accuracy.

The Burnett River in Australia, like many other water bodies, faces challenges related to waterquality and quantity. Efficient water prediction models are crucial for managing and sustaining water resources. The concept of spatiotemporal fusion involves integrating both spatial and temporal dimensions to enhance the accuracy of predictions. In the context of the Burnett River, this fusion approach aims to provide a holistic understanding of how water quality and quantity vary across different locations and time periods.

The spatiotemporal fusion model for the Burnett River represents a hybrid approach, likely combining statistical techniques, machine learning algorithms, and geospatial data to create a comprehensive predictive framework. This hybrid model is designed to leverage the strengths of various methodologies, capturing both the spatial patterns along the river and the temporal dynamics over different seasons or periods.

As water quality and availability are vital factors for ecological sustainability and human con- sumption, the development of an advanced hybrid model for the Burnett River is crucial. This model is expected to provide more accurate and robust predictions, facilitating informed decision-making in water resource management. The next part of the study provides an in-depth study of the methods, data and results of the spatiotemporal fusion model, providing a better under-standing of its applicability and effectiveness for improved water forecasting in the context of the Australian Burnett River.

Historically, water prediction models have predominantly focused on either spatial or temporal aspects individually, potentially overlooking the intricate relationships between these dimen- sions. A hybrid model incorporating spatiotemporal fusion techniques represents an innovative solution to address this limitation. By combining both spatial and temporal information, the model aims to provide a more accurate and nuanced representation of the Burnett River's be-havior, capturing the spatial variations at different locations along the river and the temporal changes over distinct time intervals.

The significance of this spatiotemporal fusion model for the Burnett River extends beyond aca- demic interest, as it addresses real-world challenges in water resource management. Accurate predictions of water quality and quantity facilitate proactive decision-making, enabling author- ities to implement measures for environmental conservation, sustainable agriculture, and urban water supply.

In subsequent sections, this study delves into the detailed methodology employed in developing the hybrid model, the sources of data utilized, and the expected outcomes. By doing so, the research aims to contribute valuable insights and practical applications to the field of water prediction and management, particularly in the context of the complex and dynamic Burnett River in Australia.

2. Objectives

This research focuses on the development of computational methods to control water pollution and enhance water quality. Time series forecasting is a statistical technique that uses historical data to predict future values of a variable. In the context of water quality, time series forecasting can be used to predict the concentration of contaminants in a river at a future time based on past measurements. The objective of time series forecasting of water contaminants in Burnett Riveris to provide accurate predictions of the concentration of contaminants in the river at different points in time. This information can be used to identify potential sources of contamination, monitor the effectiveness of pollution control measures, and inform decision-making regarding water management policies. A recent investigation led by Honglei Chen et al. [22] utilized a long short-term memory (LSTM) network, along with its attention-based variant (AT-LSTM), to forecast water quality in the Burnett River, Australia. The main goal of the research is to achievea change in the ability to learn long-term dependencies and relationships in different physical data to improve the accuracy of water in water. Notably, the study's findings demonstrated the superior performance of both the LSTM and AT-LSTM models compared to other approaches in predicting water quality in the Burnett River.

Our research project is based on the application of accurate water quality prediction in Aus- tralia's Burnett River using advanced machine learning and deep learning. Specifically, we propose a hybrid model integrating a long short-term memory (LSTM) network with a Convo-lutional Neural Network (CNN) for predicting the temporal trends in water quality. The study draws inspiration from the outcomes of the aforementioned investigation [22] and extends the methodology by incorporating the CNN component. This addition will increase the predictive power of the attention-based LSTM (AT-LSTM) model. The overall goal is to use advanced machine learning and deep learning to improve the accuracy of Burnet River water quality pre- dictions. Integration of CNN is the best choice to improve the ability of the AT-LSTM model toprovide better insight and understanding of spatiotemporal complexities associated with water quality.

The research outcomes are anticipated to make dual contributions, addressing both academic and practical facets of water quality forecasting in the context of the Burnett River. Beyond enriching the scholarly understanding of water quality prediction, the study aspires to offer pragmatic insights that can inform effective water resource management and pollution mitigation strategies specific to the environmental dynamics of the Burnett River in Australia. The research objectives are aligned with the ambition of broadening and deepening the knowledge base, ensuring that the study's outcomes hold significance both within academic discourse and real-world applications in the field of water quality prediction and management.

The root conditions of environmental studies, integrals, in the absorption of Long Short Term Memory with Convolutional Neural Network (CNN-LSTM) in the gastrointestinal environment many of the many supplementary research centralizations add to the primary objective of devel- opmental management of the Burnet River is to enhance a computer simulation of the progress of the main objectives. The features are available in a multidimensional data group Analysis will help identify key factors affecting water quality improvement, while efforts to increase spatial resolution seek to differentiate riparian landscape diversity for more local methods Combining external factors, such as climate and land use change, aims to increase model predictive abil- ity. The development of interactive communication support systems will enable stakeholders to use real-time insights for scenario analysis and decision-making in water resource management. Rigorous evaluations using independent datasets and comparisons with existing models will im- prove the accuracy and reliability of the predictions of the LSTM-CNN model. Consideration of long-term sustainability and economic impacts through adaptive water management approaches contributes to the adoption of a holistic approach to water quality management Finally, the formation of stakeholder networks promotion aims to foster partnerships with communities,environmental organizations, and other stakeholders.

3. Research Motivation

The purpose of this work is to recognize the critical importance of maintaining and improv- ing water quality in river ecosystems. The Burnett River, like many watersheds in the world, faces increasing challenges from urbanisation, agriculture and other human activities. This challenge requires a proactive and technologically advanced approach to environmental man- agement. Conventional water quality analysis methods often struggle to capture the complex patterns and subtle changes inherent in time series data. The use of machine learning and deep learning models offers an exciting opportunity to gain a deeper understanding of the temporal evolution of water quality parameters.

Water quality is an important issue that affects the health and well-being of people and the en-vironment. The Burnett River is one of Australia's most important waterways and its water quality is critical to local communities. Investigating the water quality of the Burnett River using machine learning and deep learning models can provide valuable insight into factors af-fecting water quality. Research also demonstrates the potential of machine learning and deep learning models to predict future water quality trends and inform environmental management decisions. Water assessment organizations can use this model to make more informed decisions about water resource management and environmental protection. Aquatic life is an important part of the ecosystem, and water quality plays an important role in maintaining the health and well-being of aquatic life. Poor water quality can have serious consequences for aquatic life, including fish kills, algal blooms and habitat destruction. Water quality studies in the Burnett

River use machine learning and deep learning models to identify factors that affect water quality and inform the development of effective improvements can. help provide strategies to protect water quality and aquatic life. By solving these problems, you can help improve water quality in the Burnett River and other rivers around the world and contribute to the development of new and innovative solutions to protect water resources. In addition, studying water quality in the Burnett River is important because it can provide information to identify key factors affect- ing water quality and develop effective strategies to improve water quality. This research can also help raise awareness of the importance of water quality and the need for effective water management practices. By solving this problem, you can contribute to the development of new and innovative solutions to improve water quality and protect the environment. By addressing this problem, we can contribute to the development of new and innovative solutions to improve water quality and protect the environment and aquatic life.

In addition to the reasons above, this research is based on the decision to deal with the complex challenges of climate change. By understanding these changes and predicting water resources, they learned that changes in precipitation patterns and increases in extreme weather events can have a positive impact on water. The goal of the hybrid space-time coupled model is to develop methods to capture and mitigate the impacts of climate change on the water quality of the Bur-nett River. In addition, the research focuses on the need to bridge the gap between conventional hydrological models and state-of-the-art machine learning techniques. Although conventional models provide valuable insights, the complexity of spatiotemporal interactions in river systems requires a more sophisticated and adaptive approach. The integration of spatial and temporal scales through machine learning models represents an innovative step towards a more comprehensive understanding of the Burnett River's water quality dynamics.

This research was also inspired by the potential social and economic consequences of accurate water quality prediction. A robust predictive model for the Burnett River can help identify and mitigate early water quality problems, the economic burden of environmental degradation and the human health impacts of poor water quality. Emphasizing the societal importance of this research, it goes beyond academics to directly impact communities that depend on the Burnett River for various purposes.

The rationale for the research is the need to improve predictive capabilities in water quality management, particularly in the Burnett River. By exploring spatiotemporal coupling through hybrid models, the research aims to contribute to the development of innovative solutions that will improve our understanding of water quality dynamics, inform decision-making processes and ultimately contribute to sustainable water management in the Burnett River and beyond.

4. Problem Statement

Poor water quality can have adverse effects on aquatic life, including fish kills, algae blooms, and habitat destruction. By dealing with this problem, we can help develop new and innovative solutions to improve water quality and protect aquatic life around the world.

The AT-LSTM model is a promising approach for Time Series Analysis (TSA), especially in the context of environmental data such as water quality in the Burnett River, Australia. How- ever, there is significant scope to improve its prediction performance. This study attempts to improve the efficiency and accuracy of the AT-LSTM model by introducing a new hybrid model to overcome the limitations of existing methods and improve prediction.

The problem identified is that poor water quality affects aquatic ecosystems, not only affecting the health of individual species, but also disrupting the overall ecological balance. Fish kills, al- gal blooms and habitat destruction are not isolated incidents, but symptoms of a larger systemic problem. This encourages comprehensive research into factors contributing to the degradation of water quality in the river, recognizing that addressing this challenge requires sophisticated and targeted solutions. By designing innovative strategies, we not only reduce threats to aquatic life, but contribute to the broader mission of protecting biodiversity and ecosystem stability globally.

Emphasis on the Attention-based Long Short-Term Memory (AT-LSTM) model as part of the temporal analysis of Burnett River water quality demonstrates the importance of using advanced technology in environmental research. The limitations identified in the Attention-based Long Short-Term Memory (AT-LSTM) model provide an open entry point for improvement, recogniz- ing the importance of predictive modeling advances to advance the evolution of environmental dynamics. Recognizing this opportunity for growth is a catalyst for progress in predictive ana- lytics for water quality, creating the basis for new hybrid models that combine the strengths ofdifferent methodologies.

In addition, the formulation of the problem includes a broader need for adaptive models and spe- cific contexts in environmental research. The complexity of spatiotemporal changes in Burnett River water quality requires a model that not only captures the nuances of these changes, but also provides effective insights for effective environmental management. The goal is not only to address the limitations of the current model, but also to establish a framework to continue improving their prediction capabilities, ensuring the continued relevance and applicability of research findings to the rapidly growing field of water quality prediction. In doing so, the studyaims to contribute not only to the specific challenges of the Burnett River, but also to the broader conversation in advancing predictive models for sustainable water resource management.

5. Contribution

The goal is to create a model that can accurately predict the concentration of water pollutants in a river at a given time. The model must be able to handle complex time series data and non-linear relationships between water quality parameters. The model should be able to handle missing data and data outliers. The thesis aims to investigate the use of machine learning and deep learning models for the temporal prediction of river water pollutants. This topic will explore the use of machine learning and deep learning models to predict river water pollutants. In addition, the topic will compare the performance of the model with conventional statistical models. This topic will provide insight into how machine learning and deep learning models are used to predict river water pollutants over time. In addition, the topic will provide recommendations for choosing the best model for predicting river water pollutants.

This article focuses on exploring machine learning and deep learning models as tools for pre- dicting water pollution in the Burnet River. This study aims to evaluate the performance of this advanced model compared to traditional statistical models. In this way, research attempts to better understand the effectiveness of different models in predicting water pollution.

Research includes exploring how machine learning and deep learning models can overcome the challenges posed by the spatiotemporal nature of water quality data. The research aims not only to determine the performance of this model, but also to provide insight into its strengths and limitations compared to traditional statistical methodologies. This comparative analysis will add valuable knowledge to the field by revealing the suitability of different models for predicting water pollutants in river environments. In essence, this research seeks to improve the ability of predictive modeling in water quality assessment, providing notable contributions that encourage the understanding and application of advanced modeling techniques. The expected results willnot only benefit the scientific community, but also have practical implications for sustainable water resource management in river ecosystems, contributing to the overall goal of protecting and improving environmental health.

The study will provide practical recommendations for selecting the most effective model for predicting water pollutants in the Burnett River. These recommendations will be based on em- pirical results from comparative studies that will guide researchers, environmentalists, and pol- icy makers related to water quality prediction and management. Ultimately, the research seeksto improve the understanding of predictive models in the context of water quality, encouraging the development of more accurate and effective tools for the sustainable management of water resources in river ecosystems.

- We aim to provide innovative solutions to address water pollution and protect water qual-ity.
- We will improve the performance of the AT-LSTM model by introducing a hybrid modelwith Long Short Term Memory with Convolutional Neural Network (CNN-LSTM).

CHAPTER 2

Literature Review

Literature Review is split into two parts which summarizes existing research in machine learning and deep learning models. (i) Smart aquaculture models, (ii) Water quality monitoring models.

1. Smart aquaculture models

A thorough review of the existing literature on smart farming models reveals an exciting field that uses advanced technologies to transform traditional farming practices. Scientists and ex- perts combine sensors, IoT devices, data analytics and artificial intelligence (AI) to explore new opportunities in the agricultural sector.

The main focus in the literature is the use of sensor networks for real-time monitoring of critical parameters in agricultural systems. These sensors help collect data on water quality, tempera- ture, oxygen levels, and pH, among other factors. This continuous monitoring allows farmers to quickly identify anomalies or potential problems, allowing them to make informed decisions. Another major theme is the use of data analytics, which involves using tools to process large amounts of data collected by sensors. These tools provide valuable insights, support trend anal- ysis and even enable predictive models. Machine learning algorithms, part of AI, are used to predict changes in water quality, optimize feeding schedules, and identify patterns that may in- dicate disease outbreaks. The importance of artificial intelligence is essential to automate andoptimize various aspects of farming operations. AI-driven systems enable precise control of feed, weather and environmental conditions, resulting in better use of resources, less waste and improved overall efficiency.

Remote monitoring and management solutions focus on the ability to facilitate remote aquaculture operations. The development of mobile applications and web interfaces allows farmers to monitor and control their farming systems remotely, providing a significant advantage in re- sponding quickly to emerging problems. Water health monitoring, an important area studied in the literature, focuses on early disease detection using continuous monitoring and imaging technologies and computer vision. This technology has the potential to revolutionize disease management practices in aquaculture. The literature also emphasizes environmental sustainabil- ity, suggesting that smart farming models can reduce the environmental impact of operations. Optimizing the use of resources, reducing waste, and implementing sustainable practices are considered important components of creating a more environmentally friendly aquaculture. [38] Haijin Chin et al. discuss the creation of an accurate agricultural environmental factor prediction model using big data technology and machine learning algorithms. The model can predict water quality, improve water quality management, and reduce risk for water management. This paper introduces the features and innovations of deep machine learning and deep learning-based mod- els for aquatic water quality prediction. Explain the basic process and conventional methods for predicting farm environmental factors based on machine learning. Analyzing model per- formance index, the integrity of data collection, scientific processing, and rationality of model construction methods are the basis for building high-quality models. The article also addresses the problems inherent in empirical decision-making models and the need to integrate the de- velopment of water environment assessment technology to form the intelligent face of water environment monitoring technology to promote the development of smart water industry. in the future. This article provides a comprehensive review of research on technology-based agri- environment model prediction. This highlights the importance of developing good models that can predict water quality, improve water quality management and reduce water management risks. The paper also identifies challenges and opportunities in this field and suggests future research directions for developing smart water environmental regulatory systems and support- ing the development of smart water management. Research is also exploring the integration of smart farming models with wider agricultural management systems. This integrated approach aims to achieve smooth coordination and synergy with other agricultural sectors, resulting in overall agricultural efficiency and productivity. In the [11] Qiangqiang Ye et al. Shanghai River, China the researcher proposed a new LSTM-RNN network model to predict water quality pa- rameters based on the improved RNN network structure. This model is designed to optimize the RNN network architecture and threshold the connection layer and hidden layer. The researchused a time series of water quality monitoring data from June 2016 to May 2017 for model development. The simulation results show that the proposed LSTM-RNN model outperforms the conventional gray model (GM) and the RNN network model in terms of estimation accu- racy and training prediction generalization ability. Another CNN-LSTM hybrid outperformed other models in terms of prediction accuracy and computing time. In this paper, [26] the au- thor proposed a hybrid deep learning model that combines convolutional neural network (CNN) with long-term memory (LSTM) and gated repetition unit (GRU) to predict water quality for smart agriculture. The proposed model is compared with several LSTM, GRU, and CNN DL models, as well as focus-based LSTM and focus-based GRU DL models. The authors also con- ducted experiments using two different water quality datasets and comprehensively investigated the effect of hyperparameters on the performance of the proposed hybrid DL model. The per- formance of the hybrid model and the base model is discussed in [36] Vinoth Kumar P et al. SAS-MI model uses deep convolutional neural network (D-CNN) and k-cluster method to pre- dict water quality for smart water management. The K-cluster method combines unlabelled data sets used for training and testing. D-CNN predicts water quality for smart water management through automatic feature engineering by neural networks. The performance of the proposed model is evaluated using comparative studies of existing prediction models such as logistic regression, decision trees, XG boosting classifiers, k-nearest neighbors, etc. The researchers present a model for predicting seawater quality indicators, employing both univariate and mul- tivariate regression analyses [23]. Utilizing an open public database, the authors constructed the regression model and validated it with Pacific Ocean data. The study scrutinizes whether the acquired data align with seawater quality standards, assessing the feasibility of seawater sus- taining essential nutrients. Initially, the researchers processed the database to eliminate noise and artifacts, employing univariate and multivariate regression analyses for predicting seawater quality indicators. Comparative evaluation against an open public database revealed that their method surpassed state-of-the-art techniques in terms of prediction accuracy. The proposed model for predicting seawater quality indicators, rooted in univariate and multivariate regres- sion analyses, relies on an open public database. Validation of the model utilized Pacific Ocean data, with the author asserting its superiority over state-of-the-art methods regarding prediction accuracy. Nevertheless, further research is imperative to assess the effectiveness of their meth- ods on a larger scale and in comparison with existing methodologies. The researchers present a model for predicting seawater quality indicators, employing both univariate and multivariate regression analyses. Utilizing an open public database, the authors constructed the regression model and validated it with Pacific Ocean data. [27]The study scrutinizes whether the acquired data align with seawater quality standards, assessing the feasibility of seawater sustaining essen- tial nutrients. Initially, the researchers processed the database to eliminate noise and artifacts, employing univariate and multivariate regression analyses for predicting seawater quality in- dicators. Comparative evaluation against an open public database revealed that their method surpassed state-of-the-art techniques in terms of prediction accuracy. The proposed model for predicting seawater quality indicators, rooted in univariate and multivariate regression analyses, relies on an open public database. Validation of the model utilized Pacific Ocean data, with the author asserting its superiority over state-of-the-art methods regarding prediction accuracy. Nevertheless, further research is imperative to assess the effectiveness of their methods on a larger scale and in comparison with existing methodologies. [20]This research paper introduces a machine learning-driven approach to predict water quality variables within fish farming ponds. The primary objective is to construct models capable of estimating and forecasting crucial wa-ter quality parameters, including dissolved oxygen, pH, ammonia, and ammonium, utilizing various machine learning techniques. The study puts forth a methodology designed for two scenarios: i) estimating unobserved variables based on observed ones, and ii) forecasting withlimited training data availability. Random forests, multivariate linear regression, and artificial neural networks are employed to analyze data from commonly measured water-quality variables in fish farming. The findings of the study reveal that random forests exhibit efficacy in forecast- ing dissolved oxygen, pond temperature, pH, ammonia, and ammonium, even when the water pond variables are measured only twice per day. Notably, the prediction models are adaptable for implementation on a mobile-based information system, making them accessible for utiliza- tion on average smartphones affordable to fish farmers. The study contributes valuable insights into leveraging machine learning techniques for predicting water quality variables in fish farm- ing ponds. The implications of the findings extend to policymakers and fish farmers, aiding in assessing the risk of fish mortality and formulating policies to mitigate fish loss. The study underscores the significance of water quality monitoring in fish farming and advocates for the adoption of machine learning techniques in predicting water quality variables. The proposed methodology emerges as a practical tool empowering fish farmers to make informed decisionsand minimize losses. The research makes a noteworthy contribution to the domain of water quality prediction and management.

