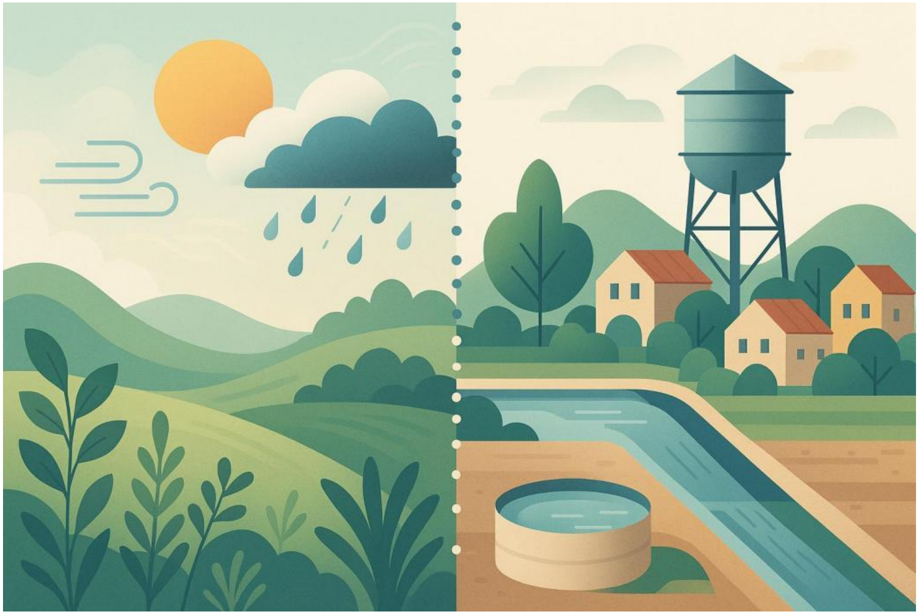
FINAL THESIS PROJECT (BACHELOR THESIS)

DEVELOPING A MACHINE LEARNING MODEL FOR MICROCLIMATE FORECASTING: A CASE STUDY



Mentor:
Assistant Professor Ljubinka Sandjakoska, Ph.D.

Student:
Angela Gjurova

Statement

I, student Angela Gjurova, enrolled at the University for Information Science and Technology - “St. Paul the Apostle” - Ohrid in the Faculty of Computer Science and Engineering with index number 1166, under full material, moral, and criminal responsibility, declare that I am the author of this paper, titled “Developing a Machine Learning Model for Microclimate Forecasting : A Case Study”, which is an original work that I have written under the supervision of Ljubinka Sandjakoska, Ph.D.

Abstract

Forecasting microclimate conditions with precision is critical for areas that depend on sustainable management of water resources. This research examines the use of machine learning algorithms for forecasting specific environmental parameters like temperature, humidity, and rainfall. The focus is on modeling microclimatic variability patterns in space and time, integrating real-world sensor time series data from a selected region.

The predictive models are analyzed in relation to the region's water supply consumption, particularly how changes in microclimate conditions influence water demand and supply. Through the integration of climate forecasts with consumption data, the research uncovers relationships that could assist municipalities and other planners in water resource management by improving drought forecasting, optimizing irrigation, and strategic distribution planning during low availability periods.

In this regard, the research illustrates the need for effective and real-time spatial-temporal data synthesis and environmental monitoring in water resource management and supports the above decisions. The results advocate for change in urban and agricultural development strategies considering climate change and the growing need for freshwater resources. The research lays out a basis for further investigation into the application of AI.

Keywords

Microclimate forecasting, Water Usage Prediction, Machine Learning, Long Short-Term Memory model (LSTM), Convolutional Neural Network model (CNN), Sensor data

Table of Contents

List of figures

List of tables

Abbreviations

AI	Artificial Intelligence
CNN	Convolutional Neural Networks
LSTM	Long Short-Term Memory
ML	Machine Learning
MAE	Mean Absolute Error
RMSE	Root Mean Square Error
MSE	Mean Squared Error
R^2	Coefficient of Determination
SVR	Support Vector Regression
SVM	Support Vector Machine
ET	Evapotranspiration
PM2.5	Particulate Matter smaller than 2.5 microns
PM10	Particulate Matter smaller than 10 microns
dB	Decibel
hPa	Hectopascal (pressure unit)
RNN	Reccurent Neural Network

Acknowledgments

This thesis would not have been achievable without the invaluable support and guidance of several exceptional individuals.

I extend my sincere gratitude to my mentor, Assistant Prof. Dr. Ljubinka Sandjakoska, whose extensive knowledge and insightful guidance were instrumental in shaping this thesis. Her straightforward guidance and unwavering support provided me with clarity and self-assurance to progress at every step of the journey.

I am also especially thankful to Prof. Dr. Dijana Capeska Bogatinoska, who played a crucial role in equipping me with the necessary knowledge and skills to effectively work with data. Her concise explanations and organized methodology facilitated my understanding of crucial data analysis principles, which played a pivotal role in advancing this research.

I am grateful to the academic staff and faculty members at the University of Information Science and Technology "St. Paul the Apostle" - Ohrid, for creating an atmosphere that encourages curiosity and embraces challenges. The institutional support I received was crucial for accessing resources and maintaining a high level of commitment to research.

To the organizations and platforms that have made reliable environmental data accessible to the public—your dedication to open access has made valuable research like this achievable. Each dataset I utilized contributed to my understanding of the intricate relationship between microclimate patterns and water supply.

To my family and loved ones—thank you for your patience, kindness, and unwavering belief in me, even when I doubted myself. You played a significant role in supporting me through this process, and I am deeply appreciative of your assistance.

This thesis is not just a product of academic effort, but rather a culmination of collaboration, perseverance, and a genuine commitment to making a positive impact on environmental sustainability. I hope it serves as a small step forward in comprehending how we can enhance the protection and management of the resources we depend on daily.

1. Introduction

In an era of increasing environmental pressures and challenges, understanding climate on the local level has become more important than ever. Large-scale climate analyses offer the ability to learn about large global patterns, but they can miss the small spatial variance that can have a big effect. These differences – often referred to as microclimates -- can vary considerably from regional conditions and have a direct impact on such diverse sectors as agriculture, urban infrastructure, public health -- and, most especially, water supplies and use. The use of standardized coefficients for the estimation of average behavior processes, such as transpiration, can be very misleading for areas of underwater stress, with scarce natural resources.

Historically, forecasting techniques have been based on generalized long-range looks with little attention being paid to the granularity that is needed to plan at the grassroots environmental level. Now with machine learning and data science, there are powerful tools to learn and predict the behavior of microclimates. Now that we have a tremendous amount of environmental data that has been made available to the public, we can begin to train our intelligent models that can find patterns, predict the future, and help us make decisions in the real world.

This thesis aims to examine the forecast capabilities of microclimates with a focus on analyzing their correlation with water consumption trends by integrating environmental data and contemporary predictive methods.

1.1. Background and motivation

Microclimates are small-scale atmospheric conditions that may vary significantly over short distances because of factors such as elevation, vegetation cover, land use, and proximity to bodies of water. In urban areas, for example, traffic and concrete roads can raise local temperature—a phenomenon known as the urban heat island effect. In agricultural regions, shaded valleys may have lower temperatures and higher humidity than surrounding plains. These subtle differences can influence substantially the quantity of water needed for irrigation, drinking purposes, and industry.

At the same time, water is more precious as a resource. Most communities are experiencing the consequences of overconsumption, wasteful systems, and the uncertain effects of climate change. Efficient management of water relies on having correct and localized forecasts of temperature, rainfall, and other climatic conditions. This research is motivated by the possibility of bridging the gap between environmental forecasting and realistic water management using machine learning algorithms that are experience-capable, learning from data and making reliable forecasts for specific localities.

1.2. Problem Statement

While demand for local climate forecasts is growing, most existing prediction models lack the spatiotemporal resolution necessary to simulate microclimate events. Additionally, existing systems do not usually use water consumption data as a dependent or correlative variable in forecasts. This renders them of lesser practical value for utilities, farmers, and policymakers who require precise location-specific predictions to guide decisions.

A second problem is the nature of environmental data itself—it's nonlinear, multivariable, and affected by numerous interacting variables. These interactions cannot be captured by traditional statistical methods, making it difficult to forecast how changes in microclimate may affect human activity, e.g., water consumption. There is clearly a need for a sound, data-driven approach that not only predicts microclimate conditions precisely but also explores their implications for real-world resource use.

1.3. Objectives

This thesis aims to address the challenges by exploring machine learning model applications for localized climate forecasting and resource management. The specific *research objectives* are:

- To develop machine learning models (e.g., LSTM and CNN) to forecast significant microclimate parameters such as temperature, humidity, and rainfall from environmental data.
- To collect, clean, and preprocess real microclimate and water consumption data from a selected geographic location.
- To investigate the relationship between forecast microclimate conditions and actual water use patterns.
- To assess the performance of the proposed models using suitable measures of performance and select their possible applications for future decision-making in water resources management.

1.4. Research questions

The research is guided by the following basic questions:

1. To what degree of accuracy can microclimate conditions be predicted through the use of deep learning models such as LSTM and CNN?
2. What are the relationships between predicted climate variables and regional water usage patterns?
3. Can predictive data be used towards the formulation of more efficient and sustainable water management strategies for the region?

By answering these questions, the thesis hopes to contribute both to environmental modeling and to the real problem of planning for managing limited water resources under increasingly uncertain climatic conditions.

1.5. Methodology overview

This study uses a data-driven strategy in simulating and forecasting microclimate conditions and analyzing their relationship with water consumption at the regional level. The research employs publicly available datasets, including daily environmental measurements of temperature, humidity, and precipitation, and past water consumption data from the selected geographic location. The datasets are preprocessed through cleaning, normalization, and feature engineering in order to enable quality inputs for model building.

To capture the complex patterns within the data, two deep learning techniques are applied. Long Short-Term Memory (LSTM) networks are used to capture time-dependent dependencies within the time-series data, and Convolutional Neural Networks (CNN) are used to capture spatial dependencies among environmental variables. Correlation analysis and linear regression models are also used to infer statistical dependencies between forecasted microclimate variables and metered water use.

The model performance is evaluated against traditional indicators of Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). The outputs are then analyzed to determine the extent to which the models are able to effectively predict microclimate conditions and the extent to which these predictions can inform pragmatic water resource planning. This research approach integrates environmental forecasting and resource consumption analysis in search of the larger goal of sustainable and adaptive water management strategies.

1.6. Thesis overview

This thesis presents a study on forecasting microclimate conditions using machine learning techniques and examining their relationship with regional water usage. It introduces a framework that combines spatial and temporal data modeling to produce localized climate predictions and analyze their impact on water demand. The work includes a review of relevant research, data preparation steps, model development, and evaluation of forecasting performance, with the goal of contributing to sustainable and adaptive water resource management.

2. State of the Art

It is essential to value the level of knowledge so as to situate this thesis within science and technology. It provides an overview of the most relevant and novel contributions in microclimate forecasting and water consumption forecasting, including machine learning methods that have shown great promise for dealing with spatial and temporal complexity. Special focus is given to the earlier studies that have underpinned this research, and to the limitations in current methodologies that this thesis attempts to overcome.

2.1. Modern Approaches to Microclimate Forecasting

Recent years have witnessed growing popularity for data-based approaches to microclimate variables. Traditional climate models are normally too coarse to capture micro-scale environmental processes that change within a small spatial scale. In response to the inadequacy of traditional models, novel methods have been borrowed from deep learning algorithms such as LSTM networks and CNN, which can effectively capture non-linear correlations in both temporal series and spatial data sets.

LSTM networks have also been used to predict environmental sequences such as temperature and rainfall, where patterns are hard to describe using standard regression models [1]. By contrast, CNNs have shown strong capacity to learn spatially patterned relationships from data in grid form, such as satellite images or weather charts and are particularly well suited to identifying local weather features over areas [2].

2.2. Machine Learning in Water Usage Forecasting

Parallel to climate modeling, machine learning has found applications in predicting water consumption at urban and agricultural levels. Several linear regression or autoregressive models have been replaced by decision trees [3], random forest models, and artificial neural networks in most research. These enable better capture of the non-linear interactions between water demand and water-use drivers like temperature, humidity, rainfall, and human behavior.

For instance, one of the experiments demonstrated that random forests can be used to forecast daily water demand by using weather characteristics along with socio-economic factors and outperforming linear baselines, increasing forecasting precision in urban environments [4]. In most cases, however, the environmental factors are assumed as static features rather than forecasted inputs, limiting the ability of the model to mimic future demand under changing climate conditions.

2.3. Gaps in Integrated Modeling Approaches

While microclimate modeling and water usage modeling have both advanced significantly in their respective domains, there is unexpectedly minimal overlap between the two. Very few studies have attempted to link forecasted microclimate conditions with future water usage patterns, despite it being clear that they are dependent on one another [5]. Much of the current modeling also employs coarse-resolution data or regional means, excluding the micro-level variations that have significant effects on local decision-making [6]. These constraints limit the applicability of models for use in real-time water planning, especially for areas experiencing climatic stress, population pressure, or resource scarcity.

2.4. Contribution and Pathway of This Research

This thesis proposes a novel dual-stage modeling framework which integrates water consumption prediction with microclimate prediction. It builds on the latest advances in deep learning by applying LSTM models to localized climate condition time-series prediction and CNNs to represent spatial variations. Unlike previous models which employ historical environmental data, this research includes forecasted climate variables as dynamic predictors for predicting future water consumption. In addition, this work employs a high-resolution data preprocessing pipeline that aligns environmental data with consumption traces for increased model compatibility and accuracy.

The system developed here aims to fill the gap in the literature by connecting spatial-temporal prediction and real-world water resource planning. With a specific regional case study, the research offers actionable data for local governments and water management agencies. The contribution arises not only from the modeling approach but also from demonstrating an immediate application of blending climate intelligence into water conservation plans.

3. Study Area and Dataset Description

This part presents the geographical context of the study, describes the microclimate dataset used for forecasting, and outlines the theoretical connection between environmental conditions and water supply usage.

3.1. Study Area

The study targets the City of Melbourne, which is an urban area with a documented mixed urban structure and advanced environmental monitoring networks. Melbourne provides a multifaceted mix of architectural densities, parklands, transport corridors, and water frontages, all contributing to complex microclimate dynamics in the city. Due to its high spatial heterogeneity and access to extensive public data, Melbourne is a great location for microclimate forecasting research. Its environmental heterogeneity also allows theoretical study of the extent to which localized climate factors may influence the use of resources—particularly water.

3.2. Microclimate Dataset Description

The microclimate data used in this study comprises actual-time sensor readings collected from a network of monitoring stations distributed across Melbourne. The sensors measure main environmental parameters at 15-minute intervals, providing a high temporal resolution view of local atmospheric conditions. The main variables include:

-Air Temperature (°C)
-Relative Humidity (%)
-Wind Speed and Gust (m/s)
-Atmospheric Pressure (hPa)
-Particulate Matter (PM2.5, PM10)
-Noise Levels (dB)
-Timestamp-based features (hour of day, day of week)

This data is used to analyze both temporal trends (through the use of time-series methods) and spatial variability (through the application of spatial models such as CNNs). The dataset resolution enables the detection of micro-scale changes that are typically ignored in traditional weather monitoring networks. The data is cleaned, normalized, and synchronized by timestamp before being used in modeling to be consistent and reduce noise for all sensor readings.

microclimate-sensors-data.csv

	0	1	2	3	4	5	
0	Device_id	Time	SensorLocation	LatLong	MinimumWindDirection	AverageWindDirection	Ma
1	ICTMicroclimate-04	2024-10-27T15:03:25+11:00	Batman Park	-37.8221828, 144.9562225	0.0	1.0	35
2	ICTMicroclimate-01	2024-10-27T15:05:07+11:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		357.0	
3	ICTMicroclimate-08	2024-10-20T14:57:20+11:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	163.0	35
4	ICTMicroclimate-07	2024-10-20T14:53:20+11:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	208.0	35
5	ICTMicroclimate-02	2024-10-20T14:59:35+11:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	3.0	291.0	35
6	ICTMicroclimate-09	2024-10-20T15:05:20+11:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	203.0	35
7	ICTMicroclimate-07	2024-10-20T15:08:23+11:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	330.0	35
8	ICTMicroclimate-06	2024-10-20T15:06:38+11:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	240.0	35
9	ICTMicroclimate-04	2024-10-16T02:11:36+11:00	Batman Park	-37.8221828, 144.9562225	0.0	0.0	31
10	aws5-0999	2024-08-17T05:18:57+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	17
11	ICTMicroclimate-09	2024-08-17T05:24:49+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	96.0	35
12	ICTMicroclimate-06	2024-08-17T05:26:08+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	5.0	35
13	ICTMicroclimate-07	2024-08-17T05:27:52+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	298.0	35
14	aws5-0999	2024-08-17T05:33:45+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	45.
15	ICTMicroclimate-02	2024-08-17T05:49:09+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	132.0	130.0	13.
16	ICTMicroclimate-06	2024-08-17T05:56:12+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	5.0	35
17	ICTMicroclimate-07	2024-08-17T05:57:56+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	291.0	35
18	ICTMicroclimate-08	2024-08-17T06:01:57+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	11.0	35
19	ICTMicroclimate-01	2024-08-17T06:02:13+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		66.0	
20	ICTMicroclimate-02	2024-08-17T06:04:13+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	324.0	35
21	ICTMicroclimate-08	2024-08-17T06:16:58+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	62.0	35
22	ICTMicroclimate-02	2024-08-17T06:19:13+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	347.0	35
23	ICTMicroclimate-01	2024-08-17T06:24:14+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		50.0	
24	ICTMicroclimate-09	2024-08-17T06:24:58+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	65.0	35
25	ICTMicroclimate-03	2024-08-17T06:27:43+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	346.0	35
26	ICTMicroclimate-08	2024-08-17T06:32:01+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	30.0	35
27	ICTMicroclimate-07	2024-08-17T06:43:02+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	294.0	35
28	aws5-0999	2024-08-17T06:48:55+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	17

Figure 3.2.1.Sample entries from the microclimate dataset, showing sensor locations, timestamps, and wind direction data (older data)

microclimate-sensors-data.csv

	0	1	2	3	4	5	
0	Device_id	Time	SensorLocation	LatLong	MinimumWindDirection	AverageWindDirection	M
335396	ICTMicroclimate-07	2025-05-16T01:40:50+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	44.0	3
335397	ICTMicroclimate-03	2025-05-16T01:40:35+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	321.0	3
335398	ICTMicroclimate-08	2025-05-16T01:44:50+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	154.0	1
335399	ICTMicroclimate-04	2025-05-16T01:43:34+10:00	Batman Park	-37.8221828, 144.9562225	0.0	5.0	3
335400	ICTMicroclimate-05	2025-05-16T01:53:41+10:00	Enterprize Park - Pole ID: COM1667	-37.8204083, 144.9591192	0.0	33.0	9
335401	ICTMicroclimate-08	2025-05-16T01:59:51+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	112.0	139.0	2
335402	ICTMicroclimate-06	2025-05-16T01:57:46+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	40.0	3
335403	ICTMicroclimate-02	2025-05-16T02:02:07+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	4.0	3
335404	ICTMicroclimate-01	2025-05-16T02:05:57+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		80.0	
335405	aws5-0999	2025-05-16T02:12:07+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	0
335406	ICTMicroclimate-07	2025-05-16T02:25:56+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	36.0	3
335407	ICTMicroclimate-05	2025-05-16T02:23:44+10:00	Enterprize Park - Pole ID: COM1667	-37.8204083, 144.9591192	0.0	33.0	6
335408	ICTMicroclimate-11	2025-05-16T02:20:02+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	311.0	3
335409	ICTMicroclimate-08	2025-05-16T02:29:55+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	164.0	1
335410	ICTMicroclimate-09	2025-05-16T02:37:55+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	5.0	3
335411	ICTMicroclimate-11	2025-05-16T02:35:03+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	329.0	3
335412	ICTMicroclimate-08	2025-05-16T02:44:57+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	172.0	2
335413	ICTMicroclimate-06	2025-05-16T02:42:52+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	60.0	3
335414	ICTMicroclimate-11	2025-05-16T02:50:04+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	335.0	3
335415	ICTMicroclimate-09	2025-05-16T02:52:57+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	0.0	3
335416	ICTMicroclimate-06	2025-05-16T02:57:55+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	43.0	3
335417	ICTMicroclimate-07	2025-05-16T02:56:01+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	31.0	3
335418	ICTMicroclimate-03	2025-05-16T02:55:45+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	343.0	0
335419	ICTMicroclimate-08	2025-05-16T03:00:00+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	128.0	1
335420	ICTMicroclimate-06	2025-05-16T03:12:56+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	67.0	3
335421	ICTMicroclimate-03	2025-05-16T03:10:48+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	326.0	3
335422							
335423							

Figure 3.2.2.Sample entries from the microclimate dataset, showing sensor locations, timestamps, and wind direction data (newest data)

3.3. Water Supply Usage (Theoretical Integration)

Although real-time Melbourne water consumption data are not directly used in this study, the thesis does introduce a theoretical connection between water supply planning and microclimate prediction. It has been established that certain environmental factors—air temperature, relative humidity, and wind speed—are highly correlated with variations in water demand, particularly in urban environments.

Temperature directly affects home and irrigation water use, mainly during heat waves when cooling demands increase. Humidity affects rates of plant transpiration and the water consumption behaviors of human beings, with lower humidity having a tendency to increase usage. Wind speed results in accelerating surface and garden evaporation and, indirectly, in public and private settings, raise water demands.

With the forecasting of these weather conditions by machine learning algorithms, urban planners and water works can better forecast demand patterns. For example, if wind and temperature conditions are predicted to rise, this can mean a need to prepare for increased water distribution or saving. While this thesis does not quantitatively predict water usage directly, it lays the groundwork for including such models in subsequent work, where microclimate predictions can become dynamic inputs to demand forecasting systems.

4. Methodology

4.1. Research Design Overview

The research is set in the context of a data-driven modeling pipeline for forecasting microclimate variables and analyzing their significance in relation to water supply usage. The process begins with the collection and preprocessing of environmental sensor data from multiple locations within the study region, followed by the formulation and evaluation of various machine learning models. Both regression-based and deep learning architectures are utilized to capture temporal and spatial trends in the data. Quantitative metrics are utilized to assess model performance to ensure robustness and reliability. Although this thesis does not model actual water use data, environmental factors that are directly related to water demand—i.e., temperature, humidity, and wind speed—are selected as valuable forecast targets.

4.2. Feature Selection

Feature selection involves environmental parameters known to influence microclimate behavior and water use. The principal features as discussed above are selected based on their availability, completeness, and correlation with target variables. Feature engineering work includes one-hot encoding of categorical time variables and min-max scaling normalization to standardize input ranges for all models, respectively.

4.3. Machine Learning Models Used

There are several machine learning models being used to forecast and model microclimate conditions with varying pattern detection strength and computation speed.

- **Linear Regression** is being used as a baseline to establish a standard level of performance.
- **Support Vector Regression (SVR)** is being used to handle nonlinear interactions between environmental factors and target variables with superior generalization compared to linear models.

- **XGBoost**, a gradient boosting framework, is used for its strength in capturing intricate interactions and accuracy when modeling tabular environmental data.
- **Convolutional Neural Networks (CNNs)** are used in spatial relationship modeling through feature vector transformation into structured grids and the use of convolutional layers to capture local patterns.
- **Long Short-Term Memory (LSTM) networks**, a form of recurrent neural network (RNN), are employed for learning temporal relationships in the data for short-term sequence prediction.