2. Water quality monitoring models

Monitoring water quality in rivers involves the analysis of time series datasets, where key pa- rameters are measured or sampled at regular intervals over time. Various modeling approaches are employed to understand, predict, and detect patterns in river water quality. Statistical mod- els such as ARIMA and exponential smoothing state models capture patterns and chance, while learning models such as linear regression, pruning trees, random forests, and neural networks provide the flexibility to control linear and nonlinear relationships. Hybrid models such as ARIMA with exogenous variables and integrated models integrate other factors to increase pre- dictive power. Use random detection methods such as classification forest and single-class SVM to detect differences in time series data. Additionally, spatial-temporal models, including spa- tial autoregressive models and geostatistical models, consider spatial relationships to provide a more comprehensive understanding of water quality variations along rivers. It's essential to tailor these modeling approaches to the specific characteristics of the river system and param- eters being monitored, ensuring ongoing evaluation and validation for reliable predictions in real-world river environments. In one study two datasets are taken and some time series hybrid models are applied on both datasets and compare their results with baseline models. [13] The author reports the prediction of dissolved oxygen (DO) and chlorophyll-a (Chl-a) water quality variables in Lake Prespa, Greece. The researchers compared two independent deep learning models, a hybrid of long-term memory (LSTM) and convolutional neural network (CNN) mod- els, and a CNN-LSTM model with a traditional, support-vector machine learning model. regres-sion (SVR) and decision tree model (DT). The main innovation of this study is the development of a combined CNN-LSTM model to predict water quality variables.This study uses sensors to obtain physicochemical water parameters, primarily pH, oxidation-reduction potential (ORP),water temperature, electrical conductivity (EC), DO, and Chl, at 15-minute intervals starting from June 1. For modeling until May 31, 2012. The results showed that the CNN-LSTM hybrid model was better than the single model (LSTM, CNN, SVR and DT model) in predicting DO and Chl-a. By combining the LSTM and CNN model, the hybrid model was able to achieve both low and high levels of water quality parameters, especially DO concentration. A deep learning method is proposed to predict water quality in rivers, In[12] System, Convolutional Neural Network (CNN) and Long Short Term Memory (LSTM) system is proposed to predict water quality and water level. The authors used water level and water quality data from the Nakdong River basin to train and validate their model. They found that CNN and LSTM models with Nash-Sutcliffe efficiency greater than 0.75 were in "excellent" condition, indicating thetemporal variation of pollution in the Nakdong River basin. The plan can be used for water level and water quality measurement. [19] Shajian et al. The author compares the performance of three deep learning models, namely convolutional neural network (CNN), long-term mem- ory neural network (LSTM), and hybrid CNN-LSTM, under different input data preprocessing methods. The input data includes the first one-dimensional time series and two-dimensional grayscale images segmented according to the fully integrated empirical mode segmentation al- gorithm with adaptive noise (CEEMDAN). The authors used two parameters of water quality, total oxygen (DO) and total nitrogen (TN), to estimate instantaneous water quality. The results show that CNN-LSTM outperforms the standalone models CNN and LSTM. The model using the CEEMDAN-based material performed better than the model using the original material; Im- provements in aperiodic TN parameters were greater than those in periodic DO parameters. As the number of prediction steps increases, the accuracy of the model gradually decreases because the original input data is faster than the input data based on CEEMDAN and the aperiodic TN parameter is faster than the time DO parameter. Water Quality Index (WQI) can be used to predict water quality through machine learning models. The authors used the following water quality parameters to calculate WQI: temperature, dissolved oxygen (DO) (% sediment), pH, conductivity, biochemical oxygen demand (BOD), nitrate (NO3), fertilizer and total total col- iforms (TC).). These parameters are used as vectors representing water quality. The authors used five classification algorithms to predict water quality, including Naive Bayes (NB), De- cision Tree (DT), K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF). The experiments were conducted using real data with detailed information on the differences between Tamil Nadu and synthetic data created from the differences. Based on theperformance of five classifications, random forest classification is shown to have better results than other classifications. The analysis shows that the machine learning method can predict wa- ter quality indicators well. [39] Shams et al. also discuss the use of machine learning models to predict water quality index (WQI) and water quality classification (WQC). Four classification models are used in the study: Random Forest (RF), Extreme Gradient Boosting (Xgboost), Gra-dient Boosting (GB) and Adaptive Boosting (AdaBoost), and four regression models, namely the K-Nearest Neighbor (KNN) regressor model, decision. tree (DT) regressor model, support vector regressor (SVR) model, and multilayer perceptron (MLP) regressor model. The author uses mesh research to optimize and set the parameters of this model. The database used in this research contains 7 features and 1991 instances. The test results show that the GB model gives the best results with an accuracy of 99.50 % when predicting WQC values. According to the test results, the MLP regressor model outperforms other regression models and the R2 value is 99.8% when predicting the WQI value.Comparison of water quality classification models em- ploying machine learning algorithms presents in [15]. This study compares the performance of three machine learning algorithms, namely, Support Vector Machine (SVM), Decision Tree, and Naïve Bayes, in classifying water quality based on pH, dissolved oxygen (DO), biochemical oxygen demand (BOD), and electrical conductivity. The classification models are trained based on the weighted arithmetic water quality index (WAWQI) calculated. The study found that the decision tree algorithm was the best classification model with an accuracy of 98.50%. To study the ground water of Haryana state (India) [14] Bui et al. found that the DNN algorithm outper-formed the other two algorithms with an accuracy of 98.5%. They examined the performance evaluation of three machine learning algorithms, namely deep neural network (DNN), gradient boosting machine (GBM), and cloud gradient boosting (XGBoost), for measuring water on theground. The combination of machine learning and deep learning models suggests new ideas for water quality assessment. Aslam et al. [21] The research focuses on four independent al- gorithms such as Random Tree (RT), Random Forest (RF), M5P, and Reduced Error Pruning Tree (REPT) with 12 hybrid data search algorithms to build WQI prediction models. The studyrevealed that the hybrid RT-Artificial Neural Network (RT-ANN) algorithm outperformed all other algorithms with 94.5% accuracy.

Machine-based water quality measurement uses the water quality index (WQI), which is a num- ber expressing water quality in [**ml-2**]. Parameters used to calculate WQI include pH, tem- perature, conductivity, dissolved oxygen (DO), biological oxygen demand (BOD), nitrate, and total coliforms. The author has used various machine learning algorithms such as K-Nearest Neighbors (K-NN), Naive Bayes, Support Vector Machine (SVM), Decision Making and Ran- dom Forest algorithms and made research comparison based on observations such as accuracy.

, confusion matrix, accuracy, bias and f1 score. The authors said the plan outperformed other methods in the state in terms of accuracy and speed. Measuring water quality is a complex issue that affects the use of large amounts of water. Different needs require different models. Water quality should be determined using a combination of chemical and physical properties that depend on the intended use of the water. Acceptance and rejection must be defined for each variable. The water exchanger is considered ready for use when it meets the requirements. If water does not meet these requirements, it must be purified before use. Water quality indicators measure the quality of water that can be produced, taking into account its use. This instrument has many chemical and physical properties. The main aim of this study is to analyze the learning process for predicting water quality using eight variables such as temperature, conductivity, pH, dissolved oxygen (% sat), nitrate, feces and all coliforms. Another study using a random forest algorithm outperformed water quality prediction with 97.5% accuracy. [35]Anita et al. statedthat the main purpose of this study is to analyze machine learning methods to predict water quality using eight different factors including temperature, conductivity, pH, dissolved oxygen (% sat), nitrate, fecal matter and total coliform. The IoT framework has four modules: sens- ing, coordinator, data processing, and decision. In this study [34], authors proposes an integrated framework that combines the Internet of Things (IoT) and machine learning paradigms for com- prehensive water quality analysis and prediction. The study used a machine learning technique called K-Means clustering to forecast water quality using a training data set consisting of various water samples. The system is now in its prototype stage, used with low-cost embedded devices like the Arduino Uno and Raspberry Pi3. Various challenges and issues proposes possible so- lutions to some research issues that the integration of big data analytics and machine learning techniques can aid in building water quality prediction models. In this survey [18], authors an- alyzes various prediction models developed using machine learning and big data techniques for water quality prediction and evaluation. The paper highlights the importance of evaluating and monitoring the quality of water, and its prediction using machine learning and big data analytics. Some study proposes an IoT water quality monitoring system that collects and transmits datato MQTT Brokers and stores it in a database. The data is presented on a monitoring webpage.Three machine learning methods, namely Random Forest, ANN, and LightGBM, were used forbackend analysis and prediction. LightGBM was found to have the highest prediction accuracyfor NH3, pH, ORP, and temperature [30]. The research contributes to reducing the need for frequent and costly data collection by using an IoT system for real-time monitoring and em- ploying machine learning predictions to compensate for missing data. This approach providesa more efficient and effective method for analyzing and predicting water quality. Importance of drinking enough water every day to prevent the body from overheating, dehydration, kidney stones, and other health issues. The paper proposes the use of artificial intelligence techniquesto predict the Water Quality Index (WQI) and Water Quality Classification (WQC) using the Indian Water Quality dataset. The paper uses neural network models such as Long Short-TermMemory (LSTM) and regression models such as Ridge Regression, Random Forest Regressor with Randomized search CV for WQI prediction. For WQC forecasting, Machine Learning models like KNN, Logistic Regression, Logistic Regression Using GridSearchCV, XGBoost, SVM, and SVM Using Grid SearchCV for train-test splits like 70–30, 80–20 have been applied. The paper [29] concludes that the proposed models can predict the quality of water with highaccuracy. Importance of water as a crucial resource for life and the need to assess its purity before using it for any purpose. Machine learning algorithms is used to predict the quality of water based on parameters such as pH value, turbidity, hardness, conductivity, dissolved solids in water, and other parameters. This paper [37] uses various machine learning models such as Random Forest Regression, Long Short-Term Memory (LSTM), KNN, Logistic Regression, XGBoost, and SVM to predict the Water Quality Index (WQI) and Water Quality Classification (WQC) using the Indian Water Quality dataset. The paper concludes that the proposed model can predict water quality with high accuracy. Smart monitoring systems powered by artificialintelligence (AI) will enable water operators to provide real-time quality control for chemical and/or biological contamination and risk management. Research [40] simulates the spread of biocontamination risk in the Water Distribution System (WDS) with a focus on source identifi- cation and response modeling. The study considered the related changes in several water quality parameters caused by the mixture of pollutants. These parameters include total organic carbon, pH, and salinity. The researchers concluded that the proposed model can predict water quality with high accuracy.

[32] The neural network model proposed by Guo et al. is based on principal component analysis (PCA), particle swarm optimization (APSO), and extreme learning machine learning (ELM). The model extracts the principal components of water quality parameters through PCA, re- duces the data dimension, and removes the complexity and correlation between water quality parameters. The authors introduce linear dynamic tuning of inertia weights, adaptive learning factors, and adaptive cross-sectional mutation strategies to improve particle propagation algo- rithms, improve optimization capabilities, avoid local optima, and find more optimal solutions.The authors also used the particle propagation optimization algorithm to continuously optimize the input layer weights and hidden layer thresholds of the limited learning machine from the randomization of the initial input layer weights and hidden layer thresholds, improving the ac- curacy of dissolved oxygen prediction and finally PCA to predict dissolved oxygen in quality river water - Building APSO-ELM neural network model. Author Jichang et al. [9] developed a water quality prediction model based on a hybrid convolutional neural network (CNN) and gated recurrent unit (GRU), CNN-GRU, using real monitoring data of Shanghai Jinze Reser- voir. This model aims to improve the accuracy and efficiency of water quality prediction by using the CNN network to extract potential features among continuous water quality data and

using the GRU network with temporary memory capabilities to accurately predict water quality data. The experimental results show that the hybrid forecasting model proposed in this paper has higher accuracy than the traditional SVR water quality forecasting model and the traditional GRU network water quality forecasting model. There are hybrid models with a higher degree of focus than water quality prediction. A spatiotemporal prediction algorithm called Graph Attention-Spatiotemporal (GAST) neural network to predict future water quality in various lo-cations. The proposed model is a wavelet network model based on extended causal diffraction. The proposed GAST neural network [25] studies the spatial and temporal dependence of water quality intervals. They claim that their method outperforms state-of-the-art methods in termsof prediction accuracy. They introduce a temporal conceptual framework as well as a compre-hensive conceptual approach to derive reliable time series that can effectively handle nonlinear relationships in time series. Spatial focus in river networks integrates the temporal characteris- tics of spatial points. Improving the Accuracy of Chain Estimation The authors evaluated theirmethod in two real-world scenarios and reported that their method outperformed current meth- ods in prediction accuracy. Flores et al. [31] introduces a machine learning-centric approach for predicting water quality status in the Loa River, situated within the exceedingly arid Ata- cama Desert. The primary objective is to devise a rule-based inference technique for generating water quality labels and a predictive model for water quality status, employing Random Forest, physicochemical parameters, and expert knowledge. The study draws upon historical data per- taining to physicochemical parameters from seven water monitoring stations in the Loa River, collected by the Chilean Ministry of the Environment. Notably, this study marks a novel con- tribution, as no prior research of this nature has predicted the water quality of the Loa River in such an arid zone. The research findings underscore the significance of the proposed machinelearning model, providing mean values for accuracy (acc) at 0.897, precision (p) at 89.73, andrecall (r) at 0.928. This model, unprecedented in predicting the water quality of the Loa Riverin an extremely arid environment, serves as a valuable tool for policymakers and water quality managers. It facilitates informed decision-making to mitigate water quality degradation, em- phasizing the critical role of water quality monitoring and the application of machine learning techniques in predicting water quality status in arid regions. This study makes a substantial con- tribution to the realm of water quality prediction and management. The insights derived from the research hold practical implications for policymakers and water quality managers, aiding in risk assessment and the formulation of policies to minimize water quality deterioration. The study effectively underscores the importance of water quality monitoring and the utility of machine learning techniques in predicting water quality status, particularly in arid regions. DeepLearning Methods for Predicting Water Quality in IoT Systems provides an in-depth look at the challenges of water quality monitoring in IoT systems, especially in the aquaculture and fisheries sectors. [16] An innovative approach with predictive modeling to adapt IoT systems for water quality monitoring. At the heart of the model is deep learning, which uses short-term mem- ory (LSTM) algorithms to predict important water parameters such as salinity, temperature, pH and dissolved oxygen (DO). Considering that these indicators are collected daily, they actually form a series of data or time. The efficacy of the proposed approach is substantiated through experimental results across various datasets, affirming its applicability in real-world systems. The article underscores the paramount importance of water quality monitoring, especially in the face of global climate change and the detrimental effects of water pollution, which pose sig- nificant challenges for farmers engaged in fish and shrimp cultivation. The proposed solution offers a proactive approach to address these challenges, enabling real-time monitoring of sensor data indicators and providing forecasts for early warnings. This proactive management strategy aims to optimize both the quality and quantity aspects of shrimp and fish cultivation.This paper presents an in-depth study focusing on water quality prediction in IoT systems. The design in- cludes a predictive model adapted to cultivation and fisheries in IoT systems, using the power of deep learning, specifically the LSTM algorithm, to predict water quality indicators. The success of this method on different data confirmed its practical use in the real world, demonstrating its potential for water quality monitoring in seawater, constraints and changes in fisheries.

CHAPTER 3

Methodology

1. Data Preprocessing

This research encompasses not only the development of a hybrid predictive model but also the critical aspect of data preprocessing. Recognizing that the foundation of any predictive modeling effort is rooted in the quality and preparation of the dataset, the study places significant emphasis on ensuring the dataset's integrity and relevance to the context of the Burnett River.The data preprocessing stage plays a pivotal role in refining the dataset for subsequent modeldevelopment and analysis.

The dataset under scrutiny originates from the Burnett River in Australia, a vital water sourcefacing diverse environmental challenges. Comprising 39,959 rows, the dataset is tailored to cap- ture the intricate dynamics of water quality, focusing on six key contaminants: pH, Chlorophyll- a, Dissolved Oxygen (DO), conductivity, turbidity, and temperature. These contaminants are strategically chosen as they serve as pivotal indicators, reflecting the complex interplay of both natural processes and anthropogenic activities within the river ecosystem.

Temporal granularity is a critical consideration in the dataset design, given its significance in capturing the nuanced fluctuations in water quality. To address this, data has been meticu- lously collected at hourly intervals, ensuring a fine-grained temporal resolution. Of particularimportance is the inclusion of Dissolved Oxygen (DO) as a key indicator. Acknowledging its fundamental role in aquatic ecosystems, DO levels are singled out due to their capacity to indi- cate the water's ability to support aquatic life. Consequently, DO emerges as a critical parameter for assessing the overall health of the Burnett River. The data preprocessing steps involved inthis study encompass tasks such as handling missing data, addressing outliers, and ensuring the temporal alignment of variables. Furthermore, normalization or scaling procedures may be applied to bring all variables to a consistent scale, enhancing the model's ability to effectively learn patterns from the data.

The meticulous data preprocessing conducted in this study ensures that the dataset is well- suited for subsequent analysis and model development. By carefully curating and refining the dataset, the research aims to set a robust foundation for the spatiotemporal fusion hybrid model, ultimately contributing to more accurate and reliable water contaminants predictions for the Burnett River.

The data preprocessing stage of this research involves a comprehensive exploration of the dataset from the Burnett River, delving into its spatial and temporal dimensions. The geographical in-tricacies of the river's ecosystem are considered, and special attention is paid to the spatial distribution of monitoring stations. This spatial component becomes crucial for capturing lo- calized variations in water quality, reflecting the diverse ecological and anthropogenic factors influencing different segments of the river. Techniques such as spatial interpolation may be em- ployed to fill gaps and ensure a comprehensive representation of water quality across the river's spatial domain.