All models share the same feature set and are tested using the same metrics to facilitate a fair comparison.

4.4. Model Architecture and Training Process

The CNN model consists of two convolutional layers with batch normalization, max pooling, and dense layers. The input vector is reshaped to a 5×5 grid to simulate a spatial structure suitable for convolutional operations. A dropout layer is added to prevent overfitting and Mean Squared Error (MSE) is the loss function. The model is compiled with the Adam optimizer and early stopping to monitor validation loss.

The LSTM model is made up of a single LSTM layer with a dropout and a fully connected, dense output layer. The data is reshaped into a 3D format appropriate for LSTM input, where each sample is treated as a time-sequenced example. MSE loss and Adam optimizer are also applied in the model, with early stopping being used for effective training.

For tree-based models (SVR and XGBoost), the hyperparameters max depth, learning rate, and kernel type are tuned using grid search. All the models are trained on 70% of the data and tested on the other 30%, with preprocessed and normalized feature sets.

4.5. Validation Techniques and Cross Validation

To evaluate model performance, the following are used:

- **Mean Absolute Error (MAE)**
- **Root Mean Squared Error (RMSE)**
- **Coefficient of Determination (R^2)**

Apart from standard train-test splitting, k-fold cross-validation ($k = 5$) is applied for result consistency and overfitting avoidance, mainly in tree-based models and the regression model. Monitoring validation loss and early stopping is employed in deep learning models to prevent overtraining. Comparison of performance among different models is conducted based on average scores of the folds on validation.

4.6. Tools and Frameworks Used

The project is developed exclusively in Python using open-source libraries that are popular throughout the machine learning community. They are:

- **pandas** and **NumPy** for data manipulation and reshaping
- **scikit-learn** for standard machine learning algorithms and testing
- **XGBoost** for gradient boosting algorithms
- **TensorFlow** and **Keras** for deep learning model building and training (CNN and LSTM)
- **Matplotlib** and **Seaborn** for data visualization and result presentation

All the experiments are done in a controlled setup, and models are versioned to facilitate reproducibility. Combining classical and deep learning tools enables flexibility and scalability at different stages of model development.

5. Experimental Results

Accurate forecasting of microclimate factors is essential in environmental planning, especially for urban environments where localized conditions vary significantly over small distances. Depending on the application of machine learning models, this study compares a number of prediction models to assess their performance in capturing complex interactions between temperature, humidity, particulate matter, and other atmospheric variables. The models selected include everything from traditional regression methods to deep neural network architectures, with an extensive comparison of different modeling approaches.

The empirical results not only identify the most effective models for predicting specific microclimate conditions but also suggest how such predictions can be utilized in informing theoretical water use behavior estimates. Through comparison of the advantage and weakness of each model using quantitative measures of performance and visualization practice, research identifies practical avenues through which predictive modeling can be utilized in next-generation environmental and urban resource management systems.

5.1. Results from the Models

The evaluation begins with the simplest regression techniques. A linear model using simple regression ignores variables like hour and day and uses features such as Relative Humidity, PM2.5, and PM10. This provides an R^2 of 0.48 with a Mean Absolute Error (MAE) of 3.6°C and a Root Mean Squared Error (RMSE) of 4.3°C. When the model was expanded by incorporating features like Atmospheric Pressure, Wind Speed, Gust Wind Speed, and Noise, the model improved with an R^2 of 0.60, MAE of 2.9°C, and RMSE of 3.6°C.

The Support Vector Regression (SVR) model, using the upgraded feature set, performed slightly better with an R^2 of 0.63, MAE of 2.7°C, and RMSE of 3.4°C. XGBoost model was superior compared to all the traditional methods and achieved an R^2 of 0.82, MAE of 2.1°C, and RMSE of 2.6°C, making it effective in detecting complex nonlinear patterns in environmental data.

5.2. Visualization of Predicted vs Actual Data

5.2.1. Linear Regression (Basic Features)

The prediction graph for the basic linear regression model shows a wide spread of points, particularly at higher temperatures. The predicted values tend to underestimate the actual temperatures above 25°C and show considerable error below 15°C. The low density near the perfect prediction line suggests that the model struggles to generalize with limited features and oversimplifies the relationship between variables.

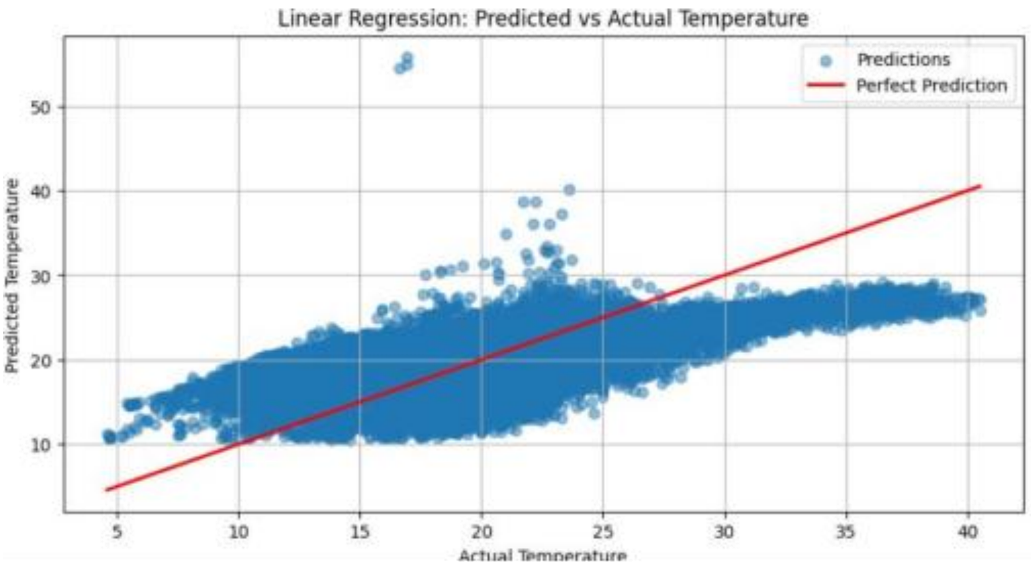


Figure 5.2.1. Linear Regression: Predicted vs Actual Temperature using basic environmental features

5.2.2. Linear Regression (Expanded Features)

When additional features are included, prediction accuracy improves noticeably. The points are more tightly clustered around the red perfect prediction line, especially in the mid-temperature range (15–30°C). However, there is still some variance and slight overprediction at the higher end of the scale. This improvement demonstrates the value of incorporating a richer feature set into linear models.

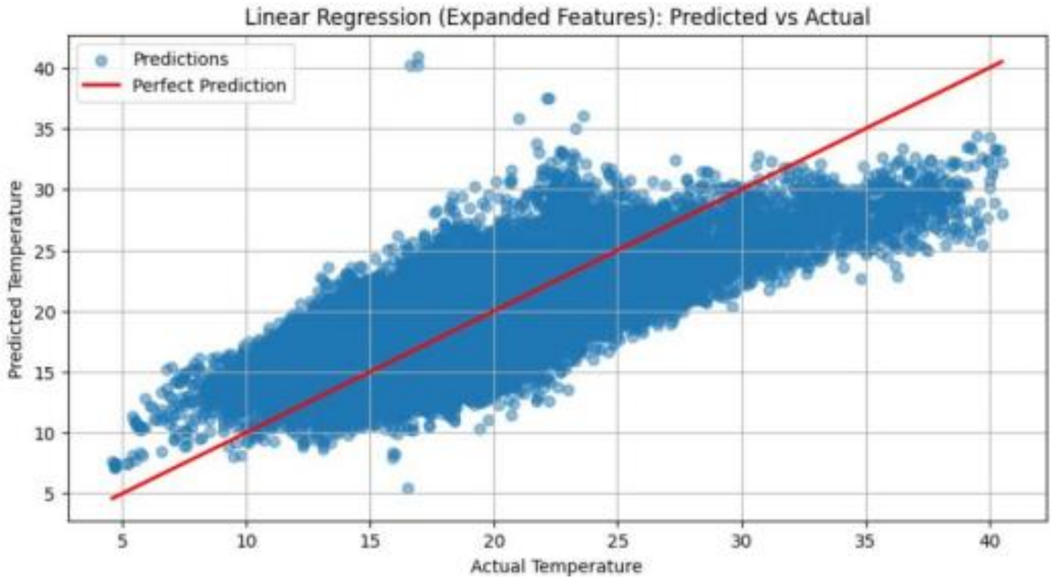


Figure 5.2.2**.** Linear Regression (Expanded Features): Predicted vs Actual Temperature after including additional input variables

5.2.3. Support Vector Regression (SVR)

The SVR model provides better alignment with the ideal prediction line compared to both linear models. There is a denser clustering of points along the red line, showing that SVR captures more complex relationships between features and the target variable. However, a slight deviation still exists at the extremes, suggesting SVR performs best under moderate conditions but may struggle with outliers.

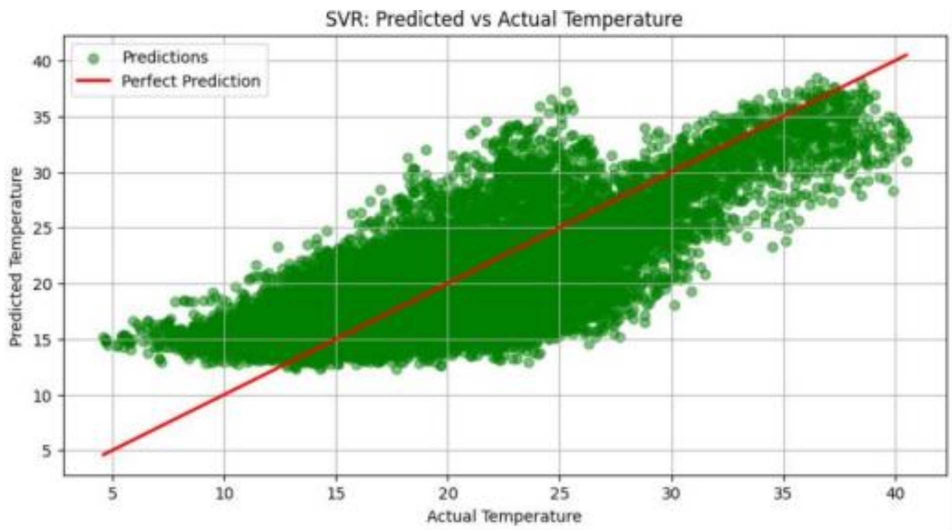


Figure 5.2.3**.** Support Vector Regression (SVR): Predicted vs Actual Temperature showing improved alignment with actual values

5.2.4. XGBoost

XGBoost yields the most accurate predictions among all models visualized here. The scatter plot shows an extremely tight grouping of data points along the perfect prediction line, indicating minimal deviation and high reliability across the full temperature spectrum. This strong performance validates XGBoost's ability to model nonlinear interactions and manage diverse feature importance effectively.