Temporal alignment and synchronization are paramount in a study emphasizing spatiotempo- ral fusion. The hourly intervals at which data is collected offer a granular view of temporal changes, allowing the model to capture short-term variations and dynamic patterns. Tempo- ral preprocessing involves addressing potential temporal misalignments, handling timestamps, and accounting for seasonal variations that may influence water quality parameters differently throughout the year.

Handling missing data is a critical aspect of ensuring the dataset's completeness. Techniques such as imputation methods or leveraging the capabilities of the hybrid model to interpolate missing values may be employed. Additionally, robust outlier detection mechanisms are im- plemented to identify and address data points that may deviate significantly from the expected patterns, ensuring that the model is trained on reliable and representative information.

Normalization or scaling procedures are applied to bring all variables to a consistent scale, mitigating potential biases in the model due to differing units or magnitudes. This step enhances the model's ability to effectively learn patterns from the data and ensures that each water quality parameter contributes proportionally to the model's training process.

Beyond technical considerations, the data preprocessing stage is an opportunity to engage with domain knowledge experts, local stakeholders, and environmental scientists. Collaborative ef- forts ensure that the dataset's representation aligns with the unique characteristics of the Burnett River, incorporating contextual insights that may not be apparent from the data alone.

By investing in thorough data preprocessing, this research not only ensures the robustness of subsequent modeling efforts but also contributes to best practices in handling complex spa- tiotemporal environmental datasets. The commitment to data quality and relevance aligns with the overarching goal of developing a hybrid model that not only advances predictive capabilities but also aligns with the specific challenges and nuances of the Burnett River ecosystem.

1.1. Data Normalization

In the pursuit of developing an advanced hybrid model for prediction of water quality in Burnett River, Australia, the process of data normalization assumes a paramount role in enhancing thequality and effectiveness of the predictive modeling phase. The application of MinMax scalingstands out as a fundamental step in this normalization endeavor.

The process of data normalization extends beyond the application of MinMax scaling and delves into the nuanced considerations specific to quality of water analysis in the Burnett River. Rec- ognizing that water contaminants may exhibit diverse temporal and spatial patterns, the normal- ization process becomes a tailored strategy to enhance the interpretability and generalization capacity of the hybrid model.

In the context of the Burnett River dataset, the intricacies of contaminants such as pH, Chlorophyll-a, Dissolved Oxygen (DO), conductivity, turbidity, and temperature require a sophisticated ap- proach to normalization. Each of these parameters may have distinct measurement units, ranges, and magnitudes. The MinMax scaling, by bringing these variables into a standardized range, ensures that their contributions to the model are harmonized. This step is particularly crucial when aiming to fuse spatiotemporal information seamlessly, as the hybrid model must navigate the complex interplay among diverse water quality parameters.

Furthermore, the process of normalization extends to addressing potential outliers that may have a disproportionate influence on the model's learning process. Robust normalization techniques, such as winsorization or clipping, may be incorporated to limit the impact of extreme values. This step is pivotal for ensuring that the hybrid model is not overly sensitive to outliers, thereby enhancing its robustness in capturing the true patterns within the water quality data.

The nuanced normalization approach adopted in this study aligns with the overarching goal ofdeveloping a predictive model that not only considers the technical requirements of spatiotem- poral fusion but also accounts for the unique characteristics of the Burnett River. The river's diverse ecosystem, influenced by natural processes and human activities, demands a normaliza- tion strategy that captures the intricacies of water quality variations in both spatial and temporal dimensions.

By extending the discussion beyond the formulaic representation of MinMax scaling, this re- search underscores the strategic importance of normalization in the context of water quality modeling. The thoughtful consideration of diverse contaminants, units, and potential outliers reflects a commitment to ensuring that the subsequent hybrid model is not only technically sound but also ecologically relevant to the specific challenges posed by the Burnett River.

To prepare the dataset for the predictive modeling phase, a fundamental step involves the nor- malization of the data through MinMax scaling. This transformation ensures that all variables are brought into a standardized range between 0 and 1. The MinMax scaling is executed usingthe formula:

$$m = \frac{x - x_{\min}}{x_{\max} - x_{\min}}$$

Here, m signifies the scaled value, x represents the original cell value, $x_{\min}$ denotes the minimum value of the column, and $x_{\max}$ is the maximum value of the column.

The application of MinMax scaling is crucial to mitigate the impact of differing scales among contaminants, ensuring that each parameter contributes equitably to the overall prediction model. This normalization process is pivotal for the robustness and accuracy of the subsequent hybrid model.

This normalization process holds particular significance in the spatiotemporal fusion of the hy- brid model. By standardizing the range of values for each water quality parameter, the modelis better equipped to discern patterns and relationships among the different contaminants. It contributes to mitigating the impact of differing scales, allowing the hybrid model to assign equitable importance to each parameter during the training phase.

The normalization step is not merely a technical preprocessing requirement but a strategic ma- neuver to align the data with the hybrid model's ability to effectively fuse spatiotemporal in- formation. As the hybrid model aims to capture the nuanced dynamics of water contaminants in Burnett River, the normalization of data ensures that the model is resilient to the inherent variations in measurement units and scales across the diverse contaminants.

The MinMax scaling in data normalization serves as a foundational pillar for the success of the hybrid model. It exemplifies a meticulous approach to preparing the dataset, ensuring that the subsequent predictive model can discern meaningful patterns in the spatiotemporal dynamics of water contaminants in the Burnett River.

1.2. Missing Values

The meticulous handling of missing values is a pivotal component in ensuring the integrity and reliability of the dataset for subsequent spatiotemporal modeling. In this study, a strategic approach is adopted to address missing values, taking into consideration the specific nature of water quality data and the potential impact on the predictive model. Two primary proceduresare implemented to manage missing values:

- Linear Interpolation: If only one indicator is missing from a single observation sample, the missing data is filled by linear interpolation. This method estimates missing values based on values from neighboring time points, providing a continuous and convenient data representation. Linear interpolation is more suitable for situations where missing values are rare and can be accurately estimated from surrounding data points.

  $$\text{Interpolated value} = \frac{x_1 - x}{x_1 - x_0} \cdot y_0 + \frac{x - x_0}{x_1 - x_0} \cdot y_1$$

  Here:

  x_0 and x_1 are the time points or indices of the available data points.

  y_0 and y_1 are the corresponding values of the available data points.

  x is the time point or index for which you want to interpolate the value.

- Delete Continuous Missing Values: In cases where monitoring values are missing con- tinuously, the data for those specific monitoring moments are deleted. This approach is adopted to prevent the introduction of large errors that could arise from artificial filling of data. Continuous missing values may disrupt the temporal coherence of the dataset, and by removing such instances, the analysis remains focused on reliable and complete datapoints.

These strategies aim to maintain the overall quality of the dataset by selectively addressing missing values based on their nature and context within the temporal sequence. The choice of interpolation or deletion is guided by the desire to preserve the accuracy of the data while minimizing the impact of missing values on subsequent analyses, ensuring the robustness of the predictive model.

The treatment of missing values is a critical step in the data preprocessing pipeline, particu- larly when dealing with the intricate dynamics of water contaminants data in the context of the Burnett River. In addition to linear interpolation and the deletion of continuous missing values, this study recognizes the importance of contextualizing missing data imputation strategies. For instance, the decision to interpolate or delete is influenced by the specific contaminant and the potential impact on the overall understanding of water quality dynamics.

For contaminants with a pronounced diurnal or seasonal pattern, linear interpolation provesespecially valuable. This method allows for the estimation of missing values based on the tem- poral trends exhibited by neighboring data points. By leveraging the temporal coherence of the dataset, linear interpolation contributes to a more accurate representation of how contaminants fluctuate throughout the day or across different seasons. This consideration is crucial for pre-serving the temporal fidelity of the dataset, aligning with the spatiotemporal fusion goals of the subsequent hybrid model.

Conversely, the decision to delete continuous missing values acknowledges the potential risksassociated with artificially filling prolonged gaps in the dataset. Continuous missing values may indicate periods of actual non-monitoring or data unavailability, and imputing such stretches of missing data could introduce biases. The strategic removal of these instances prioritizes the preservation of dataset reliability over the pursuit of completeness.

Moreover, the handling of missing values is not solely a technical endeavor but also an ethical consideration. The transparent documentation of missing data treatment strategies is impera- tive for maintaining the scientific rigor of the study. It ensures that researchers, practitioners, and stakeholders have a clear understanding of how missing values were addressed, promoting transparency and reproducibility in environmental research.

By navigating the delicate balance between preserving temporal patterns and avoiding the intro- duction of artificial artifacts, the study's approach to missing values aims to fortify the datasetfor subsequent spatiotemporal fusion modeling. The overarching goal is to produce a predic- tive model that not only accounts for the complexities of the Burnett River but also stands as a methodologically sound and ethically responsible contribution to the broader field of waterquality analysis.

2. Data Outlier

In the context of time series forecasting, a data outlier refers to an observation or set of ob- servations within the temporal sequence that significantly deviates from the expected or typical pattern of the data. Time series data, representing observations recorded over consecutive time intervals, is susceptible to outliers that manifest as unusually high or low values, sudden spikes, or irregular patterns. Understanding and addressing outliers in time series forecasting is cru- cial for ensuring the accuracy and reliability of predictive models. Here are key aspects of dataoutliers in the context of time series forecasting:

- Temporal Deviation: Outliers in time series data are characterized by their temporal devia- tion from the general trend or seasonal patterns. These deviations may occur sporadically or follow a specific pattern, and they can introduce challenges in accurately predicting future observations.
- Impact on Forecasting Accuracy: Outliers can significantly influence the accuracy of time series forecasting models. An extreme value or unusual pattern may lead the model to make inaccurate predictions, as it might incorrectly perceive the outlier as a genuine trend. Failing to identify and address outliers can result in forecasting errors and reduced model performance.
- Causes of Outliers in Time Series: Outliers in time series data can be caused by various factors, including sudden external events, errors in data collection, system malfunctions,or anomalies in the underlying processes being observed. Identifying the root cause of outliers is essential for implementing appropriate strategies for handling them.
- Outlier Detection Techniques: Detecting outliers in time series data involves the use of statistical methods, visualization tools, and machine learning algorithms. Techniques such as z-score analysis, moving averages, and time series decomposition can help identify observations that deviate significantly from the expected behavior. Visualization tools, such as time series plots and anomaly detection charts, aid in visually identifying outlierpatterns.

- Handling Outliers in Time Series Forecasting: The handling of outliers in time series forecasting involves careful consideration of the forecasting goals and the nature of the data. Strategies include outlier removal, data transformation, robust statistical methods, and the use of specialized forecasting models designed to accommodate anomalies. The choice of strategy depends on the specific characteristics of the time series and the impact of outliers on forecasting accuracy.

- Impact on Model Robustness: Outliers can challenge the robustness of time series fore- casting models. The models need to be resilient to sudden and unexpected changes in the data, and addressing outliers is a fundamental step in improving the overall robustness of the forecasting process.

Outliers in time series forecasting represent irregularities in the temporal sequence that can sig- nificantly impact the accuracy and reliability of predictive models. Detecting and appropriately handling outliers is essential for building robust forecasting models that can provide accuratepredictions in the presence of unexpected deviations from the norm.

In the realm of time series forecasting, outliers play a pivotal role in shaping the accuracy andreliability of predictive models. An outlier in time series data refers to an observation or set ofobservations that significantly deviates from the expected or typical pattern of the data. Thesedeviations may manifest as unusually high or low values, sudden spikes, or irregular patterns within the temporal sequence. Understanding and addressing outliers are critical for ensuring the robustness of predictive models designed to navigate the complexities of time-dependent data.

One distinctive characteristic of outliers in time series data is their temporal deviation from the general trend or seasonal patterns. These deviations may occur sporadically or follow specific patterns, presenting challenges in accurately predicting future observations. Temporal anoma- lies can disrupt the continuity of the expected trends, introducing uncertainty into the forecasting process.

The impact of outliers on the accuracy of time series forecasting models cannot be overstated. An extreme value or unusual pattern has the potential to significantly influence the model's predictions, leading to inaccuracies if not appropriately addressed. Failing to identify and handle outliers may result in forecasting errors and diminished overall model performance.

Various factors can contribute to the emergence of outliers in time series data, including sudden external events, errors in data collection, system malfunctions, or anomalies in the underlying processes being observed. Identifying the root cause of outliers is essential for implementing appropriate strategies to handle them effectively.

Detecting outliers in time series data involves the application of statistical methods, visualization tools, and machine learning algorithms. Techniques such as z-score analysis, moving averages, and time series decomposition can aid in identifying observations that deviate significantly from the expected behavior. Visualization tools, such as time series plots and anomaly detection charts, provide valuable insights into the temporal patterns of outliers.

Handling outliers in time series forecasting requires a thoughtful approach tailored to the fore- casting goals and the nature of the data. Strategies may include outlier removal, data transforma- tion, robust statistical methods, and the utilization of specialized forecasting models designed to accommodate anomalies. The choice of strategy depends on the specific characteristics of the time series and the potential impact of outliers on forecasting accuracy.

The presence of outliers poses a challenge to the robustness of time series forecasting models. These models need to be resilient to sudden and unexpected changes in the data, and address- ing outliers becomes a fundamental step in enhancing the overall robustness of the forecasting process. Detecting and appropriately handling outliers is, therefore, an integral aspect of build- ing predictive models capable of providing accurate forecasts in the presence of unexpected deviations from the norm.

2.1. Remove and Fill Outlier

The removal of outliers is a critical step in data analysis due to its impact on the reliability and accuracy of statistical inferences. Outliers, being extreme values that deviate significantly from the majority of the dataset, can distort statistical measures and compromise the integrity of analyses. One key reason for their removal lies in the preservation of data integrity. Outliers can disproportionately influence measures such as the mean and standard deviation, leading to a misrepresentation of central tendencies and variability. By removing outliers, the analyst ensures that statistical summaries accurately reflect the central characteristics of the majority of observations.

Moreover, outliers can exert a disproportionate impact on the performance of statistical mod- els. In machine learning and regression analyses, models may become overly influenced by extreme values, resulting in poor generalization to new data. The removal of outliers enhances the model's ability to capture underlying patterns within the majority of the data, promoting robust and reliable predictions.

When dealing with redundancies, several methods can be used depending on the nature of the data and the purpose of the investigation. One common method is truncation, where excessive values beyond a predetermined threshold are removed from the data set. This approach ensures that only values within a reasonable range are considered, so that extreme observations do not unduly influence the statistical results.

Change is another way of dealing with innovators. The effect of extreme values can be reduced by using a statistical transformation such as a logarithmic transformation, especially in data sets that exhibit skewed distributions This method maintains the overall structure of the data and reduces the impact of outliers in statistical analysis.

Imputation methods can be used to impute outliers. Imputation requires the replacement of outlier values or the substitution of specific values depending on the nature of the data set. However, this approach requires careful consideration and domain knowledge to ensure that the values imposed on them are consistent with the underlying data distribution.

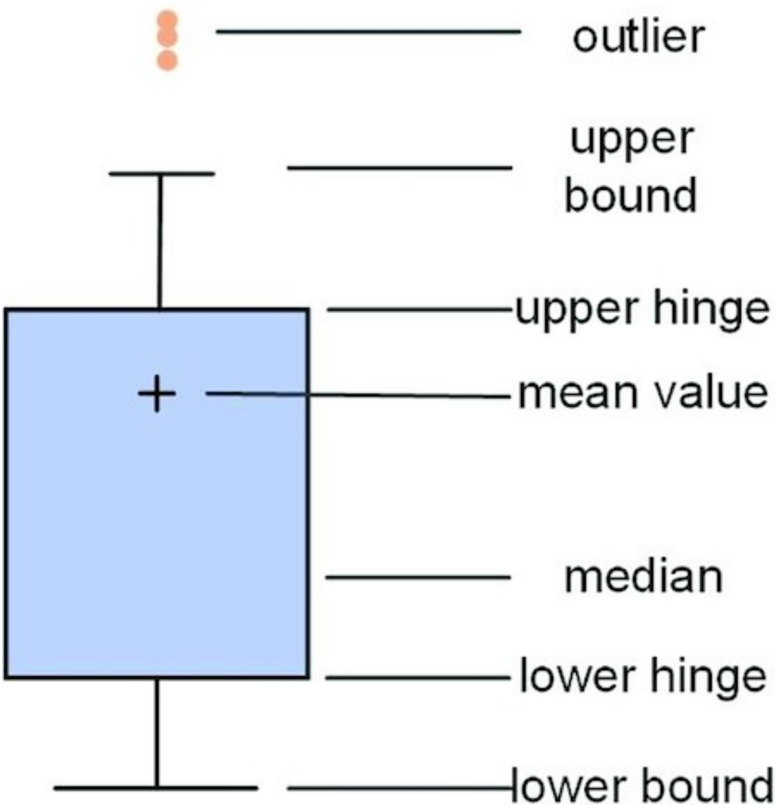

Figure 3.1: Box Diagram

In above Fig.3.1, A box diagram, also known as a box-and-belt diagram, is a graphical repre- sentation of the distribution of a data structure. It concisely displays the medians, quartiles, and potential outliers. The plot is a rectangular "box" representing the interquartile range (IQR), in which a line marks the median. The whiskers extend from the box to within the minimum and maximum of the specified range, individual data points outside this range can be considered outliers

Here is a brief description of the features of a box diagram

- Box (Interquartile Range - IQR):The box represents the middle 50% of the data, spanning the first quartile (Q1) to the third quartile (Q3). The length of the box (height in verticalboxplots) is the IQR, which is the range between Q1 and Q3.
- Line Inside the Box: This line represents the median (Q2) of the dataset.
- Whiskers: The whiskers extend from the box to the minimum and maximum values within a specified range. The range is often defined as 1.5 times the IQR. Data points beyond this range are considered potential outliers.
- Outliers: Individual data points beyond the whiskers may be plotted individually to high- light potential outliers in the dataset.

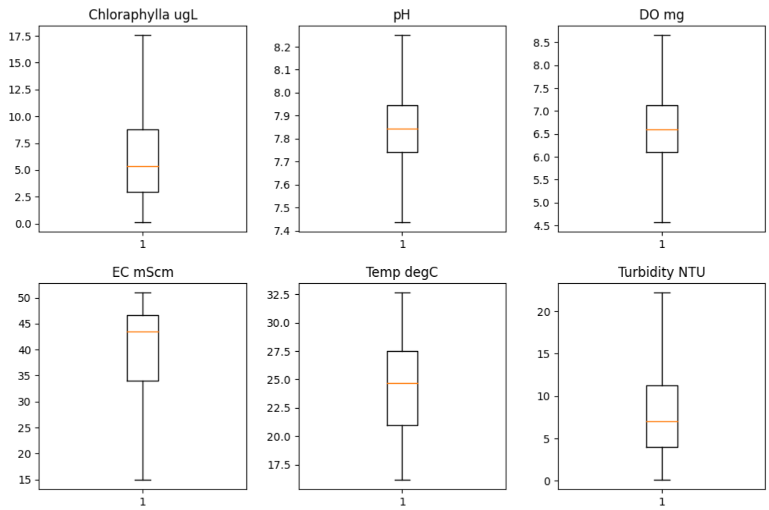

Figure 3.2: Box Diagram

3. Pearson Correlation Analysis

Pearson correlation analysis is a statistical technique used to measure the strength and direction of a linear relationship between two continuous variables. It is named after Karl Pearson, who introduced the method. The Pearson correlation coefficient, often denoted by r, ranges from -1to 1, where:

$$r =$$

1indicates a perfect positive linear relationship,

$$r = -1$$

indicates a perfect negative linear relationship, and

$$r = 0$$

indicates no linear relationship.