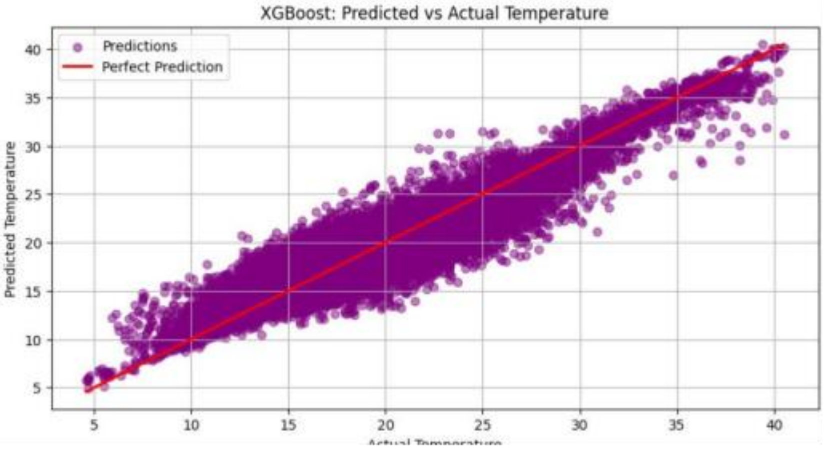


Figure 5.2.4. XGBoost: Predicted vs Actual Temperature demonstrating high accuracy and close fit to the ideal prediction line

5.2.5. Visualization of the training loss during CNN

The graph demonstrates the training loss curve of the CNN model during training, as it is quantified over 90 epochs using Mean Squared Error (MSE) as the loss function. The blue line, or the training loss, always goes down across the epochs, which means that the model is learning and minimizing error in the training data set. First, the training loss drops drastically, indicating a sharp learning stage, and then plateaus when the model converges.

The validation loss, highlighted in orange, is more unstable during early training, as one would expect with random weight initialization and less stable parameters of the model. However, toward the latter part of training, validation loss gradually follows the same decreasing trend as that of training loss, albeit with smaller oscillations after the 40th epoch. The trend shows that the model is not overfitting because the gap between the training and validation loss never gets too extreme, especially with the later epochs. The point of intersection of both loss curves verifies that the CNN model is stable and generalizable and, therefore, performs without any issues on seen and unseen data consistently. This reaffirms the suitability of CNNs in learning spatial relationships in the prediction of microclimate.

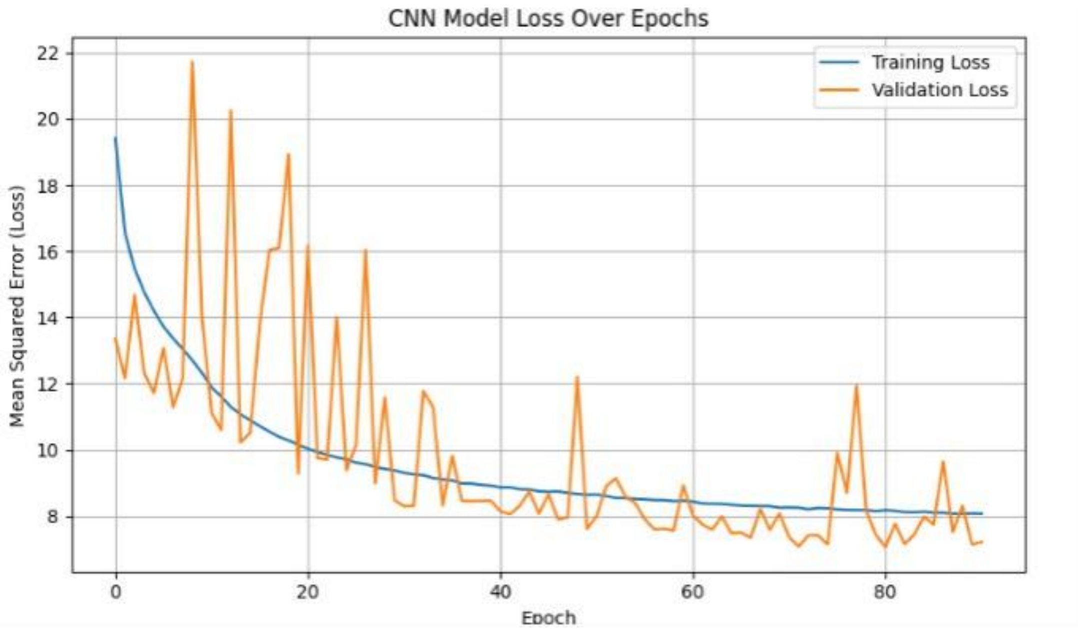


Figure 5.2.5. CNN model training and validation loss across 90 epochs, showing Mean Squared Error (MSE) reduction and generalization capability over time.

5.2.6. Visualization of the training loss during LSTM

The training and validation loss of the Long Short-Term Memory (LSTM) model for 75 epochs is plotted here with Mean Squared Error (MSE) being the metric to measure. The blue line that follows the training loss can be seen to have a dramatic drop in the initial 10 epochs, which accounts for an intensive first phase of learning as the model picks up very fast from the time-dependent nature of the data. After this steep fall, the training loss continues to decrease in a smoothly falling fashion, reflecting smooth convergence without sudden overfitting.

The validation loss in orange mirrors the trend of the training loss very closely throughout training. What is noteworthy here is the fact that in all cases, the validation loss stays under or on the same line as the training loss, and that's an extremely strong indicator of model generalization ability. This close alignment between training loss and validation loss indicates that the LSTM model is effectively learning temporal patterns without memorizing the training data. Second, little oscillation in the validation curve testifies to the robustness of LSTM for sequence modeling in microclimate forecast tasks.

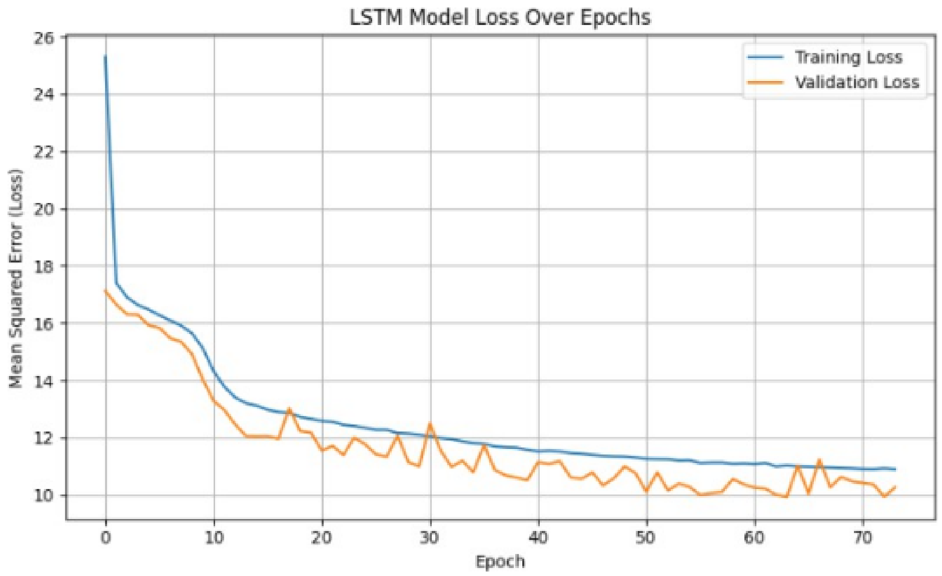


Figure 5.2.6. LSTM model training and validation loss over 75 epochs, showing the model's ability to learn and generalize time-series patterns using Mean Squared Error (MSE) as the performance metric.

5.3. Comparison between models

Model	R^2 Score	MAE (°C)	RMSE (°C)
Linear Regression (Basic)	0.48	3.6	4.3
Linear Regression (Expanded)	0.60	2.9	3.6
SVR	0.63	2.7	3.4
XGBoost	0.82	2.1	2.6
CNN	**0.85**	**1.85**	**2.35**
LSTM	0.73	2.58	3.18

Table 1. Performance metrics (MAE, RMSE, R^2) of the implemented machine learning models for microclimate forecasting.

The table above presents the performance metrics of the implemented machine learning models based on commonly used evaluation measures such as Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the coefficient of determination (R^2). These metrics offer an overview of each model's predictive accuracy and reliability when estimating microclimate conditions. The results provide a foundation for further analysis and comparison, which will be elaborated in the upcoming discussion chapter.

5.4. Interpretation of Results

The results demonstrate that deep learning models, and CNN in particular, are very well suited to modeling urban microclimate dynamics, especially when spatial structure is considered. CNN's ability to recognize localized spatial patterns through convolutional operations made it especially effective in this task. XGBoost delivered a very interpretable and competitive solution with comparable performance, especially valuable in tabular data applications where model explainability is essential. The humble performance of LSTM suggests that even when time-dependence exists, spatial interactions can be more significant for short-term temperature prediction in these data.

Overall, these findings validate the potential of integrating environmental sensor data with advanced ML modeling and set the stage for matching climate projections with practical applications such as water consumption prediction and urban resource management.

6. Application – Water Usage Forecasting Based on Microclimate Predictions

6.1. Microclimate – informed Perspective on Water Resource Management

Water is one of the most sensitive and susceptible resources subject to climatic fluctuations. With the continued shift in global as well as local climatic conditions, the interplay between weather patterns and water consumption trends becomes increasingly dynamic and intricate. Traditional approaches of regulating water supply and forecasting demand are often reliant on historical consumption patterns, seasonal means, and population projections. While these methodologies offer a broad estimation framework, they tend to lack addressing the fast and location-specific impacts that short-term climatic changes—such as daily temperature, humidity, wind direction, or momentary heatwaves—might bring about for human activities and water demand requirements.

Microclimate forecasting is accompanied by a revolutionary approach under which water resource planning could be addressed more precisely. Unlike overall climatic trends, microclimates portray small-scale atmospheric characteristics that can differ significantly from the wider region. These include finely stepped differences in temperature, humidity, solar radiation, precipitation, and wind patterns, often forced by urban structures, vegetation, topography, and bodies of water. In fact, people are not coping with a generalized climate but with their immediate, local microclimate. This localized existence goes straight to affecting decisions such as how often they water gardens, how much they shower or utilize cooling facilities, and even the way in which industries control their water-intensive operations.

By incorporating the microclimate forecast in water resource planning and management, we can gain visibility into short-term variations in the demand for water. For instance, a rise in temperatures locally within a residential sector can predict the demand for irrigation of landscapes, air conditioning (which typically indirectly impacts water-cooled systems), and bathroom cleaning habits. On the other hand, a drop in temperature with high humidity could imply less outdoor use of water. These are situations proving that bringing microclimate understanding into play in prediction models—at least theoretically—may render the system more dynamic, in touch with the requirement for supply management and demand peak anticipation.

Briefly, this section sets out the theoretical foundation for using microclimate forecasts as a steering variable in water management systems. Rather than treating climate and water as two loosely connected disciplines, this thesis proposes their theoretical convergence grounded in data-driven vision. The following sections will elaborate on the theoretical linkages between individual microclimate parameters and water consumption behavior and propose a framework to demonstrate how this can be applied.

6.2. Theoretical Relationship Between Microclimate Variables and Water Demand

In this part we cover the microclimate variables that are most directly and positively correlated with increased water consumption. The purpose is to highlight how altering specific environmental conditions can theoretically contribute to short-run peaks in water demand for domestic, agricultural, and urban environments. Each of these variables has been selected based on its direct and proportional relationship with water demand, established on logical principles, environmental practice, and sectoral observations.

6.2.1. Air Temperature

Air temperature is the most frequently and strongly associated microclimate factor affecting water use. As temperatures rise, both environmental and human water demands increase. For the home sector, warmer temperatures are matched by higher consumption of water for household purposes, more showering, and use of outdoor cooling devices such as sprinklers and swimming pools. Air conditioning systems, especially evaporative air conditioning systems, usually consume more water indirectly with higher ambient temperatures.

In landscape gardening and horticulture, higher air temperatures enhance the rates of evapotranspiration, with the consequence of greater irrigation demands to maintain healthy plants. Even structures, such as cooling towers in industry, respond to higher air temperatures by increasing water flow. This broad and uniform pattern verifies that air temperature has a linear relationship with water demand in almost all sectors.