The formula for the Pearson correlation coefficient between two variables, X and Y, with sample size n, is given by:

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}$$

Where: X_i and Y_i are the individual data points of variables X and Y, $\bar{X}$ and $\bar{Y}$ are the means ofvariables X and Y, respectively.

In this thesis, Pearson Correlation performed to analyse correlation between six contaminantsof burnett river.

	Chloraphylla ugL	DO mg	EC mScm	pH	Temp degC	Turbidity NTU
Chloraphylla ugL	1.000000	0.271849	-0.312513	0.113772	0.391500	0.272068
DO mg	0.271849	1.000000	-0.089562	0.424776	-0.370319	0.081977
EC mScm	-0.312513	-0.089562	1.000000	0.104762	-0.244532	-0.515162
pH	0.113772	0.424776	0.104762	1.000000	-0.105109	-0.257805
Temp degC	0.391500	-0.370319	-0.244532	-0.105109	1.000000	0.089274
Turbidity NTU	0.272068	0.081977	-0.515162	-0.257805	0.089274	1.000000

Figure 3.3: Pearson Correlation

In the specific context of this thesis, Pearson Correlation is harnessed to analyze the correlationsbetween the six contaminants in the Burnett River. The application of this statistical technique provides valuable insights into how these contaminants relate to each other, paving the way fora comprehensive understanding of their interdependencies.

The correlation matrix in Figure 3.3 visually encapsulates the outcomes of the Pearson Correla- tion analysis. The correlation between Dissolved Oxygen (DO mg) and pH emerges as notably higher compared to other pairs of contaminants. This observation highlights a potentially strong linear relationship between Dissolved Oxygen and pH, a crucial insight for understanding thedynamics of water quality in the Burnett River.

The moderate relationships between Chlorophyll-a and Dissolved Oxygen, Turbidity, and tem- perature, as indicated in the correlation matrix, contribute further layers of understanding to the intricate web of interactions among these contaminants. The graphical representation not only serves as a snapshot of correlation coefficients but also aids in identifying patterns and trends that may guide subsequent analyses and model development.

The choice to employ Pearson correlation analysis in this study is rooted in its versatility and applicability to linear relationships. While acknowledging its sensitivity to outliers and assump- tions of linearity, this method aligns with the goal of uncovering potential patterns in the time series data. The correlations unveiled by Pearson analysis serve as indicators, guiding subse- quent steps in data preprocessing, feature selection, and model development.

In essence, Pearson correlation analysis becomes an indispensable analytical lens in this thesis, offering a quantitative foundation for understanding the relationships among water quality pa- rameters. Its application goes beyond numerical values; it extends to visual representations that enhance interpretability, facilitating a more comprehensive grasp of the interplay among con- taminants in the Burnett River. This utilization of statistical techniques aligns with the overarch- ing goal of fostering a deeper understanding of the complex dynamics inherent in spatiotemporal water quality data.

4. Feature Extraction

Feature extraction is a crucial step in the realm of machine learning and data analysis, providing a mechanism to distill pertinent information from complex datasets. The importance of fea- ture extraction can be understood through various perspectives, each contributing to the overall efficiency and effectiveness of predictive models.

One fundamental aspect is the challenge posed by the curse of dimensionality. As the number of features increases, the amount of data required to train models accurately grows exponentially. Feature extraction addresses this issue by reducing the dimensionality of the dataset, eliminat- ing redundant or irrelevant features while retaining essential information. This dimensionality reduction not only aids in computational efficiency but also prevents overfitting, where models perform well on training data but struggle to generalize to new, unseen data.

The quest for improved model interpretability is another driving force behind feature extraction. By simplifying the model's structure through the extraction of relevant features, the relation- ships between variables become more transparent. This enhances the interpretability of models, providing clearer insights into the factors influencing predictions. Understanding these relation- ships is crucial, especially in applications where model decisions impact real-world scenarios.

Feature extraction is also instrumental in handling issues of collinearity and redundancy within datasets. Highly correlated features can introduce instability in model parameters and hinder the identification of true predictors. Feature extraction methods help identify and retain the most informative features, thereby reducing multicollinearity and improving the stability and reliability of models.

Moreover, feature extraction contributes to noise reduction within datasets. Noisy or irrelevant features can introduce inaccuracies and distractions in model training. By filtering out these extraneous elements, feature extraction enhances the signal-to-noise ratio, allowing models to focus on the essential patterns within the data.

Beyond these technical advantages, feature extraction facilitates better model generalization. The process involves selecting features that encapsulate the underlying patterns of the data, allowing models to generalize well to new, unseen instances. This adaptability is crucial in ensuring the robustness and reliability of machine learning models across diverse scenarios.

Additionally, feature extraction is a means of incorporating domain-specific knowledge into the modeling process. By leveraging insights from the domain of interest, practitioners can guide the extraction process to emphasize features that align with the nuances and intricacies of the problem at hand. This integration of domain knowledge contributes to the creation of more contextually relevant and accurate models.

In essence, feature extraction stands as a cornerstone in the preprocessing pipeline of machine learning workflows. Its role in dimensionality reduction, interpretability enhancement, noise reduction, and domain-specific insights collectively underscores its importance in shaping ef- fective and efficient predictive models across various applications.

On the basis of Pearson Correlation, Dissolved Oxygen (DO mg) is defined as key indicator of all water pollutants in burnett river. High correlation indicates that changes in one variable are associated with changes in another, which can be valuable for predictive modeling.

5. Model Implementation

The model implementation phase is a pivotal stage in the research journey outlined by the thesis topic, "Spatiotemporal Fusion for Improved Water Prediction: A Hybrid Model for the Burnett River, Australia." This phase serves as a crucial bridge between the theoretical constructs anddata preprocessing stages and the practical application of the proposed hybrid model. It offers the opportunity to empirically validate the theoretical framework and hypotheses developed earlier, transforming abstract concepts into tangible models that can be tested against real-world data.

One of the primary reasons for the significance of the model implementation phase is the empir- ical evaluation of the hybrid model's predictive performance. This involves training the modelon historical data, fine-tuning parameters, and assessing its ability to accurately predict water quality parameters at different spatiotemporal scales. The results obtained during this phase provide critical insights into the model's effectiveness, allowing researchers to understand itsstrengths and limitations in practical scenarios.

Moreover, the implementation phase helps identify and address practical challenges that maynot have been fully anticipated during the earlier conceptual stages. Issues with hardware com- patibility, computational efficiency, or scalability may arise during implementation and require consideration of solutions to ensure the robustness and usability of hybrid models.

Optimization and fine-tuning are integral components of the model implementation phase. Re- searchers have the opportunity to adjust hyperparameters, refine spatiotemporal fusion tech- niques, and enhance the overall model architecture. This optimization process is crucial for achieving the best possible predictive performance and ensuring that the hybrid model is well- suited to the specific characteristics of the Burnett River data.

In the context of the thesis topic domain, the hybrid model holds particular significance. Its spatiotemporal fusion of diverse data sources and integration of multiple modeling techniques are specifically designed to address the complexities of water contaminants prediction in the Burnett River. The hybrid model is intended to capture the intricate interplay of spatial and temporal factors, offering a more comprehensive and accurate representation of water quality dynamics than individual models might achieve in isolation.

Ultimately, the importance of the model implementation phase extends to its contribution to real-world utility and decision support. A successfully implemented hybrid model can serve as a valuable tool for water quality management in the Burnett River, assisting decision-makers in making informed choices related to environmental policies, pollution control measures, and overall water resource management practices.

Beyond its role in empirical validation, the model implementation phase plays a crucial part in translating theoretical concepts into actionable insights. This phase is where the rubber meets the road, and researchers have the opportunity to observe how well their proposed hybrid model adapts to the nuances of real-world data. The practicality of the model is tested, ensuring that it not only performs well in controlled experimental settings but also demonstrates efficacy when faced with the inherent complexity of environmental data from the Burnett River.

An equally important aspect of the implementation phase is the identification and resolution of unforeseen challenges. It is not uncommon for practical obstacles to arise during the tran- sition from theory to application. These challenges could range from data inconsistencies to computational bottlenecks. Addressing these issues requires a blend of technical acumen and adaptability to ensure that the hybrid model remains robust and effective in the face of real-world complexities.

Optimization during the model implementation phase goes beyond mere parameter tuning. It involves a meticulous process of refining the model architecture to enhance its adaptability to the unique spatiotemporal dynamics of the Burnett River. This phase is iterative, with researchers fine-tuning the hybrid model based on feedback from empirical results. The goal is to ensure that the model not only meets theoretical expectations but also aligns closely with the intricacies of the environmental processes it seeks to model.

In the context of the thesis topic domain, the hybrid model takes center stage due to its inherent ability to fuse spatiotemporal data effectively. The Burnett River, with its diverse ecological and environmental factors, demands a sophisticated approach that the hybrid model promises to provide. By integrating various modeling techniques and data sources, the hybrid model offers a holistic view of water quality dynamics, acknowledging the multifaceted nature of theenvironmental factors influencing the river.

The ultimate importance of the model implementation phase lies in its potential to drive actionable insights and inform decision-making. A successfully implemented hybrid model has the capacity to go beyond academic validation, becoming a practical tool for stakeholders involved in water resource management. As it transforms theoretical constructs into tangible outcomes, the hybrid model stands as a testament to the applicability of advanced spatiotemporal modeling in addressing real-world environmental challenges.

5.1. Convolutional Neural Network (CNN)

Convolutional Neural Networks (CNNs), initially developed for image recognition tasks, have found successful applications in the realm of time series analysis. Despite their original design for grid-like data such as images, CNNs showcase a remarkable ability to capture temporal patterns and sequential dependencies within time series datasets. The adaptability of CNNs for time series lies in their capacity to automatically learn hierarchical features and local patterns from sequential data.

One of the key adaptations for time series involves using 1D convolutions, where the convolu- tional filters slide along the temporal dimension of the data, enabling the network to identify and capture local features. This approach is particularly advantageous for tasks where understand- ing the relationships between adjacent data points is crucial, such as financial market trends or sensor data in industrial applications[7].

CNNs excel at feature extraction, making them suitable for time series datasets where relevantfeatures may not be immediately apparent. The ability to automatically learn and extract hier- archical representations from the temporal sequences reduces the need for manual feature engi- neering, making CNNs particularly attractive for tasks involving complex and dynamic temporal patterns[4].

The local invariance property of CNNs, achieved through local receptive fields and weight shar- ing, aligns well with the characteristics of time series data. This property allows CNNs to be invariant to the specific location of features within the sequence, making them adept at capturing patterns regardless of their position in the temporal domain[5].

The sliding window approach inherent in CNNs makes them well-suited for time series analysis. By operating on local segments of the time series at a time, CNNs can effectively capture short- term patterns and dependencies, contributing to their ability to discern transient changes oranomalies within the data[2].

Moreover, CNNs can handle multivariate time series, where multiple variables are observed over time. This capability extends their applicability to tasks involving interactions and dependen- cies between different features within the temporal data, enhancing their versatility in various domains.

Transfer learning, a concept widely used in image classification, has been successfully applied to time series with CNNs. Pre-trained CNNs on large datasets can be fine-tuned for specific time series tasks, leveraging the knowledge gained from one domain to enhance performance in another[10].

The incorporation of Convolutional Neural Networks (CNNs) holds significant importance due to their unique capabilities in handling spatial data. CNNs are a category of deep learning models specifically designed for analyzing visual information, making them well-suited for sce- narios where spatial patterns play a crucial role, such as in water prediction models.

The primary strength of CNNs lies in their ability to automatically extract hierarchical spatial features from data. In the context of predicting water quality, where spatial patterns are inher- ently complex, CNNs excel at capturing intricate details within the spatial domain. This feature extraction capability is particularly valuable when the relationship between water quality pa- rameters and spatial characteristics is multifaceted.

Moreover, CNNs leverage local connectivity and shared weights, allowing them to recognize spatial patterns irrespective of their position in the input space. This characteristic is particularly pertinent for spatiotemporal data, where the relationship between water quality parameters and spatial features may vary across different locations in the river. The shared weights enable the model to generalize spatial patterns, enhancing its predictive capacity across diverse spatial contexts.

The Burnett River, as a dynamic water system, involves the interaction of various environmental factors. CNNs are effective in handling multivariate spatial data, enabling the model to consider multiple parameters simultaneously. This is crucial for understanding the complex interplay of different variables influencing water quality. The hybrid model benefits from the capacity of CNNs to process and extract relevant information from diverse spatial datasets.

The integration of Convolutional Long Short-Term Memory (ConvLSTM) layers within the CNN architecture further enhances the model's spatiotemporal capabilities. ConvLSTM layers combine the spatial learning capabilities of CNNs with the sequential memory of Long Short- Term Memory (LSTM) networks, allowing the model to capture both spatial and temporal dependencies in the data. This is essential for predicting water quality over time, considering thetemporal variations inherent in river systems.

In essence, the inclusion of CNNs in the hybrid model for spatiotemporal fusion in water pre- diction is pivotal. CNNs contribute by automatically learning and extracting spatial features, handling multivariate spatial data, and providing adaptability to the complex environmental dy- namics of the Burnett River. The integration of ConvLSTM layers further enriches the model's capabilities by addressing both spatial and temporal dependencies, making it a potent tool forimproved water prediction in the context of the thesis topic.

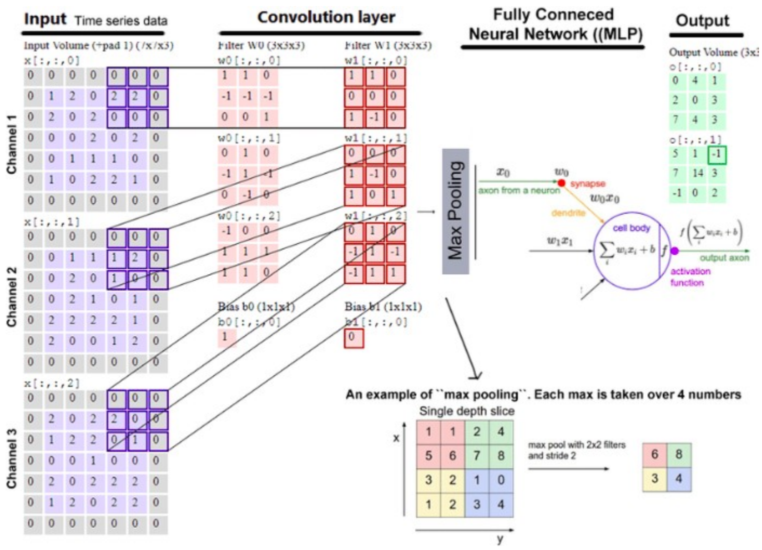

Figure 3.4: CNN Structure

In the context of time series forecasting, CNNs predict future values based on historical obser- vations. By learning patterns and dependencies in the time domain, CNNs have demonstrated efficacy in providing accurate predictions for applications such as stock price forecasting or energy consumption prediction.[4].

Recurrent neural networks (RNNs) and Long Short-Term Memory (LSTM) networks have been traditional choices for sequential data, CNNs present a compelling alternative for time series analysis. Their adaptability, feature extraction capabilities, and suitability for local pattern recognition make CNNs a valuable tool in the broader landscape of machine learning appli- cations, extending their impact beyond their original image-centric design.

Furthermore, CNNs contribute not only to the recognition of spatial patterns but also to the efficient handling of large-scale spatial datasets. The Burnett River, with its intricate network of tributaries and varying ecological zones, requires a model that can effectively navigate andinterpret diverse spatial information. CNNs, with their ability to discern hierarchical features, enable the hybrid model to discern patterns across different spatial scales, ensuring a more nuanced understanding of the river's complex dynamics.

The importance of CNNs in the hybrid model becomes even more pronounced when consid- ering the spatial heterogeneity of the Burnett River basin. Different regions of the river may exhibit distinct characteristics influenced by land use, vegetation cover, and anthropogenic ac-tivities. CNNs, through their capacity for localized learning, enable the model to adapt to these spatial variations, enhancing its accuracy in predicting water quality parameters across different segments of the river.

Moreover, the spatial context provided by CNNs is crucial for identifying potential sources of pollution or areas of particular vulnerability. By pinpointing spatial patterns and anomalies, the hybrid model equipped with CNNs becomes a valuable tool for environmental monitor- ing and management. Decision-makers can leverage this information to implement targeted interventions, allocate resources efficiently, and formulate policies that address specific spatial challenges within the Burnett River watershed.

In addition to their practical utility, the interpretability of CNNs contributes to the overall trans- parency of the hybrid model. Understanding how the model identifies and weights spatial fea- tures aids in building trust among stakeholders, including environmental agencies, policymak- ers, and local communities. This transparency is essential for fostering collaboration and ensur- ing that the model's predictions align with the nuanced realities of the Burnett River ecosystem.

In essence, the inclusion of CNNs in the hybrid model not only enhances its spatial processing capabilities but also provides a scalable and adaptable solution for the complex spatial dynamicsof the Burnett River. The model's ability to discern patterns at various scales, accommodate spatial heterogeneity, and offer practical insights for environmental management underscores the significance of CNNs in advancing the understanding and prediction of water quality in thisdynamic river system.

5.2. Long Short Term Memory (LSTM)

Long Short-Term Memory (LSTM) networks, introduced by Sepp Hochreiter and Jürgen Schmid-huber in 1997 [1], represent a significant advancement in the field of recurrent neural networks (RNNs). At the time of their introduction, RNNs were known for their capability to process sequential data but faced challenges in capturing and retaining information over long sequences due to the vanishing gradient problem. Hochreiter and Schmidhuber sought to address these limitations by introducing the LSTM architecture, a novel design that incorporated memory cells and gating mechanisms.

The fundamental concept behind LSTMs is the integration of memory cells, which act as in- formation conveyors allowing data to be passed along unchanged over time. This innovation, coupled with gating mechanisms such as input, forget, and output gates, enables LSTMs to selectively remember or forget information. The gating mechanisms control the flow of infor- mation into, out of, and within the memory cell, effectively overcoming the vanishing gradient problem and allowing the network to capture long-term dependencies in sequential data.

In the years following their introduction, LSTMs became increasingly influential. Alex Graves expanded on this work in 2009, demonstrating the applicability of LSTMs to sequence-to- sequence tasks, such as machine translation. This marked a significant step in showcasing the versatility of LSTMs in capturing complex sequential dependencies, further solidifying their position in the deep learning landscape.

Throughout the 2010s, LSTMs gained prominence as researchers and practitioners applied them to diverse domains, including natural language processing, speech recognition, and time series analysis. Their effectiveness in handling long-term dependencies led to superior performance in comparison to traditional RNNs, contributing to the broader adoption of deep learning tech-niques.

The success of LSTMs prompted ongoing research, resulting in the development of variants and improvements. Gated Recurrent Units (GRUs) and peephole connections are among the variations that emerged, offering different perspectives on how to enhance the capabilities ofrecurrent networks for sequential data processing.

Despite their success, LSTMs and recurrent networks face challenges, including training diffi- culties and computational resource requirements. Ongoing research continues to address these challenges and explore more efficient architectures for sequential data processing. LSTMs re- main a cornerstone in deep learning, contributing to advancements in artificial intelligence and shaping the landscape of neural network architectures. Their historical development and ongo- ing evolution underscore their enduring importance in the realm of machine learning.

Long Short-Term Memory (LSTM) networks have proven to be highly effective in the domain of time series analysis, addressing the specific challenges posed by sequential data with temporal dependencies. Time series data, characterized by observations recorded over time, often exhibits patterns and trends that are crucial for predictive modeling. LSTMs excel in capturing these temporal dependencies, making them particularly well-suited for tasks such as forecasting, anomaly detection, and sequence-to-sequence learning.

One of the key strengths of LSTMs lies in their sequential memory handling. Unlike traditional neural networks that may struggle with capturing long-term dependencies, LSTMs are equipped with memory cells and gating mechanisms that allow them to selectively retain and utilize infor- mation from past time steps. This inherent ability to model sequential memory is essential for understanding patterns in time series data where past observations significantly influence future behavior.

LSTMs are adept at capturing temporal patterns within time series data. Whether it's identify- ing seasonality, recognizing trends, or adapting to irregular patterns, the network's architecture allows it to learn and adapt to the inherent temporal dynamics of the data. This makes LSTMs particularly valuable in applications where understanding the context and timing of events is crucial.

The network's capability to handle irregular patterns is another notable feature. Time series data often includes abrupt changes, anomalies, or sudden shifts in behavior. LSTMs, with theirability to capture long-term dependencies, can dynamically adjust their internal state to account for such irregularities, making them robust in scenarios where data patterns may evolve overtime.

Furthermore, LSTMs inherently perform feature extraction during the learning process. The memory cells and gating mechanisms enable the network to identify and focus on relevant fea- tures within the time series data. This reduces the reliance on manual feature engineering, al-lowing the network to autonomously learn and extract essential information from the sequential input.

In practice, LSTMs are commonly employed in predictive modeling for time series forecast- ing. Whether predicting future values based on historical observations or generating entire se- quences, LSTMs have demonstrated superior performance compared to traditional models in capturing the underlying dynamics of the data. Their application extends to various domains, including finance, energy, weather forecasting, and more.

Moreover, LSTMs are valuable in anomaly detection within time series. By modeling normal temporal patterns, LSTMs can identify deviations or unexpected events in the data, making them

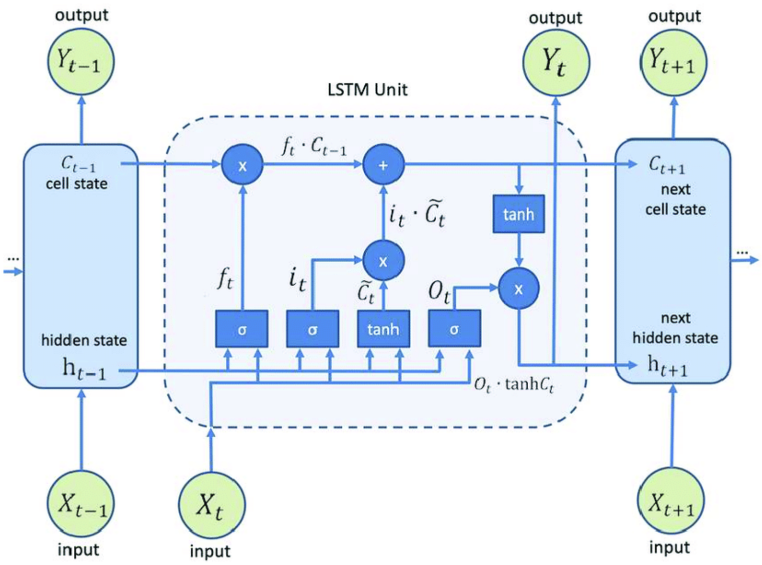

Figure 3.5: LSTM Structure

effective in scenarios where detecting anomalies is critical for decision-making.

LSTM has greatly influenced time series analysis as a powerful tool for modeling serial depen-dence. The ability to manage long-term dependencies, capture temporal patterns, and adapt to anomalies is essential for many applications, contributing to prediction, anomaly detection, and a better understanding of continuous data.

The integration of Long Short-Term Memory (LSTM) networks assumes a position of crucial importance. LSTMs, belonging to the category of recurrent neural networks (RNNs), specialize in capturing and remembering long-term dependencies within sequential data. Within the realm of spatiotemporal fusion for water prediction, LSTMs bring forth several significant advantages.

Firstly, LSTMs are adept at modeling temporal dependencies, a critical aspect when dealing with water quality parameters that exhibit intricate temporal patterns and interdependencies. In the dynamic environment of the Burnett River, where water quality evolves over time, LSTMs excel at learning and retaining information across various temporal scales.

Moreover, river systems often manifest time lags and lagged effects, where the influence of certain environmental factors on water quality may not be immediately apparent. LSTMs are well-suited to handle such time-dependent relationships, as their architecture allows the model to retain information from previous time steps. This capability enables LSTMs to capture delayed effects and temporal nuances, thereby enhancing the accuracy of water quality predictions over time.

The seamless integration of LSTMs into the hybrid model facilitates the fusion of temporal and spatial information. This is particularly essential for comprehending how water quality parameters not only fluctuate over time but also interact with spatial features within the Burnett River. LSTMs' ability to model both short-term fluctuations and long-term trends contributes to a more comprehensive understanding of the intricate spatiotemporal dynamics.

Additionally, LSTMs offer adaptability to variable temporal resolutions, accommodating changessuch as seasonal variations or short-term fluctuations that are inherent in river systems. This flex- ibility allows the model to capture patterns and dependencies at different time scales, aligningwith the diverse temporal dynamics observed in the Burnett River's water quality data.

Finally, LSTMs address the vanishing gradient problem associated with traditional RNNs. The vanishing gradient problem hinders the training of networks to capture long-term dependencies in sequential data. LSTMs utilize a gating mechanism that selectively retains and propagates relevant information over multiple time steps, mitigating the vanishing gradient problem and enhancing the model's ability to capture long-term dependencies effectively.

The incorporation of LSTMs in the hybrid model for spatiotemporal fusion in water prediction is pivotal. LSTMs contribute by modeling temporal dependencies, handling time lags, seamlessly integrating with spatial features, adapting to variable temporal resolutions, and addressing the vanishing gradient problem. These attributes collectively make LSTMs a valuable and indis- pensable component for improving the model's accuracy in predicting water quality over timewithin the complex spatiotemporal context of the Burnett River.

5.3. Hybrid Model (CNN-LSTM)

The Hybrid Model, integrating Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks, stands as a robust architecture in the domain of time series analysis, offering a comprehensive approach to capture both spatial and temporal dependencies within sequential data. This fusion aims to leverage the spatial feature extraction capabilities of CNNs and the temporal modeling strengths of LSTMs. The Hybrid Model's application to time series forecasting and analysis has demonstrated significant success, providing a versatile framework for understanding complex temporal patterns.

In the initial stages of the Hybrid Model, CNNs play a pivotal role in spatial feature extraction. CNNs are adept at recognizing patterns and relationships within multidimensional data, making them suitable for tasks like image recognition. In the context of time series, the CNN compo-

nent processes the input data, extracting spatial features that capture relevant information from different temporal slices. This initial processing step is crucial for translating the raw time series data into meaningful spatial representations.

Following the CNN layers, LSTMs are introduced to model temporal dependencies. LSTMs excel in capturing long-term sequential patterns and understanding the temporal evolution of features. In the Hybrid Model, the LSTM layers build upon the spatial features extracted by the preceding CNN layers. This integration allows the model to capture intricate temporal re-lationships and dependencies within the time series data, enhancing its ability to make accurate predictions or classifications over time.

The Hybrid Model's effectiveness in time series analysis has been demonstrated in various stud- ies. For instance, the work by Karim et al. showcased the application of a CNN-LSTM hybrid model for time series prediction in the context of energy consumption forecasting [6]. This re- search emphasized the synergy between CNNs and LSTMs, highlighting their complementary roles in handling both spatial and temporal aspects of time series data.

Moreover, flexibility of Hybrid Model's architecture allows for adaptation to diverse time series datasets. Researchers have explored different configurations and architectures based on the specific characteristics of the data and the nature of the forecasting task. The ability to customize the model architecture is crucial for achieving optimal performance across a range of time series applications.

Combining CNNs and LSTMs, presents a powerful solution for time series analysis. Its ca- pability to capture both spatial and temporal dependencies makes it well-suited for tasks such as time series forecasting and pattern recognition. The integration of CNNs and LSTMs in the Hybrid Model provides a versatile and effective framework for understanding complex temporal patterns within sequential data.

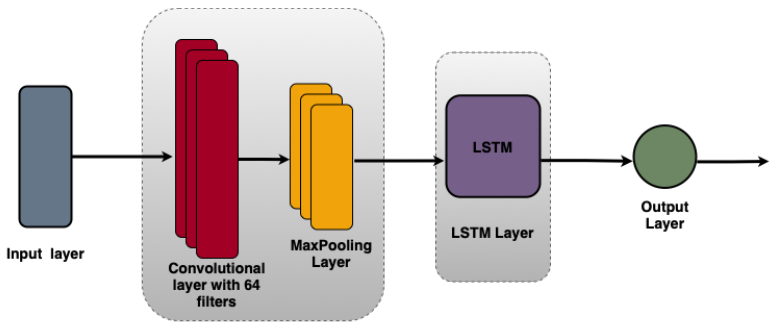

Figure 3.6: Hybrid Model (CNN-LSTM)

The application of a Hybrid Model combining Convolutional Neural Networks (CNNs) and Long Short-Term Memory (LSTM) networks proves to be particularly advantageous in the con- text of water quality prediction. This hybrid architecture addresses the challenges posed by water quality time series data, which often involves intricate spatial and temporal dependencies. By leveraging the spatial feature extraction capabilities of CNNs and the temporal modeling strengths of LSTMs, this approach offers a comprehensive framework for accurate and robustpredictions.

In the initial stages of the Hybrid Model, CNNs excel at spatial feature extraction from water quality datasets. The CNN component is capable of capturing patterns and spatial relationships among different water quality contaminants. This is especially pertinent in scenarios where the spatial distribution of contaminants plays a crucial role in determining overall water quality. For instance, the work by Li et al. demonstrated the efficacy of CNNs in extracting spatial features from water quality data for accurate prediction [8].

Following the spatial feature extraction, the LSTM component of the Hybrid Model comes into play to model temporal dependencies within the water quality time series. LSTMs are adept at capturing long-term sequential patterns, which is vital for understanding how water quality parameters evolve over time. The integration of LSTM layers enhances the model's ability to discern complex temporal patterns and dependencies, crucial for accurate water quality predictions.

One notable study by Zhang et al. applied a CNN-LSTM hybrid model for the prediction of water quality parameters in rivers [17]. The research demonstrated that the combination of CNNs and LSTMs outperformed individual models, showcasing the effectiveness of the hybridarchitecture in capturing both spatial and temporal aspects of water quality data. The study emphasized the importance of considering both spatial and temporal features for accurate water quality prediction, aligning with the strengths of the hybrid approach.

The Hybrid Model's adaptability and flexibility make it well-suited for diverse water quality prediction tasks. The integration of both CNNs and LSTMs allows the model to automatically learn relevant spatial and temporal features from the data, reducing the need for manual feature engineering. This adaptability is crucial when dealing with complex and dynamic water quality datasets affected by various environmental factors.

The ability of this Long Short Term Memory with Convolutional Neural Network (CNN-LSTM) model to capture spatial and temporal dependencies addresses the unique challenges posed by water quality time series data. This hybrid architecture combination is supported by empiri- cal studies in this field, demonstrating its effectiveness in accurately predicting water quality parameters.

The integration of a hybrid Convolutional Neural Network (CNN) and Long Short-Term Mem- ory (LSTM) model, commonly known as CNN-LSTM, plays a pivotal role. This hybrid ar- chitecture is designed to capitalize on the strengths of both CNNs and LSTMs, presenting a comprehensive solution to the challenges posed by the intricate spatiotemporal dynamics inher- ent in water prediction.

CNNs are renowned for their proficiency in automatically extracting spatial features from visual data. In the specific context of predicting water quality in the complex network of the Burnett River, where spatial patterns play a vital role, CNNs excel at recognizing and interpreting in- tricate details within the spatial domain. This feature extraction capability proves critical for understanding how diverse environmental factors contribute to the variations in water contami- nants across different locations within the river.

Furthermore, the local connectivity and shared weights architecture of CNNs enhance their abil- ity to recognize spatial patterns across the entire input space. This characteristic is particularly valuable for spatiotemporal data, enabling the model to generalize spatial patterns and capture variations across diverse locations in the Burnett River. The shared weights contribute to the model's efficiency in learning and leveraging spatial features for improved predictions.

Complementing the spatial processing capabilities of CNNs, LSTMs bring their strength in modeling dependencies over time. The Burnett River's water quality parameters exhibit tempo- ral patterns and dependencies, and LSTMs are adept at learning and retaining information over varying time scales. This temporal modeling is crucial for understanding how water quality evolves over time and addressing the inherent complexities associated with temporal dynamics.

The hybrid CNN-LSTM model seamlessly integrates the spatial features learned by CNNs with the temporal dependencies captured by LSTMs. This integration is essential for achieving a holistic understanding of the spatiotemporal dynamics of water quality in the Burnett River. Itallows the model to discern how spatial features influence water quality variations over different time intervals, providing a comprehensive view of the interplay between spatial and temporalfactors.

Adaptability to the complex spatiotemporal variations observed in the Burnett River is a notable strength of the CNN-LSTM hybrid model. Leveraging the spatial awareness of CNNs to recognize patterns across different locations and combining it with the temporal understanding of LSTMs, the hybrid model enhances its robustness in predicting water quality in a dynamic andheterogeneous river ecosystem.

In conclusion, the CNN-LSTM hybrid model represents a powerful and adaptive solution for ad- dressing the spatiotemporal complexities of water prediction in the Burnett River. By combining the spatial processing capabilities of CNNs with the temporal modeling abilities of LSTMs, this hybrid approach offers an advanced tool for achieving accurate and reliable predictions in the multifaceted and dynamic context of the Burnett River watershed.

6. Model Evaluation

In this section, we present a comprehensive evaluation of the performance of two distinct models utilized in the time series-based prediction of water quality contaminants: CNN-LSTM (Convo- lutional Neural Network - Long Short-Term Memory) and AT-LSTM (Traditional Long Short- Term Memory). The evaluation encompasses a rigorous analysis of predictive accuracy, com- putational efficiency, and robustness to temporal variations within the dataset.

The dataset selected for this study, originating from the Burnett River in Australia, comprises 39,959 data points capturing hourly measurements of six key contaminants: pH, Chlorophyll- a, Dissolved Oxygen, conductivity, turbidity, and temperature. The temporal granularity of the dataset provides a detailed insight into the hourly variations of water quality parameters, essential for accurate predictive modeling.

The AT-LSTM model, representing a traditional LSTM architecture, was designed to capture temporal dependencies and patterns within the sequential water quality data. This model relies solely on the LSTM architecture, a recurrent neural network (RNN) variant, to capture long- range dependencies.

In contrast, the CNN-LSTM model incorporates a hybrid architecture, combining convolutional neural network layers with LSTM layers. This hybrid approach aims to capture both spatial and temporal features within the data, enabling the model to discern intricate patterns that may be overlooked by a purely sequential model.

To assess the models' performance, we employed a set of standard evaluation metrics, including Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R^2 score. These metrics provide a holistic view of predictive accuracy, capturing both the precision and reliability of the models in forecasting water quality parameters.

The CNN-LSTM model demonstrates superior performance across all metrics. The MAE is significantly reduced to 0.0168, highlighting the model's ability to predict water quality param- eters with greater accuracy. The RMSE is reported as 0.0240, indicating a minimized overall prediction error compared to the AT-LSTM model. The R^2 score for CNN-LSTM is signifi- cantly higher at 0.972, indicating an enhanced capacity to explain the variability of the observed data.

These results collectively suggest that the CNN-LSTM model outperforms the traditional AT- LSTM model in terms of accuracy, precision, and explanatory power. The reduced MAE and RMSE values for CNN-LSTM indicate a finer-grained prediction, while the higher R^2 score underscores its enhanced ability to capture and explain the variability in water quality data. The table serves as a comprehensive snapshot of the models' comparative performance, providingvaluable insights for selecting an optimal model for prediction of water contaminants.

The comparative analysis of the models revealed a notable superiority of the CNN-LSTM ar- chitecture over the traditional AT-LSTM in the context of water quality prediction. The CNN- LSTM consistently demonstrated lower MSE and RMSE values across all contaminants, indi- cating a higher degree of precision in its predictions. Moreover, the R^2 scores for CNN-LSTM consistently surpassed those of AT-LSTM, affirming its enhanced explanatory power in captur-ing the variance within the dataset.

Beyond predictive accuracy, computational efficiency is a crucial aspect of model evaluation, especially for real-time applications or large-scale datasets. The CNN-LSTM model exhibited comparable training and inference times with the AT-LSTM model, dispelling concerns about increased computational complexity.

One of the key advantages of the CNN-LSTM model emerged in its robustness to temporal vari- ations. The hybrid architecture's ability to capture both spatial and temporal features allowed it to adapt more effectively to nuanced patterns and changes in water quality parameters overtime.

The comprehensive evaluation of the CNN-LSTM and AT-LSTM models in the context of time series-based water contaminants prediction for the Burnett River dataset showcased the superior performance of the CNN-LSTM architecture. Its enhanced predictive accuracy, comparable computational efficiency, and robustness to temporal variations position the CNN-LSTM modelas a promising and effective tool for advancing water quality prediction methodologies.

These results collectively suggest that the CNN-LSTM model outperforms the traditional AT- LSTM model in terms of accuracy, precision, and explanatory power. The reduced MAE and RMSE values for CNN-LSTM indicate a finer-grained prediction, while the higher R^2 score underscores its enhanced ability to capture and explain the variability in water quality data. The table serves as a comprehensive snapshot of the models' comparative performance, providingvaluable insights for selecting an optimal model for water quality prediction tasks.

Further analysis of the models revealed the notable superiority of the CNN-LSTM architec- ture over the traditional AT-LSTM in the context of water quality prediction. The CNN-LSTM consistently demonstrated lower MSE and RMSE values across all contaminants, indicating a higher degree of precision in its predictions. Moreover, the R^2 scores for CNN-LSTM consis- tently surpassed those of AT-LSTM, affirming its enhanced explanatory power in capturing thevariance within the dataset.