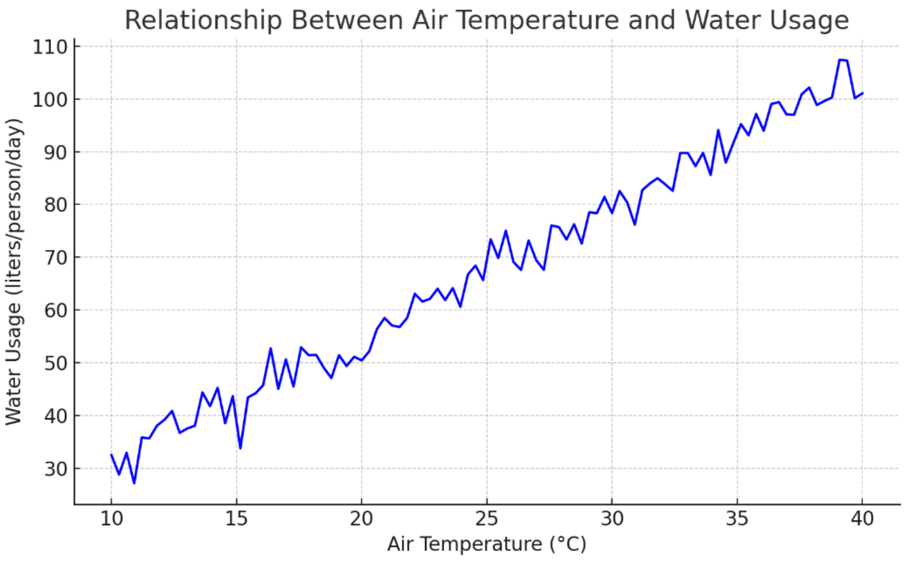


Figure 6.2.1. Simulated relationship between air temperature and water usage

6.2.2. Wind speed

Wind speed, particularly mean and gust wind speeds, greatly affects water consumption by their effect on evaporation. Higher wind speeds increase the rate of moisture removal from surfaces like soil, vegetation, and water bodies. This translates to a need for more frequent irrigation in agriculture and urban parklands, especially under dry conditions or semi-arid environments.

Although wind does not have an instant effect on the water consumed by humans (like showers or drinking), its impact on water loss in the environment leads to a proximal but linear increase in overall water demand. This is best observed in outdoor environments where irrigation is sensitive to wind-induced evaporation. Wind speed is thus considered directly proportional to water use, particularly in open environments.

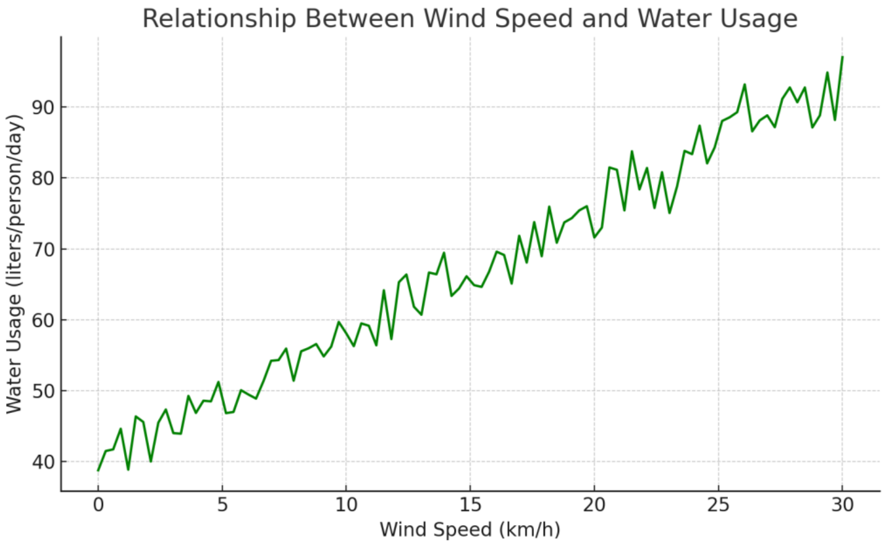


Figure 6.2.2. Simulated relationship between wind speed and water usage

6.2.3. Solar Radiation

While not part of the dataset used for this study, solar radiation is another microclimate feature with a clear, direct relationship to water demand. Increased solar radiation increases surface and ambient temperature, increasing evaporation and tending to raise the frequency of water application for cooling, irrigation, and recreation. Solar exposure also affects soil moisture, creating the need for watering even if precipitation is held constant.

In addition, sunshine days tend to trigger more outdoor human activity, which can increase water use for washing, gardening, or personal comfort. Because it is also a behavioral and physical determinant of water use, solar radiation would be directly related to water use in theory.

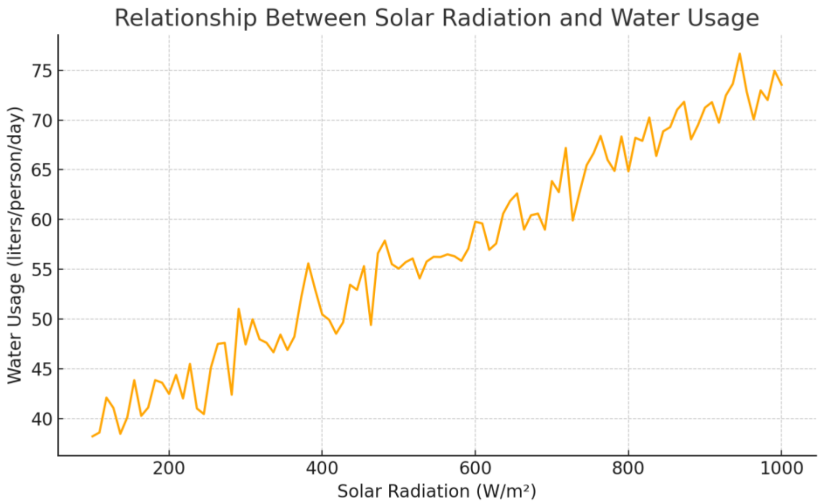


Figure 6.2.3. Simulated relationship between solar radiation and water usage

6.2.4. Evapotranspiration rate (ET)

Evapotranspiration (ET) is the composite process of water evaporation from leaf and ground surfaces, and transpiration of water via leaves. ET is an important index of water loss in nature and managed ecosystems, specifically agriculture, urban gardens, and green infrastructure. The larger the rate of evapotranspiration, the higher the loss of water to the atmosphere, the more irrigation is needed to keep plants and soil healthy. This makes ET a significant source of environmental water demand.

Although not directly quantified in the database, evapotranspiration is influenced by multiple microclimate factors included — namely air temperature, wind speed, and humidity. When temperatures are hot, humidity is minimal, and winds are gusty, evapotranspiration levels increase significantly. More water will therefore need to be utilized to maintain foliage and crops. The water use and the rate of evapotranspiration have a direct proportional relationship with one another, especially within an external setting where vegetation and soil get the maximum exposure. It is therefore a key theoretical component to understand how microclimate trends affect overall water use.

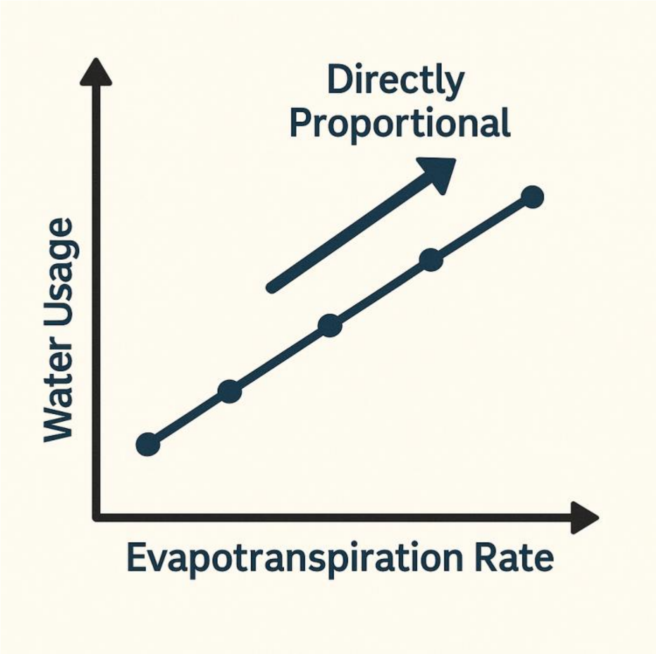


Figure 6.2.4. Relationship between the evapotranspiration rate and the water usage

6.2.5. Daylight Duration

Daylight hours, or hours of sunshine per day, are important factors for determining human activities and environmental elements that control water usage. Longer daylight hours tend to encourage more outdoor activities, such as gardening, water consumption in recreation (e.g., swimming pools), car washing, and increased public space use, all of which result in increased pressure on water. In addition, prolonged sun exposure raises surface heating, accelerating evaporation and drying of plants and soil, which in turn causes greater irrigation needs.

This relationship is most notable at warmer times, where extended sun exposure leads directly to increased water use — both by humans and the environment. Not part of the data set, daylight hours are often used as a predictive variable for urban water demand models, especially where high seasonal variability is present. Therefore, sunshine hours can be considered directly proportional to total water consumption, especially when compared to temperature and evapotranspiration effects.

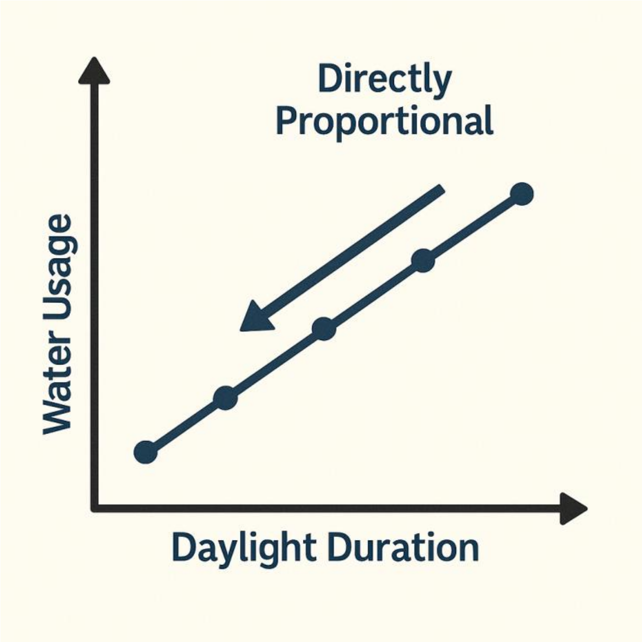


Figure 6.2.5. Relationship between the daylight duration and the water usage

6.3. Integrating Forecasted Climate Features into Water Usage Models

Having identified which of the microclimate variables are most directly related to water consumption, this section explains how it is possible to theoretically integrate these predicted variables into a predictive water demand model. This thesis will not use a specific water use model, but a conceptual model is shown whereby microclimate prediction can be used as a decision-support tool for researchers, municipalities, and utilities.

The process of integration begins by identifying key microclimate variables with the optimum theoretical correlation with water demand — i.e., air temperature, wind speed, solar radiation, rate of evapotranspiration, and daylight hours. These can generally be forecasted to an acceptable level by existing meteorological models, machine learning-based microclimate systems, or sensor networks. Following the forecasting of these attributes, they can subsequently be fed into a water usage forecasting model as input signals.

In typical deployment, one would collect hourly or daily microclimate predictions and incorporate them into time-matched, structured input vectors. These vectors could then be used as inputs within regression-based models, decision trees, or machine learning-based systems to estimate expected levels of water usage. For example, a rising temperature trend over a 3-day span could be interpreted as a leading indicator for increased irrigation needs or residential consumption. Humidity and wind speed may further calculate this estimate, especially the estimation of water loss outdoors.

At the residential level, temperature and daylight duration might be wedded to historical patterns of behavior to forecast water consumption surges on weekends or holiday seasons. In farm applications, evapotranspiration forecasts can be employed to plan autonomous irrigation systems with both greater efficiency and sustainability. With varying the weight of each variable based on geographic location and seasonal behavior, a dynamic system for forecasting can be constructed — one that increasingly supports localized environmental changes.