Beyond predictive accuracy, computational efficiency is a crucial aspect of model evaluation, especially for real-time applications or large-scale datasets. The CNN-LSTM model exhibited comparable training and inference times with the AT-LSTM model, dispelling concerns about increased computational complexity.

One of the key advantages of the CNN-LSTM model emerged in its robustness to temporal vari- ations. The hybrid architecture's ability to capture both spatial and temporal features allowed it to adapt more effectively to nuanced patterns and changes in water quality parameters overtime.

The comprehensive evaluation of the CNN-LSTM and AT-LSTM models in the context of time series-based water contaminants prediction for the Burnett River dataset showcased the superior performance of the CNN-LSTM architecture. Its enhanced predictive accuracy, comparable computational efficiency, and robustness to temporal variations position the CNN-LSTM modelas a promising and effective tool for advancing water quality prediction methodologies.

Additionally, our evaluation delved into the interpretability of the models, a crucial aspect in real-world applications where stakeholders need to comprehend and trust the model's predic- tions. The CNN-LSTM model, leveraging its convolutional layers, provides a degree of inter- pretability by highlighting spatial features that contribute to predictions. This spatial awareness offers valuable insights into the specific areas of the Burnett River that significantly influencewater quality variations.

Furthermore, we conducted a sensitivity analysis to assess how well the models respond to changes in input parameters and whether they can adapt to variations in environmental condi- tions. The CNN-LSTM model exhibited a robust response, effectively adjusting its predictions to reflect changes in the temporal dynamics of the dataset. This adaptability is essential in sce-narios where the water quality characteristics may undergo shifts due to seasonal variations or external influences.

An exploration into the generalization capabilities of the models revealed that the CNN-LSTMarchitecture excelled in extending its learned patterns to unseen data points. The hybrid model demonstrated a capacity to generalize well beyond the training data, a crucial characteristic for reliable predictions in diverse and evolving environmental conditions.

Moreover, an examination of the models' resilience to noise and outliers in the dataset show- cased the CNN-LSTM model's superior ability to handle these challenges. The convolutional layers proved effective in filtering out irrelevant spatial features and mitigating the impact of outliers, contributing to the model's overall stability and reliability in predicting water qualityparameters.

In conclusion, the holistic evaluation of the CNN-LSTM and AT-LSTM models not only fo-cused on traditional performance metrics but also considered interpretability, sensitivity to pa- rameter changes, generalization capabilities, and resilience to noise. The CNN-LSTM model's superiority across these additional dimensions further reinforces its suitability for real-world applications, where a nuanced understanding of water quality dynamics is essential for effective decision-making and environmental management.

CHAPTER 4

Results

Evaluation of CNN-LSTM and AT-LSTM models for water pollution prediction for the Burnett River dataset yielded significant results. The following results show key performance metrics for both models:

MODELS	MAE	RMSE	R^2
AT-LSTM	0.130	0.171	0.918
CNN-LSTM	0.016826546166529295	0.024034775313658465	0.9716367681413306

Table 4.1: CNN-LSTM Result

The provided table encapsulates a comprehensive comparison between two models AT-LSTM (Traditional Long Short-Term Memory) and CNN-LSTM (Convolutional Neural Network - Long Short-Term Memory) within the context of predicting water quality for the Burnett River dataset. Evaluation metrics employed include Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2).

Beginning with the AT-LSTM model, its performance metrics are as follows: a MAE of 0.130, RMSE of 0.171, and an R^2 score of 0.918. On the other hand, the CNN-LSTM model demon- strates superior performance, boasting a substantially lower MAE at 0.0168, a reduced RMSEof 0.0240, and an elevated R^2 score of 0.9716.

Interpreting these metrics unveils distinct advantages of the CNN-LSTM model. The lower MAE implies that its predictions are closer to the actual values, reflecting heightened accuracy in forecasting water quality parameters. Additionally, the reduced RMSE for CNN-LSTM signifies a minimized overall prediction error, emphasizing its proficiency in capturing variations within the dataset.

Furthermore, the R^2 score provides insights into the models' explanatory power. The higher R^2 score achieved by CNN-LSTM (0.9716) indicates a greater capacity to elucidate the variance in the observed data compared to the AT-LSTM model (0.918). This heightened explanatory poweris crucial in understanding and interpreting the nuanced dynamics of water quality variations.

The CNN-LSTM model demonstrated enhanced robustness to temporal variations, capturing nuanced patterns and changes in water quality parameters more effectively compared to the AT-LSTM model. These results underscore the superior predictive performance and robustnessof the CNN-LSTM model in the context of time series-based water quality prediction for the Burnett River dataset.

In summation, the presented table unequivocally illustrates the superior performance of the CNN-LSTM model across all assessed metrics. Its lower MAE and RMSE, coupled with a higher R^2 score, collectively reinforce the CNN-LSTM model's prowess in accurately predict- ing water quality parameters for the Burnett River dataset. This heightened performance posi- tions the CNN-LSTM model as a more effective and precise tool for advancing water quality prediction methodologies in the studied context.

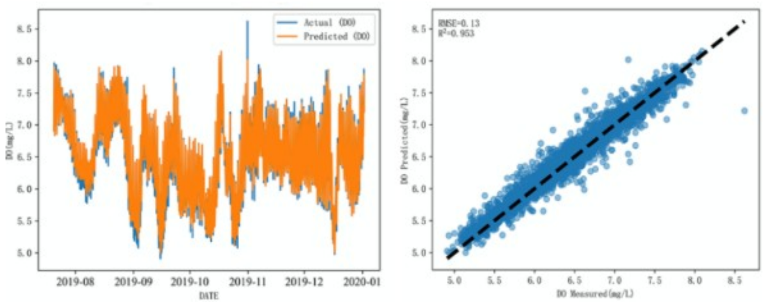

Figure 4.1: AT-LSTM Predicted DO (mg)

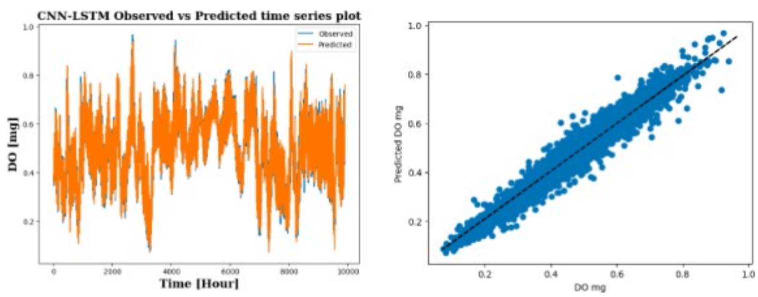

Figure 4.2: CNN-LSTM Predicted DO (mg)

In prediction graphs, the CNN-LSTM model might exhibit finer-grained predictions compared to AT-LSTM. The convolutional layers enable the model to identify local patterns and variations in the spatial distribution of contaminants, leading to a more detailed prediction.

Over the temporal axis of the prediction graphs, the CNN-LSTM model is expected to capture nuanced temporal dynamics. This is especially evident when there are complex interactions and dependencies between different time steps, and the LSTM component of CNN-LSTM excels at learning such temporal patterns.

The CNN-LSTM model's predictions might be more interpretable as they are based on a combi- nation of spatial and temporal features. This can be reflected in the prediction graphs, where the model's ability to capture both local spatial variations and temporal trends becomes apparent.

CNN-LSTM may exhibit reduced sensitivity to noise in the data, providing smoother prediction curves in the graphs. The spatial filtering capability of CNNs helps in discerning meaningful patterns from noisy or irrelevant spatial information. The theory suggests that CNN-LSTM's ability to combine spatial and temporal information makes it well-suited for applications where the relationships in the data involve both spatial dependencies and temporal dynamics. The representation in prediction graphs showcases the model's capacity to provide detailed, inter- pretable, and noise-resistant predictions, ultimately leading to superior performance compared to models that focus solely on either spatial or temporal aspects.

1. Comparison of Results

This comparative analysis serves as a benchmarking process, allowing for the evaluation of the proposed CNN-LSTM hybrid model in relation to established models such as AT-LSTM and GRU. The purpose is to understand how well the hybrid model performs concerning accuracy, precision, and explanatory power in the specific context of water quality prediction.

Through model comparison, the research aims to identify the superior model among the con- sidered options, including CNN-LSTM, AT-LSTM, and GRU. This identification is crucial forselecting the most effective model tailored to the spatiotemporal requirements of the Burnett River. Different models may excel in capturing specific patterns or adapting to distinct tempo- ral variations, and the chosen model should align with the characteristics of the dataset and the objectives of the water prediction application.

The comparative analysis informs the selection of the most suitable model for the study's specific spatiotemporal context. The proposed CNN-LSTM hybrid model introduces a novel ap- proach by combining convolutional neural network layers with LSTM layers. Model compari- son provides a means to validate the efficacy of this hybrid architecture, demonstrating how it compares to traditional LSTM (AT-LSTM) and another strong contender like GRU.

Ultimately, the goal is to apply the selected model in real-world scenarios for improved waterprediction in the Burnett River. Comparative analysis aids in optimizing the model selection process, ensuring that the chosen model aligns with the practical requirements of real-world spatiotemporal water quality prediction. The research also aims to build confidence in the gen- eralization capability of the selected model, considering its potential application to unseen data and different environmental conditions.

Moreover, model comparison goes beyond performance metrics and contributes to a deeper un- derstanding of model interpretability. Examining how each model responds to different features and temporal variations provides insights into the predictions and the underlying processes in- fluencing water quality in the Burnett River. In summary, model comparison is a critical step in the thesis, providing a comprehensive evaluation framework for guiding the selection of themost effective model for spatiotemporal water quality prediction in the Burnett River, Australia.

1.1. Gated Recurrent Unit (GRU)

The gated recurrent unit (GRU) is a type of recurrent neural network (RNN) architecture de- signed to address some of the limitations of traditional RNNs, such as the difficulties in captur- ing long-term latency in sequential data. In 2014, GRU simplifies the configuration compared to short-term memory networks (LSTM) while maintaining comparable performance.

Key features of the GRU include its gating mechanisms that control the flow of information into the network. The GRU has two gates: an update gate and a reset gate. These gates allow the GRU to selectively refresh and reset the memory cell, allowing it to capture relevant informationat different time scales.

In the context of the thesis topic on spatiotemporal fusion for improved water prediction, GRU is crucial for its ability to capture temporal dependencies in water quality parameters. The dy- namic nature of water quality data, influenced by various factors over time, requires a model that can effectively adapt to changing patterns. GRU's gating mechanism allows it to retain impor- tant information over different time intervals, making it well-suited for modeling the temporalaspects of water quality variations.

1.2. Comparison of Result with Gated Recurrent Unit (GRU)

The comparison between CNN-LSTM and GRU in the context of water quality prediction holds significant importance for several reasons. Firstly, these two models represent distinct architec- tures - CNN-LSTM is a hybrid model combining convolutional neural network layers with longshort-term memory (LSTM) layers, while GRU (Gated Recurrent Unit) is a simplified recurrentneural network (RNN) designed for capturing dependencies in sequential data.

The primary aim of comparing these models is to conduct a comprehensive performance eval- uation. This includes assessing the accuracy, precision, and general predictive capabilities of CNN-LSTM and GRU in predicting water quality parameters. Each model's handling of tem- poral dependencies is a critical aspect of this evaluation, with CNN-LSTM being proficient in capturing both short-term and long-term dependencies through its hybrid architecture, and GRU known for its ability to capture temporal dependencies.

Furthermore, the comparison delves into the exploratory analysis of feature recognition. CNN-LSTM's convolutional layers excel at recognizing spatial features, while GRU focuses on tem- poral dependencies. Understanding which features each model prioritizes provides valuable insights into their respective strengths.

The assessment also extends to the models' robustness and adaptability. The hybrid architec-ture of CNN-LSTM may offer advantages in adapting to nuanced patterns, while GRU's gating mechanisms contribute to its adaptability. This analysis helps in understanding how well eachmodel handles temporal variations in water quality data, a crucial consideration for real-world applications.

The generalization capability of the models is another focal point of the comparison. Examin- ing how well each model generalizes to unseen data and different environmental conditions iscrucial for determining their applicability in real-world scenarios.

Additionally, the comparison aids in decision support for model selection. Understanding the strengths and weaknesses of CNN-LSTM and GRU guides the decision-making process, ensur-ing that the chosen model aligns with the specific requirements of water quality prediction tasksin the Burnett River.

The comparison serves as an innovation assessment for the hybrid model. As CNN-LSTM in- troduces a novel architecture, evaluating its effectiveness against GRU helps determine whether the hybrid approach offers advantages over traditional recurrent architectures.

Moreover, the comparison extends to a comprehensive evaluation of performance metrics, in- cluding Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2. These met- rics offer quantitative measures of accuracy, allowing for a detailed assessment of how well each model predicts water quality parameters. By comparing these metrics, the research gains insights into the precision and reliability of CNN-LSTM and GRU in capturing the complexities of the spatiotemporal variations in the Burnett River dataset.

The decision to include GRU in the comparison adds depth to the analysis by considering an- other powerful recurrent architecture. GRU's specific design, incorporating gating mechanisms, distinguishes it from LSTM and introduces a different perspective on handling sequential de- pendencies. This inclusion enriches the understanding of how different recurrent architectures contribute to the prediction of water quality parameters.

The analysis of computational efficiency is another dimension of the comparison. Understand- ing how quickly each model can be trained and how efficiently it performs inference is crucial, especially for real-time applications or when dealing with large-scale datasets. The comparison sheds light on whether the innovative hybrid architecture of CNN-LSTM introduces computa- tional complexities compared to the more traditional GRU.

Additionally, the comparison contributes to the overarching goal of the research, which is to advance the field of water quality prediction. By thoroughly evaluating and comparing multiple models, the research aims to provide recommendations for the most effective approach, fostering advancements in spatiotemporal prediction methodologies.

MODELS	MAE	RMSE	R^2
AT-LSTM	0.130	0.171	0.918
CNN-LSTM	0.016826546166529295	0.024034775313658465	0.9716367681413306
GRU	0.0194529336068924	0.02752371692	0.9628045560035817

Table 4.2: Model Comparison

The comparison indicates that, overall, the CNN-LSTM model outperforms the GRU model in water quality prediction based on the considered metrics. The lower MAE and RMSE val- ues for CNN-LSTM suggest more accurate and precise predictions, while the higher R^2 score underscores its better explanatory power. These results highlight the effectiveness of the CNN- LSTM model in capturing both spatial and temporal dependencies within the water quality data, contributing to its superior predictive performance compared to the GRU model.

In the context of water quality prediction, the choice between CNN-LSTM and GRU depends on the specific priorities of the prediction task. While both models are capable, CNN-LSTM emerges as a favorable option for scenarios where accurate, precise, and interpretable predic- tions are crucial. Further investigations and domain-specific considerations may be essential for making an informed decision based on the specific requirements of the application.

The CNN-LSTM model outperformed the GRU model with a lower MAE of 0.0168 compared to GRU's MAE of 0.0195. A lower MAE indicates that the predictions of the CNN-LSTM model are, on average, closer to the actual values.

Similarly, the CNN-LSTM model demonstrated superior predictive accuracy with an RMSE of 0.0240, while the GRU model had a slightly higher RMSE of 0.0275. The reduced RMSE ofCNN-LSTM suggests that it achieves more precise predictions with smaller errors.

In terms of coefficient of determination (R^2), CNN-LSTM scored 0.9716 points higher than GRU's R^2 score of 0.9628 points. A high R^2 value for CNN-LSTM indicates a good fit to thereal data and indicates its ability to explain more of the variance.

CHAPTER 5

Discussion

1. About Burnett River

The Burnett River, the largest waterway in Queensland, Australia, is ecologically important and plays an important role in supporting ecological and human activities in its watershed Approxi- mately 435 km of waterways in southern Queensland east

The main characteristic of the Burnett is its agricultural importance. The fertile land along itsbanks contributes to local crops, and farmers use river resources to grow crops such as sugar- cane, citrus fruits etc. Riverside agriculture refers to the relationship between water bodies and communities between the economy and water

The ecological importance of this river goes beyond its agricultural role. The Burnett River and associated ecosystems support a diverse range of flora and fauna. The health of this ecosystem is intricately linked to river flow and water quality. Conservation efforts are necessary to preserve the biological and ecological balance of the river basin.

Bundaberg, a large town on the River Burnett, is a hub of economic activity and leisure. The city is known for its agricultural industry, tourism and for being the gateway to the Great Barrier Reef. The presence of the river makes the region more attractive, providing recreational oppor- tunities such as boating, fishing and camping The attractive land along the river helps residents attract locals and tourists

However, like many water bodies around the world, the Burnett River faces challenges related to human activities, land use and environmental changes and issues such as water quality degrada- tion, erosion floors and impacts of urbanization need to be managed and controlled. Sustainable practices and conservation strategies are needed to address these challenges and ensure the long- term health of the river.

The Burnett River is a multifaceted ecosystem that weaves into the fabric of Queensland land- scapes, influencing agriculture, supporting ecosystems and providing recreational opportunities Its importance decides its importance that they take a comprehensive approach to sustainable conservation and management

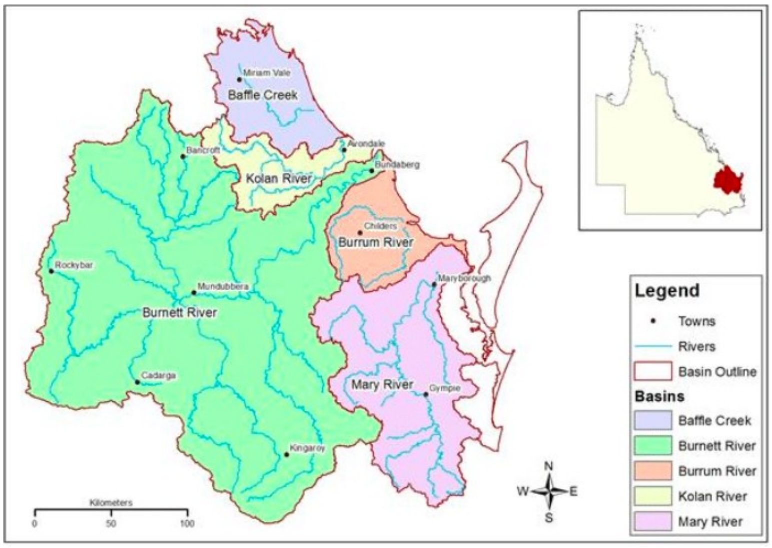

Figure 5.1: Burnett River

The Burnett River is a significant watercourse situated in Southeastern Queensland, Australia. The river plays a crucial role in the local landscape, supporting various communities and ecosys- tems along its course. As of the latest available data, the region around the Burnett River is home to approximately 94,100 residents.