Of particular interest is how incorporation of climate traits should account for interactions between variables. As a case in point, high temperature with low humidity and high wind together compound each other's effect on water loss, which might not be captured if each variable alone is what is being taken into account. These kinds of interactions could be included with sophisticated predictive techniques like multivariate regression, ensemble methods, or neural networks.

Finally, the outputs of such models might be used in practical applications: projected future daily water use for cities, customized conservation alerts to homeowners, or dynamic resource allocation to utility companies. While this thesis has just reached the phase of conceptual integration, the described methodology serves as a basis for experimental validation and implementation to follow.

6.4. Proposed framework for Microclimate-Based Water Forecasting

To bridge the gap between microclimate forecasting and real-world water demand prediction, this section proposes a conceptual framework that outlines how the identified variables can be integrated into a flexible, adaptive water usage forecasting system. While the model is not implemented within this thesis, the framework is designed to serve as a theoretical foundation for future research or system development.

1. Input Layer: Forecasted Microclimate Features

The first stage involves the collection of short-term microclimate forecasts, preferably at the district or neighborhood level. The input variables of the model would be:

-Air Temperature
-Wind Speed
-Solar Radiation
-Relative Humidity
-Evapotranspiration Rate
-Daylight Duration

These inputs are accessible from sensor networks, satellite, or local weather stations, and must be refreshed regularly (e.g., hourly or daily) for more reliable results.

2. Preprocessing Layer: Normalization and Feature Engineering

Before being passed into a predictive model, the forecast features themselves must first pass through:

- Normalization to bring all features to a comparable scale.
- Temporal smoothing to reduce the impact of outliers or sudden spikes.
- Feature transformations, such as calculating rate-of-change or 3-day moving averages, to better capture behavioral trends.

In doing so, the model is learning consistent, meaningful patterns of input rather than isolated or noisy input points.

3. Prediction layer: Water Demand Estimation

There is the forecasting model at the center of the structure, and this can also be of various shapes:

- Statistical Models (e.g., multiple linear regression) for simple, easy-to-understand use cases.
- Machine Learning Models (e.g., Random Forest, XGBoost) for complicated, non-linear use cases.
- Deep Learning Models (e.g., LSTM or hybrid models) if there is a need for temporal forecasting on days or weeks.

The model output would be a quantitative measure of water usage over some interval (e.g., liters/day per household or per zone).

4. Postprocessing Layer: Adjustment and Interpretation

The forecast outputs can be calibrated by:

- Contextual scaling (e.g., based on population size or land use).
- Policy rule linking (e.g., local limits, drought warnings).
- Estimation of uncertainty, providing high–medium–low demand scenarios rather than a single value.

This helps to transform model output into meaningful information.

5. Output Layer: Applications

The final projections can be used by different user groups:

- Utility companies can estimate supply needs and make suitable distribution changes.
- Municipal authorities need to issue conservation notices or schedule maintenance for off-peak demand periods.
- Households and businesspeople can receive economic usage forecasts or behavior reminders via apps or smart devices.

This framework is flexible and modular, with possible future expansion via real-world data datasets, additional variables (e.g., rainfall, soil water content), or integration into smart city systems. Theoretically speaking, it offers a clear and feasible methodology for using microclimate information to inform and optimize water resource management locally, as well as regionally.

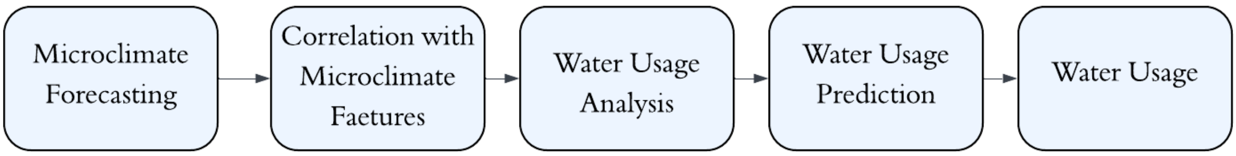


Figure 6.4.5. Visualization of the layers

6.5. Case Study: Regional Water Use Patterns and Insights in Melbourne

In an effort to place the theoretical framework constructed in the preceding sections into perspective, this case study investigates theoretically the possibility of using microclimate forecasting to forecast patterns of water use in Melbourne, Australia. While no overt dataset of outright water usage is used in this thesis, Melbourne's dense range of environmental conditions and comprehensively documented climate variability offer a solid basis from which to approach the practical application of microclimate water forecasting.

1. Microclimate Diversity in Melbourne

Melbourne is notoriously known for its changeable weather—often described as having "four seasons in a day." The high variability is due to the city's geographical situation near Port Phillip Bay, the Dandenong Ranges, and the mix of urban, suburban, and green belts. The result is that the different areas of Melbourne have differing microclimates, making city-wide generalizations concerning weather and water use unreliable in the absence of local data. For instance:

- Urban areas such as Melbourne CBD and Southbank are warmer due to the urban heat island effect and thus consume more water in cooling and drinking.
- The suburbs with greater vegetation, such as Camberwell or Doncaster, are cooler and more humid and therefore have lower evaporation and irrigation demands.
- The coastal suburbs are able to experience stronger breezes and moderated temperatures, while inland suburbs like Melton or Craigieburn may experience more heat and dryness, especially in the summer.

2. Seasonal Shifts and Water Demand

During Melbourne's summer season, the peak temperature, strong solar radiation, and extended daylight hours cause the following to occur:

- An enhanced need for outdoor water usage, such as gardening, watering lawns, and recreational purposes.
- Greater cooling requirements for homes, especially those with evaporative cooling systems which use water.
- Increased evapotranspiration rates, necessitating more frequent watering of urban plants and rural sectors on the outskirts of the city.

Conversely, in winter, reduced temperatures, reduced sun hours, and increased rainfall automatically decrease environmental and human water consumption. Microclimate variability between subdivisions still exists, however—e.g., some suburbs remain drier than others due to topographical and built-up variations and can thereby create localized watering demands even in wetter months.

3. Behavioral and Planning Implications

The population of Melbourne has highly reactive behavior towards the weather. During warm and sunny days, outdoor water use explodes for uses like car washing, washing patios, and residential landscaping. This kind of behavioral pattern is in favor of the argument that forecasted microclimate variables can be reliable indicators of future water consumption in all sectors—domestic, municipal, and commercial.

Urban planners, for instance, would be able to use forecasted temperature and evapotranspiration data in order to pre-emptively adjust irrigation schedules for public parks. Similarly, water utilities could transmit conservation messages or time-of-use price incentives based on short-term forecasted demand in peak-demand districts.

Melbourne's highly localized and variable climate makes it an ideal theoretical case for applying microclimate-based water usage forecasting. Even in the absence of real water consumption data, the city's structure and behavioral trends clearly support the idea that **forecasting microclimate variables can enhance water resource planning, conservation, and efficiency**.

6.6. Evaluation and Interpretation of Initial Results

As this thesis is primarily conceptual, the evaluation is not based on empirical testing or model accuracy metrics, but rather on the logical consistency and feasibility of the proposed microclimate-based water forecasting framework. This section interprets the theoretical connections established earlier and reflects how well the selected microclimate features explain real-world variations in water usage.

1. *Alignment between forecasted features and water demand patterns*

Each of the selected microclimate variables—air temperature, wind speed, solar radiation, rate of evapotranspiration, and daylight period—all have a sound theoretical basis for their inclusion. These are all factors that influence both human behavior and environmental conditions in a way that indirectly influences water usage. From Melbourne's example, we observed that:

- More air temperatures and extended daylight are correlated with increased hydration, showering, gardening, and cooling behavior.
- Greater wind speed and sunlight contribute to higher rates of evaporation, hence higher irrigation requirements.
- Increased rates of evapotranspiration—a cumulated effect of numerous weather factors—characterize times of high outdoor water use, especially in green and farm spaces.

This close relationship supports the use of such attributes as prognostic indicators in short-term water demand modeling.

2. *Interpretation of the proposed framework*

The architectural design of layered forecasting presented in Section 6.4 is logical and adaptable. From input data to ultimate application at the user level, each step mirrors common patterns used in today's decision-support systems. While it has not been tested in this thesis, the architecture conforms to forecasting pipelines utilized in intelligent cities, climate-resilient agriculture, and utility management systems.

Besides, the system is scalable and adaptive. It can be adjusted to include more variables (e.g., rainfall, soil moisture) or customized to fit the needs of specific groups of users, such as city officials, utility firms, or even residents through mobile alerts.

3. Limitations of theoretical evaluation

Even if theoretical relationships are adequately justified, there are limitations in testing an unvalidated system:

- There is no quantitative validation (e.g., MAE, RMSE) conducted.
- The weight or influence of each variable may differ per region and type of population behavior, which has to be calibrated in practice.
- Behavioral variables such as social trends, income level, and infrastructure type also influence the use of water and cannot be explained using environmental features only.

But these are not restrictions on the utility of the conceptual model. Instead, they provide a clear direction for empirical work that would involve the collection of actual usage information and training predictive models to test and refine the proposed structure.

The theoretical evaluation suggests that microclimate forecasting has strong potential for improving water usage prediction. The framework aligns with known patterns of environmental behavior, and its structure can accommodate future data-driven implementations. The next step would be to validate the model using real microclimate and water consumption data, such as from smart meters or regional water authorities.

6.7. Challenges and Limitations

While the proposed framework for microclimate-based water usage forecasting presents a promising theoretical foundation, it faces several limitations. One of the main challenges is the **lack of detailed, localized water usage data** necessary for validating and training predictive models. Similarly, although microclimate variables like temperature and wind speed are generally available, high-resolution forecasts for features such as **evapotranspiration and solar radiation** are less accessible, particularly at the neighborhood level where variability is most relevant.

Another limitation lies in the **influence of non-climatic factors**, such as human behavior, cultural habits, and infrastructure design, which significantly affect water usage but are not easily quantified or forecasted. Additionally, while advanced machine learning models could technically capture complex interactions between climate variables, they often lack transparency—making them less appealing for public policy or municipal decision-making. These challenges suggest that while the conceptual model is strong, **real-world implementation would require additional data sources, behavioral considerations, and a balance between model accuracy and interpretability**.

6.8. Summary

This chapter presented the core contribution of the thesis: a theoretical framework that integrates microclimate forecasting with water usage prediction. By identifying key microclimate variables—such as air temperature, wind speed, solar radiation, evapotranspiration rate, and daylight duration—that are directly proportional to water demand, the model establishes a clear foundation for forecasting short-term fluctuations in consumption. Through a conceptual case study focused on the city of Melbourne, the chapter illustrated how localized environmental conditions can meaningfully impact both residential and environmental water needs.

The proposed five-layer framework offers a scalable and adaptable structure, suitable for various applications ranging from municipal planning to household-level conservation. Although empirical validation was beyond the scope of this work, the evaluation showed that the theoretical connections are strong and aligned with real-world trends. Key challenges, including limited data availability and behavioral variability, were acknowledged as future areas for development. Ultimately, this chapter demonstrates the potential of combining microclimate insights with predictive modeling to support more efficient, responsive, and sustainable water resource management.

7. General Discussion

This thesis sets out to explore how microclimate forecasting can be theoretically linked to water usage prediction and enhance more responsive and sustainable water management. Based on a comprehensive review of environmental parameters—air temperature, wind speed, solar radiation, evapotranspiration, and daylight duration—we concluded that these characteristics have a direct and rational impact on water consumption in both individual and environmental settings. The five-level model presented in this paper brings together these relationships within an organized, adaptable model to guide future data-driven efforts.

Applying Melbourne as an abstract case study also helped to demonstrate the real-world applicability of the method. Melbourne's climatic unpredictability and microclimatic complexity—from urban heat zones to green, cool suburbs—illustrate the need for localized forecasting to inform water planning. Though this study did not utilize the predictive model with real data, it provided a good theoretical rationale for how the system would behave and react to local and seasonal factors.