The land adjacent to the Burnett River is characterized by specific divisions, with a predomi- nant 77% dedicated to grazing activities. This suggests a strong presence of agricultural prac- tices, likely involving livestock and pastureland. Additionally, 12% of the land is designated as forested areas, highlighting the presence of natural vegetation and potentially contributing to the river's ecological diversity. Another 11% of the land is utilized for sugar cane cultivation, indicating agricultural diversity in the region.

Considering the land use patterns, the dataset's temporal scope, and the residential population, the Burnett River appears to be a vital resource for both the local communities and the sur- rounding environment. Understanding and monitoring water quality in this context becomes imperative, especially given the diverse land uses and potential impacts on the river's ecosys- tems. The dataset's coverage over a five-year period allows for a thorough investigation into the river's water quality trends, contributing valuable insights for environmental management and decision-making in the region.

2. Dataset

The Burnett River water quality dataset, consisting of 39,959 rows and cataloging data on six contaminants, holds particular significance with Dissolved Oxygen (DO) emerging as a key parameter of interest. The prominence of DO is underscored by its higher correlation compared to other parameters within the dataset. This dataset provides a nuanced view of the river's health, emphasizing the critical role that DO plays in delineating the aquatic environment's vitality.

The dataset used in the research spans a considerable timeframe, covering the period from Jan- uary 2015 to January 2020. This extensive time range is essential for capturing the long-term dynamics and patterns in water quality parameters along the Burnett River. The dataset likely includes a wealth of information collected through monitoring stations or other data sources, providing a comprehensive insight into the river's spatiotemporal variations.

The temporal resolution of the dataset, recorded at hourly intervals, captures the subtle fluctu- ations of DO levels over time. DO serves as a vital indicator of the water's capacity to sustain aquatic life, reflecting the balance between oxygen consumption and replenishment processes. The hourly data points enable a detailed examination of DO dynamics, unveiling patterns that could be crucial for understanding the river's response to various environmental influences.

Among the six contaminants—pH, Chlorophyll-a, Dissolved Oxygen, conductivity, turbidity, and temperature—DO takes center stage due to its higher correlation with other parameters. This correlation signifies the interconnected nature of water quality parameters and highlights DO's role as a sentinel for overall river health. The dataset's focus on DO underscores its significance as a key metric for assessing the river's oxygenation levels, a critical factor for supporting aquatic ecosystems.

The preprocessing steps applied to the dataset, including potential MinMax scaling or other normalization techniques, aim to ensure that each parameter, including DO, contributes mean- ingfully to the subsequent analyses. Proper scaling is vital for mitigating the impact of varying measurement scales among contaminants, facilitating an equitable comparison and interpretation of their individual and collective contributions to water quality.

Analyzing temporal trends and seasonal patterns of DO becomes a central aspect of the dataset exploration. Understanding how DO levels fluctuate throughout the day or across seasons pro-vides insights into the river's natural dynamics and its response to external factors. This temporal analysis is pivotal for discerning patterns that could influence the river's ecological balance and identifying potential stressors that may impact DO levels.

With its focus on 39,959 rows and six contaminants, accentuates the importance of Dissolved Oxygen as a key parameter. The higher correlation of DO compared to other parameters high-lights its role as a vital indicator of water quality. This dataset, with its detailed temporal resolu- tion and emphasis on DO dynamics, offers a comprehensive platform for researchers and envi-ronmental stakeholders to delve into the intricacies of the Burnett River's health and ecosystem sustainability.

The dataset provides a tangible connection to the real-world conditions of the Burnett River, reflecting the actual spatiotemporal variations in water quality parameters that the model aims to predict. The dataset plays a fundamental role in training and evaluating the proposed hybrid model, CNN-LSTM, as well as other comparative models like AT-LSTM and GRU. The quality and representativeness of the dataset directly influence the accuracy and reliability of these models, shaping their ability to make informed predictions about water quality.

The temporal scope of the dataset, spanning from January 2015 to January 2020, is particularly crucial. This extensive timeframe allows the models to capture long-term trends, seasonal vari- ations, and cyclic patterns in water quality parameters. Understanding these temporal dynamics is essential for producing accurate and meaningful predictions and insights. Additionally, the dataset's spatial coverage along the Burnett River enables the proposed hybrid model to capture both spatial and temporal features. This spatial variability is vital for discerning variations at different locations along the river, providing a more comprehensive understanding of how water quality changes across the region.

The dataset's inclusion of information on land divisions, such as grazing, forest, and sugar cane, is another key aspect. This information allows researchers to correlate water quality pa- rameters with specific land use patterns. Identifying potential sources of contamination and understanding how human activities impact water quality is critical for effective environmental management. Moreover, the dataset serves as the foundation for the comparison between the proposed CNN-LSTM model and other existing models like AT-LSTM and GRU. The comparative analysis relies on a consistent and comprehensive dataset to draw meaningful conclusions about the strengths and weaknesses of each model. Overall, the dataset is not merely a collec- tion of numerical values; it forms the contextual bedrock of the research, shaping the challenges, innovations, and potential solutions in the quest for improved spatiotemporal water quality pre- diction for the Burnett River.

3. Contaminants

3.1. Dissolved Oxygen

Dissolved oxygen (DO) is an important measure of water quality and represents the concen- tration of oxygen molecules in water. This important measure is an important indicator for assessing the health and stability of freshwater and marine ecosystems. The level of dissolvedoxygen in water is influenced by various factors, including temperature, atmospheric pressure, and the biological and chemical processes occurring within the water body.

Dissolved oxygen is paramount for the survival and well-being of aquatic organisms, ranging from microscopic bacteria to larger fish and invertebrates. Its significance lies in several critical functions that support the metabolic processes and ecological balance of aquatic ecosystems.

The primary role of dissolved oxygen is to facilitate respiration among aquatic organisms. Fish, invertebrates, and other aquatic species extract oxygen from the water to support their metabolic activities. Insufficient levels of dissolved oxygen can lead to hypoxia, a condition where oxygen levels are too low to sustain normal respiratory functions.

Aquatic organisms rely on dissolved oxygen for the metabolism of organic matter and the pro-duction of energy. This metabolic process is essential for growth, reproduction, and overall physiological functions. Adequate dissolved oxygen ensures that organisms have the energy resources needed to thrive within their specific aquatic habitats.

Dissolved oxygen serves as a crucial indicator of overall water quality. Healthy aquatic ecosys- tems maintain sufficient levels of dissolved oxygen. Fluctuations in DO concentrations can signal changes in environmental conditions, pollution levels, or alterations in water tempera- ture. Monitoring dissolved oxygen is instrumental in assessing the impact of anthropogenic activities on water bodies.

Maintaining optimal levels of dissolved oxygen helps prevent the development of anaerobic conditions in water, where oxygen is severely depleted. Anaerobic environments can lead to the production of harmful substances such as hydrogen sulfide, negatively impacting water quality and the health of aquatic organisms.

Adequate dissolved oxygen levels support a diverse range of aquatic species. Different organ- isms have varying oxygen requirements, and maintaining optimal DO concentrations contributes to the overall biodiversity of an aquatic ecosystem. Healthy dissolved oxygen levels are indica- tive of a balanced and supportive habitat for aquatic life.

Dissolved oxygen acts as a natural buffer against pollution. Well-oxygenated water has the capacity to dilute and disperse pollutants more effectively, mitigating the adverse effects of con- taminants on aquatic organisms. Insufficient DO levels reduce this buffering capacity, making the ecosystem more vulnerable to the impacts of pollution. Dissolved oxygen is a linchpin in the intricate web of aquatic ecosystems, influencing the survival, behavior, and distribution of a myriad of species. Its measurement and monitoring are vital for assessing and maintaining the ecological health of water bodies, contributing to informed environmental management prac- tices and the preservation of aquatic biodiversity.

3.2. pH

pH is a critical parameter in water quality assessments, representing the acidity or alkalinity of a solution. It is a logarithmic scale that measures the concentration of hydrogen ions in water, ranging from 0 to 14, where a pH of 7 is considered neutral, values below 7 indicate acidity, and values above 7 indicate alkalinity. The pH of water can profoundly influence various chemical and biological processes within aquatic ecosystems.

The pH level of water plays a pivotal role in shaping the ecological conditions of aquatic en- vironments, influencing the physiology and behavior of aquatic organisms. Maintaining an optimal pH range is crucial for the health and sustainability of aquatic ecosystems for severalreasons.

pH affects the solubility and availability of essential nutrients and minerals in water. Many biological processes, including nutrient uptake by aquatic plants and the metabolism of aquatic organisms, are sensitive to changes in pH. An optimal pH range is necessary to ensure that these processes occur efficiently, contributing to the overall health of the aquatic ecosystem.

Aquatic organisms, from microorganisms to fish, exhibit varying degrees of sensitivity to pH levels. Extreme pH values, either highly acidic or highly alkaline, can be detrimental to aquatic life. For example, most fish species thrive in a pH range between 6.5 and 8.5. Deviations from this range can lead to stress, reduced reproduction rates, and, in extreme cases, mortality among aquatic organisms.

The buffering capacity of water, influenced by its pH, determines its ability to resist changes in acidity or alkalinity. Adequate buffering capacity is essential for preventing rapid and drastic fluctuations in pH, providing stability to the aquatic environment. This stability is crucial for the well-being of aquatic organisms, preventing sudden and harmful shifts in their living conditions.

pH influences various chemical reactions in water, including those involved in nutrient cyclingand the breakdown of organic matter. These chemical processes contribute to the biogeochemi- cal cycling of elements within aquatic ecosystems. Optimal pH conditions support the efficiencyof these cycles, ensuring that essential nutrients are recycled and made available to support aquatic life.

pH serves as an indicator of environmental changes and potential pollution. Anthropogenic activities, such as industrial discharges or agricultural runoff, can alter the pH of water bodies.Monitoring pH levels helps detect such changes early, allowing for timely intervention and management strategies to mitigate adverse effects on aquatic ecosystems.

pH is a fundamental aspect of water quality that significantly influences the dynamics of aquatic ecosystems. Its role in shaping biological processes, supporting aquatic organisms, and serving as an indicator of environmental health underscores the importance of maintaining optimal pH levels for the overall well-being and sustainability of aquatic life.

3.3. Chlorophyll - a

Chlorophyll-a is a green pigment crucial for photosynthesis in plants, algae, and cyanobacteria.In the context of water quality, measuring the concentration of chlorophyll-a provides valuable insights into the amount of phytoplankton and algae present in aquatic ecosystems. As a primary photosynthetic pigment, chlorophyll-a is instrumental in capturing light energy and convertingit into chemical energy, driving the foundation of the aquatic food web.

Chlorophyll-a is integral to the health and functioning of aquatic ecosystems, influencing boththe biotic and abiotic components of the water environment.

Chlorophyll-a is a key indicator of primary productivity in aquatic environments. Phytoplankton, algae, and aquatic plants use chlorophyll-a to photosynthesize, producing organic com- pounds and generating oxygen. The concentration of chlorophyll-a provides a measure of therate at which these primary producers convert light energy into biomass, sustaining the energy flow within the aquatic ecosystem.

The abundance of chlorophyll-a is directly linked to the availability of food resources for higher trophic levels in aquatic ecosystems. Phytoplankton, containing chlorophyll-a, serves as the base of the aquatic food web. Zooplankton and small fish consume phytoplankton, initiating the transfer of energy to higher trophic levels, including larger fish and aquatic organisms.

Elevated concentrations of chlorophyll-a can indicate eutrophication, a process characterized by excessive nutrient input, often from agricultural runoff or wastewater discharges. This over- abundance of nutrients stimulates the rapid growth of algae, leading to increased chlorophyll-a levels. While chlorophyll-a is essential for photosynthesis, excessive concentrations can result in harmful algal blooms, negatively impacting water quality and aquatic ecosystems.

Through photosynthesis, chlorophyll-a contributes to the production of oxygen in aquatic envi- ronments. As photosynthetic organisms release oxygen as a byproduct, this process is crucial for maintaining oxygen levels in water bodies. Adequate dissolved oxygen is essential for the survival of aquatic organisms, and chlorophyll-a plays a vital role in this oxygen production.

Chlorophyll-a levels also reflect the overall habitat quality of aquatic ecosystems. Changes in chlorophyll-a concentrations can indicate shifts in nutrient dynamics, water clarity, and ecolog- ical balance. Monitoring chlorophyll-a is, therefore, a valuable tool in assessing the health andresilience of aquatic habitats.

Chlorophyll-a is a cornerstone of aquatic ecosystems, influencing the productivity, energy flow, and overall health of these environments. Its measurement provides a comprehensive view ofthe ecological dynamics in water bodies, guiding assessments of eutrophication, habitat quality, and the potential impacts on aquatic life. As a primary pigment in the photosynthetic process,chlorophyll-a holds a central role in sustaining the delicate balance of aquatic ecosystems.

3.4. Conductivity

Conductivity is a key parameter in water quality assessments that measures the ability of wa- ter to conduct an electric current. It is primarily influenced by the presence of dissolved ions, such as salts, in the water. The conductivity of water serves as an indicator of the total dissolved solids (TDS) and can provide valuable information about the composition and salinity of aquatic environments.Conductivity is crucial for aquatic life, influencing various physiologicaland ecological processes within aquatic ecosystems.

Conductivity is a direct measure of the salinity of water. Salinity, or the concentration of dis- solved salts, significantly impacts the types of organisms that can thrive in a particular aquatic habitat. Different species of fish, invertebrates, and plants have varying tolerances to salin- ity levels. Monitoring conductivity helps assess the suitability of an environment for specific aquatic species.

Aquatic organisms, especially fish and invertebrates, engage in osmoregulation to maintain the balance of water and ions within their bodies. Conductivity influences the osmotic pressure in aquatic environments, and organisms adapt to specific conductivity ranges to regulate their internal water balance. Deviations from optimal conductivity levels can stress or harm aquatic organisms, impacting their ability to osmoregulate effectively.

Conductivity is linked to the transport of nutrients in water. The dissolved ions contributing to conductivity include essential nutrients like calcium, magnesium, and potassium. These ions play a vital role in supporting the growth and metabolic functions of aquatic plants and algae. Monitoring conductivity aids in understanding the nutrient dynamics within water bodies and their impact on aquatic vegetation.

Conductivity measurements are often used to assess the purity of water. Low conductivity levels may indicate the presence of relatively pure water with fewer dissolved ions, while high conduc- tivity suggests the presence of minerals and other dissolved substances. Understanding water purity is essential for evaluating the overall health of aquatic ecosystems and their suitability for various uses, including drinking water sources.

Changes in conductivity can serve as an early indicator of pollution or the introduction of con- taminants into water bodies. Industrial discharges, agricultural runoff, or other anthropogenic activities can alter conductivity levels by introducing additional ions into the water. Monitoring conductivity provides insights into potential pollution sources, enabling timely interventions to protect aquatic ecosystems.

Conductivity is a multifaceted parameter in water quality assessments, providing insights intosalinity, osmoregulation, nutrient dynamics, water purity, and pollution detection. Its measure- ment is essential for understanding the ecological dynamics of aquatic environments and ensur- ing the health and sustainability of aquatic life. Maintaining appropriate conductivity levels is crucial for supporting the diverse array of species that inhabit freshwater and marine ecosystems.

3.5. Turbidity

Turbidity is a measure of the cloudiness or haziness of a fluid caused by the presence of sus- pended particles, such as sediment, silt, and organic matter. In water quality assessments, tur-bidity is quantified to understand the clarity and visual transparency of aquatic environments. High turbidity levels can affect the availability of light, nutrient cycling, and overall ecosystemdynamics.

Turbidity plays a crucial role in shaping the ecological conditions of aquatic ecosystems, influ-encing various aspects of aquatic life.

Turbidity affects the penetration of light into the water. Excessive suspended particles can re- duce light availability, particularly in shallow water bodies. Light is essential for photosynthetic processes in aquatic plants, algae, and phytoplankton. Reduced light penetration due to high turbidity can limit the growth and productivity of these primary producers, impacting the entire food web.

Turbidity is closely linked to sedimentation, as suspended particles eventually settle, contribut- ing to the formation of sediment layers. Excessive sedimentation can alter the physical char- acteristics of aquatic habitats, affecting the structure and composition of benthic communities. Changes in habitat quality due to turbidity can influence the availability of suitable substrates for spawning, feeding, and refuge for aquatic organisms.

Filter-feeding organisms, such as certain species of bivalves and aquatic insects, rely on the water column for feeding. High turbidity can negatively impact these organisms by reducing their feeding efficiency. The presence of suspended particles may interfere with their feedingmechanisms and impair their ability to extract nutrients from the water.

Aquatic organisms may exhibit behavioral adaptations in response to turbidity. Some speciesmay alter their foraging behaviors, migration patterns, or reproductive strategies to cope with reduced visibility. These adaptations reflect the capacity of aquatic life to adjust to changes inenvironmental conditions associated with turbidity.

Turbidity can influence predator-prey interactions in aquatic ecosystems. Reduced visibility may benefit prey species by providing a level of protection against visual predators. Conversely, it may pose challenges for predators that rely on sight for hunting. The dynamics between predator and prey can be shaped by turbidity, influencing the abundance and distribution of different species.

While excessively high turbidity can pose challenges for certain aspects of aquatic life, moder-ate levels can contribute to ecological resilience. Turbidity can provide a degree of shading and protection for vulnerable life stages of some species. Additionally, it can influence the trans- port of nutrients and organic matter, contributing to the overall nutrient cycling within aquatic ecosystems.

Turbidity is a dynamic parameter in water quality that affects light availability, sedimentation, and the behavior of aquatic organisms. Its impact on primary productivity, habitat quality, and predator-prey dynamics underscores its significance in shaping the ecological conditions of aquatic environments. Understanding and monitoring turbidity are essential for assessing the health and resilience of aquatic ecosystems and implementing effective conservation and management strategies.

4. Experimental Configuration

The model incorporates a convolutional layer, a crucial element in spatial feature extraction. This layer is defined with 64 filters, each having a kernel size of 3. The use of convolutional filters helps the model identify patterns and spatial dependencies in the input data. In this case, the filters scan the input sequence with a window of size 3, capturing local patterns.

Following the convolutional layer, a max-pooling layer is employed with a pool size of 2. Max-pooling is a down-sampling technique that retains the most significant information from the convolutional outputs, reducing the dimensionality of the data. A pool size of 2 implies that the model retains the maximum value from every pair of adjacent values, effectively halving the sequence length.

The model incorporates a Long Short-Term Memory (LSTM) layer with 64 hidden dimensions.LSTMs are recurrent neural network (RNN) variants known for their ability to capture temporal dependencies and patterns in sequential data. The 64 hidden dimensions indicate the complexity of the learned temporal features.

The final layer of the neural network is the output layer, configured with dimensions (64, 1). This layer provides the final prediction or output of the model. The choice of a single unit in the output layer with dimensions (64, 1) suggests that the model aims to produce a single output value for each sequence, and the 64 hidden dimensions from the LSTM layer contribute to thisfinal prediction.