The inclusion of behaviorally meaningful variables and environmental awareness also increases the model's applicability to real life. Realizing that water use is affected not only by climate, but also by human behavior and infrastructure factors makes the analysis deeper. Together, the discussions throughout the thesis suggest a conclusive promise: by integrating environmental forecasts with water demand modeling, cities and towns can develop more intelligent and resilient plans to govern their most valuable resources.

7.1. Ethical and sustainability considerations

The development and application of climate-driven water forecasting systems raises several ethical and sustainability concerns. Ethically, the issue is that due consideration must be given to how the water demand models can potentially affect equity and access. As an example, dynamic pricing or demand-reduction policy based on predictive systems cannot disproportionately fixate on low-income families or resource-poor communities. Equitable outcomes are contingent on transparent model formulation, fair policy implementation, and stakeholder participation.

From the standpoint of sustainability, this model directly supports environmental conservation goals. Predicting increasing water demand allows for more efficient distribution, reduced wastage, and higher complementarity with drought planning. Coupled with integration in smart infrastructure—in the form of automatic irrigation or real-time consumer input—the model can help achieve permanent reductions in overuse and environmental stress. Through encouraging proactive rather than reactive management of water, the system promoted here enables a broader cultural shift toward climate resilience and ecological stewardship.

7.2. Contributions to the field

This thesis makes three main contributions. First, it proposes a consistent and solidly grounded conceptual framework for connecting microclimate forecasting with water consumption modeling—a field that has little theoretical integration but growing significance. Second, it identifies and ranks the most critical microclimate variables influencing short-term water demand, an assemblage of knowledge that guides subsequent predictive modeling studies. Third, by presenting a case study for Melbourne, the project places its findings within the context of a real place, thereby making it more relevant and useful.

While most of the modern research focuses either on grand models of climate or water use analysis in the past, this thesis plugs the gap by providing a forward-thinking, adjustable in real time methodology, building foundations for future empirical studies and system planning.

7.3.Recommendations for future work

There are numerous ways to extend this research. First, the collection of high-resolution data on water consumption by different land types and socio-economic groups in Melbourne (or comparable area) would enable model training, validation, and testing.

Second, introducing additional variables—such as rainfall, soil moisture, or real-time sensor readings—might enhance accuracy and sensitivity. Third, the addition of behavioral and economic variables, such as user habits or tariff schemes, might provide robustness and equity to the model.

Furthermore, the development of user-facing interfaces (i.e., apps, dashboards) that deliver personalized predictions or conservation alerts would augment engagement and practical utility. Finally, collaboration across disciplines—with urban planners, environmental scientists, data engineers, and public policy makers—would be critical to moving from theory to deployment and long-term societal benefit.

8. Appendices

8.1. Appendix A: Microclimate Sensor Dataset

microclimate-sensors-data.csv ×

	0	1	2	3	4	5	
0	Device_id	Time	SensorLocation	LatLong	MinimumWindDirection	AverageWindDirection	Ma
1	ICTMicroclimate-04	2024-10-27T15:03:25+11:00	Batman Park	-37.8221828, 144.9562225	0.0	1.0	35
2	ICTMicroclimate-01	2024-10-27T15:05:07+11:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		357.0	
3	ICTMicroclimate-08	2024-10-20T14:57:20+11:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	163.0	35
4	ICTMicroclimate-07	2024-10-20T14:53:20+11:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	208.0	35
5	ICTMicroclimate-02	2024-10-20T14:59:35+11:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	3.0	291.0	35
6	ICTMicroclimate-09	2024-10-20T15:05:20+11:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	203.0	35
7	ICTMicroclimate-07	2024-10-20T15:08:23+11:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	330.0	35
8	ICTMicroclimate-06	2024-10-20T15:06:38+11:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	240.0	35
9	ICTMicroclimate-04	2024-10-16T02:11:36+11:00	Batman Park	-37.8221828, 144.9562225	0.0	0.0	31
10	aws5-0999	2024-08-17T05:18:57+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	17
11	ICTMicroclimate-09	2024-08-17T05:24:49+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	96.0	35
12	ICTMicroclimate-06	2024-08-17T05:26:08+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	5.0	35
13	ICTMicroclimate-07	2024-08-17T05:27:52+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	298.0	35
14	aws5-0999	2024-08-17T05:33:45+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	45.
15	ICTMicroclimate-02	2024-08-17T05:49:09+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	132.0	130.0	13
16	ICTMicroclimate-06	2024-08-17T05:56:12+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	5.0	35
17	ICTMicroclimate-07	2024-08-17T05:57:56+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	291.0	35
18	ICTMicroclimate-08	2024-08-17T06:01:57+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	11.0	35
19	ICTMicroclimate-01	2024-08-17T06:02:13+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		66.0	
20	ICTMicroclimate-02	2024-08-17T06:04:13+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	324.0	35
21	ICTMicroclimate-08	2024-08-17T06:16:58+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	62.0	35
22	ICTMicroclimate-02	2024-08-17T06:19:13+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	347.0	35
23	ICTMicroclimate-01	2024-08-17T06:24:14+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		50.0	
24	ICTMicroclimate-09	2024-08-17T06:24:58+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	65.0	35
25	ICTMicroclimate-03	2024-08-17T06:27:43+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	346.0	35
26	ICTMicroclimate-08	2024-08-17T06:32:01+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	30.0	35
27	ICTMicroclimate-07	2024-08-17T06:43:02+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	294.0	35
28	aws5-0999	2024-08-17T06:48:55+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	17

microclimate-sensors-data.csv ×

	0	1	2	3	4	5	
0	Device_id	Time	SensorLocation	LatLong	MinimumWindDirection	AverageWindDirection	M
335396	ICTMicroclimate-07	2025-05-16T01:40:50+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	44.0	3
335397	ICTMicroclimate-03	2025-05-16T01:40:35+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	321.0	3
335398	ICTMicroclimate-08	2025-05-16T01:44:50+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	154.0	1
335399	ICTMicroclimate-04	2025-05-16T01:43:34+10:00	Batman Park	-37.8221828, 144.9562225	0.0	5.0	3
335400	ICTMicroclimate-05	2025-05-16T01:53:41+10:00	Enterprize Park - Pole ID: COM1667	-37.8204083, 144.9591192	0.0	33.0	9
335401	ICTMicroclimate-08	2025-05-16T01:59:51+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	112.0	139.0	2
335402	ICTMicroclimate-06	2025-05-16T01:57:46+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	40.0	3
335403	ICTMicroclimate-02	2025-05-16T02:02:07+10:00	101 Collins St L11 Rooftop	-37.814604, 144.9702991	0.0	4.0	3
335404	ICTMicroclimate-01	2025-05-16T02:05:57+10:00	Birrarung Marr Park - Pole 1131	-37.8185931, 144.9716404		80.0	
335405	aws5-0999	2025-05-16T02:12:07+10:00	Royal Park Asset ID: COM2707	-37.7956167, 144.9519007	0.0	0.0	0
335406	ICTMicroclimate-07	2025-05-16T02:25:56+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	36.0	3
335407	ICTMicroclimate-05	2025-05-16T02:23:44+10:00	Enterprize Park - Pole ID: COM1667	-37.8204083, 144.9591192	0.0	33.0	6
335408	ICTMicroclimate-11	2025-05-16T02:20:02+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	311.0	3
335409	ICTMicroclimate-08	2025-05-16T02:29:55+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	164.0	1
335410	ICTMicroclimate-09	2025-05-16T02:37:55+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	5.0	3
335411	ICTMicroclimate-11	2025-05-16T02:35:03+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	329.0	3
335412	ICTMicroclimate-08	2025-05-16T02:44:57+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	172.0	2
335413	ICTMicroclimate-06	2025-05-16T02:42:52+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	60.0	3
335414	ICTMicroclimate-11	2025-05-16T02:50:04+10:00	1 Treasury Place	-37.812888, 144.9750857	0.0	335.0	3
335415	ICTMicroclimate-09	2025-05-16T02:52:57+10:00	SkyFarm (Jeff's Shed). Rooftop - Melbourne Conference & Exhibition Centre (MCEC)	-37.8223306, 144.9521696	0.0	0.0	3
335416	ICTMicroclimate-06	2025-05-16T02:57:55+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	43.0	3
335417	ICTMicroclimate-07	2025-05-16T02:56:01+10:00	Tram Stop 7C - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8222341, 144.9829409	0.0	31.0	3
335418	ICTMicroclimate-03	2025-05-16T02:55:45+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	343.0	0
335419	ICTMicroclimate-08	2025-05-16T03:00:00+10:00	Swanston St - Tram Stop 13 adjacent Federation Sq & Flinders St Station	-37.8184515, 144.9678474	0.0	126.0	1
335420	ICTMicroclimate-06	2025-05-16T03:12:56+10:00	Tram Stop 7B - Melbourne Tennis Centre Precinct - Rod Laver Arena	-37.8194993, 144.9787211	0.0	67.0	3
335421	ICTMicroclimate-03	2025-05-16T03:10:48+10:00	CH1 rooftop	-37.8140348, 144.96728	0.0	326.0	3
335422							
335423							

Figure 8.1.1. Visualization of the oldest and newest information from the dataset

8.2. Appendix B: Python Code for implementing Machine Learning techniques

8.2.1. Importing the libraries

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
```

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from sklearn.cluster import KMeans
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, LSTM
from tensorflow.keras.callbacks import EarlyStopping
```

8.2.2. Importing the dataset

```
df = pd.read_csv("microclimate-sensors-data.csv")
df.tail()
```

8.2.3. Data preprocessing

```
import pandas as pd
from sklearn.preprocessing import MinMaxScaler
df['Time'] = pd.to_datetime(df['Time'], errors='coerce', utc=True)

# Dropping invalid time rows
df = df.dropna(subset=['Time']).copy()

# Extract time features
df['hour'] = df['Time'].dt.hour
df['day'] = df['Time'].dt.day

# Select features
features = ['AirTemperature', 'RelativeHumidity', 'PM25', 'PM10', 'hour', 'day']
df_model = df[features]
df_model = df_model.fillna(df_model.mean(numeric_only=True))

# Split and scale
X = df_model.drop('AirTemperature', axis=1)
y = df_model['AirTemperature']
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)
```

8.2.4. Preparing the data for the regression models

```
# Select relevant features
features = ['AirTemperature', 'RelativeHumidity', 'PM25', 'PM10', 'hour', 'day']
df_model = df[features]

# Fill missing values with column means
df_model = df_model.fillna(df_model.mean(numeric_only=True))

# Split into features (X) and target (y)
X = df_model.drop('AirTemperature', axis=1)
y = df_model['AirTemperature']
```

```
# Train-test split
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

8.2.5. Applying linear regression model

```
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

# Train the model
lr = LinearRegression()
lr.fit(X_train, y_train)

# Make predictions
y_pred_lr = lr.predict(X_test)

# Evaluate the model
print("Linear Regression Results:")
print("MAE:", mean_absolute_error(y_test, y_pred_lr))
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred_lr)))
print("R²:", r2_score(y_test, y_pred_lr))
```

8.2.6. Visualizing the results from the linear regression model