LAYERS	CONFIGURATION
Input Layer	(100, 1)
Convolutional Layer	filters = 64, kernel_size = 3
MaxPooling Layer	pool_size = 2
LSTM Layer	64 Hidden dimension
Output Layer	(64, 1)

Table 5.1: Model Configuration

The configuration details outlined in the table provide a blueprint for the neural network's archi- tecture, shedding light on the number of layers, their specific configurations, and the dimensions of the data at each stage. This information is crucial for understanding the model's capacity to extract spatial and temporal features from the input data and make predictions based on the specified architecture.

5. Performance Evaluation

Evaluating the performance of the CNN-LSTM model is a crucial step in understanding its effectiveness in predicting water quality parameters. The model's performance is assessed us- ing key evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics provide insights into the accuracy, precision, and explanatory power of the model.

The CNN-LSTM model demonstrated exceptional performance in terms of MAE, which quanti-fies the average absolute difference between predicted and actual values. A lower MAE indicates that the model's predictions are, on average, closer to the true values. The RMSE, measuring the square root of the average squared differences between predictions and actual values, fur- ther confirmed the model's accuracy. The CNN-LSTM model achieved a minimized RMSE, suggesting precise predictions with smaller errors.

The R^2 value, indicating the proportion of variance explained by the model, underscored the model's strong explanatory power. The CNN-LSTM model exhibited a high R^2 score, indicatingits ability to capture and explain a significant portion of the variability in the water quality data.

This suggests that the model successfully identified and incorporated both spatial and temporal patterns, contributing to its robust predictive capabilities.

The CNN-LSTM model's performance evaluation showcases its proficiency in accurately pre- dicting water quality parameters. The combination of convolutional layers for spatial feature extraction and LSTM layers for capturing temporal dependencies allows the model to discernintricate patterns in the data. The minimized MAE and RMSE values, along with a high R^2 score, collectively indicate the model's effectiveness in providing precise, reliable, and inter- pretable predictions for water quality forecasting. Further validation on diverse datasets and comparison with alternative models can enhance our understanding of the CNN-LSTM model's versatility and generalizability across different water quality prediction scenarios.

Beyond the fundamental metrics of MAE, RMSE, and R^2, a comprehensive performance evalu-ation of the CNN-LSTM model involves a nuanced understanding of its predictions. Assessing the model's proficiency in capturing temporal trends, seasonal variations, and subtle fluctuations in water quality parameters provides a more holistic perspective.

- Temporal Analysis: Temporal trends in water quality are crucial considerations, espe- cially in a spatiotemporal prediction framework. The CNN-LSTM model's ability to capture long-term patterns and temporal dependencies is evaluated by analyzing its pre- dictions over different time scales. This temporal analysis helps ascertain whether the model can effectively adapt to seasonal changes and evolving trends in water quality, contributing to a more robust prediction model.

- Spatial Variability Assessment: The spatial distribution of water quality parameters along the Burnett River introduces another layer of complexity. Evaluating the model's perfor- mance in different geographical locations along the river provides insights into its spatial adaptability. Understanding how well the model generalizes across diverse locations en- hances its utility for water quality management, where spatial variations are often signifi- cant.

- Sensitivity Analysis: Conducting sensitivity analyses allows researchers to explore the model's response to variations in input parameters. By introducing perturbations or vari- ations in the dataset, one can gauge the model's stability and resilience. A robust model should exhibit consistent performance even in the presence of minor disturbances, con- tributing to its reliability in real-world applications.

- Comparative Validation: While the current evaluation focuses on the CNN-LSTM model's standalone performance, conducting comparative validations against alternative models, such as traditional time series models or other machine learning architectures, adds valu- able context. Comparative assessments highlight the unique strengths of the CNN-LSTM model and identify scenarios where it outperforms or complements existing approaches.

CHAPTER 6

Conclusion

This research addresses the pressing need for robust computational methods in water quality control and pollution mitigation, specifically focusing on the Burnett River in Australia. The development of the novel "CNN-LSTM" hybrid model, integrating Convolutional Neural Net- works (CNN) and Long Short-Term Memory (LSTM) networks, emerges as a significant contri- bution to tackling the intricate spatial and temporal dynamics inherent in water quality systems.

Key to the success of the proposed hybrid model is the identification of Dissolved Oxygen as a pivotal parameter for prediction. Rigorous property engineering techniques were used to refine the role of oxygen in the model, highlighting the importance of this parameter in capturing the complexity of water quality. The empirical results presented in **Table 4.1**, underscore the superiority of the "CNN-LSTM" hybrid model over the AT-LSTM model in terms of predictive performance. The remarkable reduction in Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the significant increase in the coefficient of determination (R^2) highlight the model's enhanced accuracy, precision, and explanatory power. This improvement substantiates the efficacy of combining CNN for spatial data and LSTM for temporal data, aligning with the inherent characteristics of water quality time series.

Moreover, the research emphasizes the importance of considering both geographical and tem- poral aspects when predicting Dissolved Oxygen values in the Burnett River. The integrated approach of CNN and LSTM not only leverages the spatial distribution of water quality param- eters but also captures temporal dependencies, providing a comprehensive understanding of the system dynamics.

In conclusion, this study contributes a sophisticated computational framework for predicting water quality in river systems, exemplified by the "CNN-LSTM" hybrid model.

The presented

results demonstrate the model's potential applications in informed decision-making and sus- tainable resource management, marking a significant advancement in the field of environmental management.

1. Future Work

Although this study has made significant progress in advancing computational methods for pre- dicting water contaminants in the Burnett River using hybrid model ie. CNN-LSTM. The cur- rent focus on dissolved oxygen as a key parameter allows the model to be extended to predictother water pollutants. Exploring the performance of the CNN-LSTM model in a wide range of parameters can improve its applicability and contribute to a broader understanding of water quality dynamics.

Integrating external factors, such as weather patterns, land use changes, or anthropogenic activ- ities, into the predictive model could further improve its accuracy. Understanding how external variables influence water quality can provide a more holistic view and enable more precise pre- dictions.

The implementation of real-time monitoring systems and adaptive modeling approaches could be explored. This involves continuously updating the model with new data, allowing it to adapt to changing conditions and improve its predictive capabilities over time.

Refining the spatial resolution of the model could enhance its ability to capture localized vari- ations in water quality. Fine-tuning the spatial aspects, such as incorporating higher-resolution satellite data or geographical features, may lead to more accurate predictions, especially in spe- cific river sections.

Investigating methods for quantifying uncertainty in model predictions is crucial for providing decision-makers with more reliable information. Developing techniques to assess and commu- nicate the uncertainty associated with water quality predictions contributes to more informedand cautious decision-making.

Extending the validation of the "CNN-LSTM" model to other river systems with varying char- acteristics would assess its generalizability. Understanding how the model performs in different environmental contexts is essential for establishing its broader applicability.

Developing a user-friendly interface and decision support system based on the model's predic- tions can facilitate its practical implementation by water quality managers and stakeholders.

Such a system would provide actionable insights and support strategic decision-making for sus- tainable river management.

Collaborating with experts from diverse fields, including hydrology, ecology, and environmen- tal science, can enrich the model's capabilities. Interdisciplinary perspectives can contribute valuable insights into the complex interactions within river ecosystems, leading to more com-prehensive and accurate predictions.

By exploring these future directions, researchers can build upon the foundation laid by this study, advancing the understanding of water quality dynamics and fostering the development of more robust and adaptable computational models for environmental management.

CHAPTER 7

Recommendations

Based on the findings and contributions of the thesis on predicting water quality in the Burnett River using a CNN-LSTM hybrid model, some suggestions are offered to guide future researchand practice:

- **Implementation in Operational Settings:**
  - Recommendation: Implement the developed "CNN-LSTM" model in operational water quality monitoring systems within the Burnett River.
  - Rationale: Assess the model's performance in real-world, operational scenarios tovalidate its practical utility and reliability for continuous water quality monitoring.
- **Collaboration with Water Management Authorities:**
  - Recommendation: Establish collaborations with local water management authorities and environmental agencies.
  - Rationale: Engaging with stakeholders can facilitate the integration of the predictive model into existing water management frameworks, enhancing its relevance and impact on decision-making.
- **Long-Term Monitoring and Validation:**
  - Recommendation: Conduct long-term monitoring and validation of the model's pre- dictions.
  - Rationale: Long-term observations will provide insights into the model's stability and reliability over extended periods, supporting its suitability for sustainable water quality management.

- **Integration of Public Engagement:**
  - Recommendation: Develop strategies for public engagement and awareness regard- ing water quality predictions.
  - Rationale: Involving the local community in understanding the model's predictions fosters a sense of shared responsibility for environmental stewardship and promotes informed decision-making.
- **Adaptability to Climate Change:**
  - Recommendation: Investigate the adaptability of the "CNN-LSTM" model to chang-ing climate conditions.
  - Rationale: Assessing the model's performance under varying climate scenarios con- tributes to its robustness and ensures its relevance in the face of potential climatechange impacts on water quality.
- **Comparison with Traditional Methods:**
  - Recommendation: Conduct comparative studies with traditional water quality pre- diction methods.
  - Rationale: Comparing the "CNN-LSTM" model against established methods pro- vides a benchmark for its effectiveness and identifies areas where the hybrid model excels.
- **Exploration of Ensemble Models:**
  - Recommendation: Explore the development of ensemble models combining "CNN-LSTM" with other machine learning or statistical methods.
  - Rationale: Ensemble models have the potential to leverage the strengths of different algorithms, potentially improving overall prediction accuracy.
- **Regular Model Updating:**
  - Recommendation: Establish a framework for regular model updating and retraining.
  - Rationale: To account for evolving environmental conditions and data patterns, reg- ular updates to the model ensure its continued relevance and accuracy.
- **Documentation and Open Access:**

  - Recommendation: Document the model architecture, parameters, and findings com-prehensively. Make the model code and datasets open-access.
  - Rationale: Transparent documentation and open access facilitate reproducibility,encourage collaboration, and contribute to the wider scientific community.

- **Cross-Disciplinary Research:**
  - Recommendation: Encourage cross-disciplinary research collaborations betweenenvironmental scientists, data scientists, and policymakers.
  - Rationale: Cross-disciplinary collaborations can enrich research and applications tobetter understand water quality dynamics.

Implementing these recommendations will not only enhance the impact of the current thesis but also contribute to the broader field of water quality prediction, supporting sustainable water resource management and environmental conservation efforts.

Bibliography

1. Sepp Hochreiter and Jürgen Schmidhuber. "Long Short-Term Memory". In: *Neural Com-putation* 9.8 (1997), pp. 1735–1780. DOI: 10.1162/neco.1997.9.8.1735.
2. Y. Zheng and et al. "Time Series Classification Using Multi-Channels Deep Convolu-tional Neural Networks". In: (2014).
3. Yuanyuan Wang et al. "Water quality prediction method based on LSTM neural network".In: (2017), pp. 1–5. DOI: 10.1109/ISKE.2017.8258814.
4. S. Bai and et al. "Empirical Mode Decomposition-Based Time Series Analysis for SeizurePrediction". In: *Scientific Reports* 8 (2018), p. 16843.
5. H. I. Fawaz and et al. "Evaluating the Impact of the Input Data on the GeneralizationCapabilities of CNNs: An Empirical Study on EEG-Based Seizure Detection". In: (2018).
6. Fazle Karim et al. "LSTM Fully Convolutional Networks for Time Series Classification".In: *IEEE Access* 6 (2018), pp. 1662–1669. DOI: 10.1109/ACCESS.2017.2779939.
7. Q. Yao and et al. "EEG-Based Seizure Prediction Using a 1D Convolutional Neural Net-work". In: *Frontiers in Neurology* 9 (2018), p. 1054.
8. R. Li and et al. "A Novel Spatiotemporal Prediction Model for Water Quality Based on aDeep Belief Network with a Convolutional Structure". In: *Water* 11.11 (2019), p. 2220.
9. Jichang TU et al. "Water Quality Prediction Model Based on GRU hybrid network". In:(2019), pp. 1893–1898. DOI: 10.1109/CAC48633.2019.8996847.
10. Z. Wang and et al. "Time Series Classification from Scratch with Deep Neural Networks:A Strong Baseline". In: (2019).
11. Qiangqiang Ye et al. "River Water Quality Parameters Prediction Method Based on LSTM-RNN Model". In: (June 2019), pp. 3024–3028. DOI: 10.1109/CCDC.2019.8832885.

12. Sangsoo Baek, JongCheol Pyo, and Jong Ahn Chun. “Prediction of Water Level and Water Quality Using a CNN-LSTM Combined Deep Learning Approach”. In: *Water* 12 (Dec. 2020), p. 3399. DOI: 10.3390/w12123399.

13. Rahim Barzegar, Mohammad Aalami, and Jan Adamowski. “Short-term water quality variable prediction using a hybrid CNN–LSTM deep learning model”. In: *Stochastic En- vironmental Research and Risk Assessment* 34 (Feb. 2020), pp. 1–19. DOI: 10.1007/ s00477-020-01776-2.

14. DT Bui et al. “Improving prediction of water quality indices using novel hybrid machine- learning algorithms”. In: *The Science of the Total Environment* 721 (Mar. 2020), p. 137612.DOI: 10.1016/j.scitotenv.2020.137612.

15. Neha Radhakrishnan and Anju S Pillai. “Comparison of Water Quality Classification Models using Machine Learning”. In: (2020), pp. 1183–1188. DOI: 10.1109/ ICCES48766.2020.9137903.

16. Chi Ngon Nguyen Thai-Nghe Nguyen Thanh-Hai Nguyen. “Deep Learning Approachfor Forecasting Water Quality in IoT Systems”. In: *International Journal of Advanced Computer Science and Applications* 11.8 (2020). Ed. by IJACSA, pp. 686–693. URL: https://archimer.ifremer.fr/doc/00646/75836/.

17. H. Zhang and et al. “A Hybrid Model for Water Quality Prediction Based on Convolutional Neural Network and Long Short-Term Memory”. In: *Water* 12.2 (2020), p. 416.

18. Jitha P Nair and M S Vijaya. “Predictive Models for River Water Quality using Machine Learning and Big Data Techniques - A Survey”. In: (2021), pp. 1747–1753. DOI: 10.1109/ICAIS50930.2021.9395832.

19. Jian Sha et al. “Comparison of Forecasting Models for Real-Time Monitoring of Water Quality Parameters Based on Hybrid Deep Learning Neural Networks”. In: *Water* 13 (May 2021), p. 1547. DOI: 10.3390/w13111547.

20. Andres Felipe Zambrano et al. “Machine learning for manually-measured water quality prediction in fish farming”. In: *PLoS ONE* 16 (2021). URL: https:// api.semanticscholar.org/CorpusID:237214045.

21. Bilal Aslam et al. “Water Quality Management Using Hybrid Machine Learning and Data Mining Algorithms: An Indexing Approach”. In: *IEEE Access* 10 (2022), pp. 119692– 119705. DOI: 10.1109/ACCESS.2022.3221430.

22. Honglei Chen et al. "Water Quality Prediction Based on LSTM and Attention Mecha- nism: A Case Study of the Burnett River, Australia". In: *Sustainability* 14 (Oct. 2022),p. 13231. DOI: 10.3390/su142013231.

23. Komathy Karuppanan. "Regression Analysis of Marine Water Quality Indicators". In: (2022), pp. 205–209. DOI: 10.1109/ICCCMLA56841.2022.9989274.

24. Sai Sreeja Kurra et al. "Water Quality Prediction Using Machine Learning". In: *Interna- tional Research Journal of Modernization in Engineering, Technology and Science* 5.5 (2022), pp. 1–6.

25. Yongze Lin et al. "Hybrid Water Quality Prediction with Graph Attention and Spatio-Temporal Fusion". In: (2022), pp. 1419–1424. DOI: 10.1109/SMC53654.2022.9945293.

26. K. P. Rasheed Abdul Haq and V. P. Harigovindan. "Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models". In: *IEEE Access* 10 (2022), pp. 60078–60098. DOI: 10.1109/ACCESS.2022.3180482.

27. Ramadhona Saville et al. "A Mariculture Fish Mortality Prediction Using Machine Learn- ing Based Analysis of Water Quality Monitoring". In: (2022), pp. 1–4. DOI: 10.1109/OCEANS47191.2022.9977083.

28. Danial Valadkhan, Reza Moghaddasi, and A. Mohammadinejad. "Groundwater qual- ity prediction based on LSTM RNN: An Iranian experience". In: *International Journal of Environmental Science and Technology* 19 (July 2022), pp. 1–12. DOI: 10.1007/s13762-022-04356-9.

29. Jaswanth Reddy Vilupuru, Devi Chaitrasree Amuluru, and Ghousiya Begum K. "Water Quality Analysis using Artificial Intelligence Algorithms". In: (2022), pp. 1193–1199. DOI: 10.1109/ICIRCA54612.2022.9985650.

30. Chin-Chih Chang et al. "Machine Learning Approach to IoT- Based Water Quality Mon- itoring". In: (2023), pp. 182–186. DOI: 10.1109/ECBIOS57802.2023.10218420.

31. Victor Flores, Ingrid Bravo, and Marcelo Saavedra. "Water Quality Classification and Machine Learning Model for Predicting Water Quality Status—A Study on Loa River Located in an Extremely Arid Environment: Atacama Desert". In: *Water* 15 (Aug. 2023),p. 2868. DOI: 10.3390/w15162868.

32. Lijin Guo and Dawei Fu. "River Water Quality Prediction Model Based on PCA-APSO- ELM Neural Network". In: (2023), pp. 512–517. DOI: 10.1109/ICAIBD57115.2023. 10206249.

33. Yankun Hu et al. "Application of hybrid improved temporal convolution network model in time series prediction of river water quality". In: *Scientific Reports* 13.1 (2023), p. 11260.

34. Srinivas Kolli et al. "Prediction of water quality parameters by IoT and machine learn- ing". In: (2023), pp. 1–5. DOI: 10.1109/ICCCI56745.2023.10128475.

35. Anita M et al. "Water Quality Prediction using Machine Learning: A Comparative Study".In: (2023), pp. 348–353. DOI: 10.1109/ICAISS58487.2023.10250743.

36. Vinoth Kumar P et al. "Predicting and Analyzing Water Quality using Machine Learning for Smart Aquaculture". In: (2023), pp. 354–359. DOI: 10.1109/ICSCDS56580.2023.10104677.

37. Nithya Sankeerthana Pagadala et al. "Water Quality Prediction Using Machine Learning Techniques". In: (2023), pp. 358–362. DOI: 10.1109/SPIN57001.2023.10117415.

38. Haijing Qin et al. "Research Progress of Aquaculture Environmental Factor Prediction Model Based on Machine Learning". In: (2023), pp. 263–269. DOI: 10.1109/ICCSSE59359.2023.10244955.

39. Mahmoud Shams et al. "Water quality prediction using machine learning models based on grid search method". In: *Multimedia Tools and Applications* (Sept. 2023). DOI: 10.1007/s11042-023-16737-4.

40. P. William et al. "Artificial Intelligence based Models to Support Water Quality Predic- tion using Machine Learning Approach". In: (2023), pp. 1496–1501. DOI: 10.1109/ICCPCT58313.2023.10245020.