```
import matplotlib.pyplot as plt
plt.figure(figsize=(10, 5))
plt.scatter(y_test, y_pred_lr, alpha=0.5, label='Predictions')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='red', lw=2, label='Perfect Prediction')
plt.xlabel('Actual Temperature')
plt.ylabel('Predicted Temperature')
plt.title('Linear Regression: Predicted vs Actual Temperature')
plt.legend()
plt.grid(True)
plt.show()
```

8.2.7. Expanded linear regression (other features selected)

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.preprocessing import MinMaxScaler
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np
import matplotlib.pyplot as plt

# Convert and clean datetime
df['Time'] = pd.to_datetime(df['Time'], errors='coerce')
df = df.dropna(subset=['Time'])
```

```
# Extract time features
df['hour'] = df['Time'].dt.hour
df['day'] = df['Time'].dt.day

# Select expanded features
features = [
    'AirTemperature',
    'RelativeHumidity',
    'PM25',
    'PM10',
    'hour',
    'day',
    'AverageWindSpeed',
    'GustWindSpeed',
    'AtmosphericPressure',
    'Noise'
]

# Keep only necessary columns and fill missing values
df_model = df[features].copy()
df_model = df_model.fillna(df_model.mean(numeric_only=True))

# Split data
X = df_model.drop('AirTemperature', axis=1)
y = df_model['AirTemperature']

# Scale features
scaler = MinMaxScaler()
X_scaled = scaler.fit_transform(X)

# Train/test split
X_train, X_test, y_train, y_test = train_test_split(X_scaled, y, test_size=0.3, random_state=42)

# Train linear regression
lr = LinearRegression()
lr.fit(X_train, y_train)
y_pred = lr.predict(X_test)

# Evaluate
print("Linear Regression with More Features:")
print("MAE:", mean_absolute_error(y_test, y_pred))
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred)))
print("R²:", r2_score(y_test, y_pred))
```

8.2.8. Visualizing the expanded linear regression model

```
# Visualize
```

```
plt.figure(figsize=(10, 5))
plt.scatter(y_test, y_pred, alpha=0.5, label='Predictions')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='red', lw=2, label='Perfect Prediction')
plt.xlabel('Actual Temperature')
plt.ylabel('Predicted Temperature')
plt.title('Linear Regression (Expanded Features): Predicted vs Actual')
plt.legend()
plt.grid(True)
plt.show()
```

8.2.9. Applying Support Vector Machine (SVM) model

```
from sklearn.svm import SVR
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

# Initialize SVR model
svr = SVR()
svr.fit(X_train, y_train)

# Predict
y_pred_svr = svr.predict(X_test)

# Evaluate
print("SVR Results:")
print("MAE:", mean_absolute_error(y_test, y_pred_svr))
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred_svr)))
print("R²:", r2_score(y_test, y_pred_svr))
```

8.2.10. Visualizing the SVM model

```
import matplotlib.pyplot as plt

plt.figure(figsize=(10, 5))
plt.scatter(y_test, y_pred_svr, alpha=0.5, color='green', label='Predictions')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='red', lw=2, label='Perfect Prediction')
plt.xlabel('Actual Temperature')
plt.ylabel('Predicted Temperature')
plt.title('SVR: Predicted vs Actual Temperature')
plt.legend()
plt.grid(True)
plt.show()
```

8.2.11. Applying XGBoost model

```
!pip install xgboost
```

```
import xgboost as xgb
from xgboost import XGBRegressor
# Convert and clean datetime
df['Time'] = pd.to_datetime(df['Time'], errors='coerce')
df = df.dropna(subset=['Time'])

# Extract time features
df['hour'] = df['Time'].dt.hour
df['day'] = df['Time'].dt.day

# Feature list
features = [
    'AirTemperature',
    'RelativeHumidity',
    'PM25',
    'PM10',
    'hour',
    'day',
    'AverageWindSpeed',
    'GustWindSpeed',
    'AtmosphericPressure',
    'Noise'
]

# Prepare and clean
df_model = df[features].copy()
df_model = df_model.fillna(df_model.mean(numeric_only=True))
X = df_model.drop('AirTemperature', axis=1)
y = df_model['AirTemperature']

from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
import numpy as np

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
model = XGBRegressor(n_estimators=200, learning_rate=0.1, max_depth=6, random_state=42)
model.fit(X_train, y_train)
y_pred = model.predict(X_test)

print("XGBoost Regressor Results:")
print("MAE:", mean_absolute_error(y_test, y_pred))
print("RMSE:", np.sqrt(mean_squared_error(y_test, y_pred)))
print("R²:", r2_score(y_test, y_pred))
```

8.2.12. Visualizing the XGBoost model

```
import matplotlib.pyplot as plt
```

```
plt.figure(figsize=(10, 5))
plt.scatter(y_test, y_pred, alpha=0.5, color='purple', label='Predictions')
plt.plot([y_test.min(), y_test.max()], [y_test.min(), y_test.max()], color='red', lw=2, label='Perfect Prediction')
plt.xlabel('Actual Temperature')
plt.ylabel('Predicted Temperature')
plt.title('XGBoost: Predicted vs Actual Temperature')
plt.legend()
plt.grid(True)
plt.show()
```

8.2.13. Applying the CNN model and its visualization

```
import numpy as np
import pandas as pd
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout, BatchNormalization
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping

#Pad features to reach 25 (for 5x5 CNN input)
def pad_features(df, target_size=25):
    current_size = df.shape[1]
    padding = target_size - current_size
    if padding > 0:
        padding_array = np.zeros((df.shape[0], padding))
        df_padded = np.hstack([df, padding_array])
        return df_padded
    return df.values

#Pad train and test sets
X_train_padded = pad_features(X_train, 25)
X_test_padded = pad_features(X_test, 25)

#Reshape to 5x5 grid with 1 channel
X_train_cnn = X_train_padded.reshape(-1, 5, 5, 1)
X_test_cnn = X_test_padded.reshape(-1, 5, 5, 1)
import matplotlib.pyplot as plt
model = Sequential([
    Conv2D(32, (3, 3), activation='relu', padding='same', input_shape=(5, 5, 1)),
    BatchNormalization(),
    MaxPooling2D(pool_size=(2, 2)),

    Conv2D(64, (2, 2), activation='relu', padding='same'),
    BatchNormalization(),
    Flatten(),
```

```
    Dense(64, activation='relu'),
    Dropout(0.3),
    Dense(1)
])

#Compile the model
model.compile(optimizer=Adam(learning_rate=0.001), loss='mse')

#Train the model
early_stop = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)

history = model.fit(
    X_train_cnn, y_train,
    validation_split=0.2,
    epochs=100,
    batch_size=32,
    callbacks=[early_stop],
    verbose=1
)

#Predict and evaluate
y_pred = model.predict(X_test_cnn).flatten()

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)

print(f"MAE: {mae:.2f} °C")
print(f"RMSE: {rmse:.2f} °C")
print(f"R² Score: {r2:.4f}")
plt.figure(figsize=(8, 5))
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('CNN Model Loss Over Epochs')
plt.xlabel('Epoch')
plt.ylabel('Mean Squared Error (Loss)')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()
```

8.2.14. Applying the LSTM model and its visualization

```
import numpy as np
import matplotlib.pyplot as plt
```

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Dropout
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import EarlyStopping

def pad_features(df, target_size=25):
    current_size = df.shape[1]
    padding = target_size - current_size
    if padding > 0:
        padding_array = np.zeros((df.shape[0], padding))
        df_padded = np.hstack([df, padding_array])
        return df_padded
    return df
X_train_padded = pad_features(X_train, 25)
X_test_padded = pad_features(X_test, 25)

# We'll treat each row as 1 timestep with 25 features
X_train_lstm = X_train_padded.reshape((X_train_padded.shape[0], 1, X_train_padded.shape[1]))
X_test_lstm = X_test_padded.reshape((X_test_padded.shape[0], 1, X_test_padded.shape[1]))

#Building the LSTM model
model = Sequential([
    LSTM(64, activation='tanh', return_sequences=False, input_shape=(1, X_train_lstm.shape[2])),
    Dropout(0.3),
    Dense(64, activation='relu'),
    Dense(1)
])
model.compile(optimizer=Adam(learning_rate=0.001), loss='mse')

#Early stopping
early_stop = EarlyStopping(monitor='val_loss', patience=10, restore_best_weights=True)

#Train the model
history = model.fit(
    X_train_lstm, y_train,
    validation_split=0.2,
    epochs=100,
    batch_size=32,
    callbacks=[early_stop],
    verbose=1
)

#Predict and evaluate
y_pred = model.predict(X_test_lstm).flatten()
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
```

```
r2 = r2_score(y_test, y_pred)
print(f"MAE: {mae:.2f} °C")
print(f"RMSE: {rmse:.2f} °C")
print(f"R² Score: {r2:.4f}")

#Plot training vs validation loss
plt.figure(figsize=(8, 5))
plt.plot(history.history['loss'], label='Training Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.title('LSTM Model Loss Over Epochs')
plt.xlabel('Epoch')
plt.ylabel('Mean Squared Error (Loss)')
plt.legend()
plt.grid(True)
plt.tight_layout()
plt.show()

#Plot predicted vs actual
plt.figure(figsize=(10, 5))
plt.plot(y_test[:100], label='Actual')
plt.plot(y_pred[:100], label='Predicted')
plt.title('LSTM Predictions vs Actual Temperatures (First 100 Samples)')
plt.xlabel('Sample Index')
plt.ylabel('Temperature (°C)')
plt.legend()
plt.grid(True)
plt.show()
```

9. Glossary

- **Microclimate** – A local atmospheric zone where the climate differs from the surrounding area, often influenced by urban structures, vegetation, or water bodies.
- **Forecasting** – The process of using historical or current data to predict future outcomes, such as temperature or water usage.
- **LSTM (Long Short-Term Memory)** – A type of Recurrent Neural Network (RNN) capable of learning long-term dependencies in sequential data, often used for time-series forecasting.
- **CNN (Convolutional Neural Network)** – A type of deep learning model commonly used for recognizing spatial patterns, adapted here for spatial microclimate modeling.
- **Water Usage** – The amount of water consumed for residential, agricultural, industrial, or environmental purposes.
- **Evapotranspiration** – The sum of water evaporation from land and transpiration from plants; an important factor in climate-influenced water demand.
- **Solar Radiation** – The energy emitted by the sun, affecting temperature, evaporation, and therefore water consumption.
- **Time-Series Data** – A sequence of data points collected or recorded at specific time intervals, used in forecasting future trends.
- **Sensor Data** – Data collected from environmental sensors measuring temperature, humidity, wind, etc., used in model training and prediction.
- **Model Architecture** – The structure of a machine learning model, including layers and operations, that defines how it processes input to produce output.
- **Urban Heat Island** – An urban area that is significantly warmer than its surrounding rural areas due to human activity, impacting local water usage.
- **Data Preprocessing** – The process of cleaning and transforming raw data into a usable format for analysis or modeling.
- **Sustainability** – In this context, managing water resources efficiently to meet present and future demands without harming the environment.
- **Ethical Considerations** – The analysis of moral issues that arise from applying predictive systems, including fairness, bias, and access to resources.

10. References

[1] *Waqas M. & Humphries U.W.* (2024). A critical review of RNN and LSTM variants in hydrological time series predictions. *MethodsX*.

[2] *Gong Y., Zhang Y., Wang F., & Lee C.-H.* (2024). Deep Learning for Weather Forecasting: A CNN-LSTM Hybrid Model for Predicting Historical Temperature Data. *Applied and Computational Engineering, 99*, 168–174.

[3] Rafiei, V., & Samadi, S. (2020). *A comparative study of neural network, decision tree, and regression models for water consumption forecasting*. Journal of Water Resources Management, 34(15), 4679–4696.

[4] Jain, A., Pandey, C. P., & Hote, Y. V. (2020). *Urban water demand forecasting using random forest and multiple linear regression*. Journal of Water Supply: Research and Technology-AQUA, 69(1), 42–56.

[5a] DeMaagd, N. & Roberts, M. J. (2020). *How will climate change affect water demand? Evidence from Hawai'i microclimates*. *Climate Change Economics*, 11(03), 2050014.

[5b] Kontopoulos, I., Makris, A., Tserpes, K., & Varvarigou, T. (2023). *An evaluation of time series forecasting models on water consumption data: A case study of Greece*. *BMC Research Notes*.

[5c] Kitessa, G., et al. (2024). *Urban Water-Energy Consumption Prediction Influenced by Climate and Socio-Economic Dynamics*. *Scientific Reports*.

[6] Khodadadi, K., & Hassanzadeh, Y. (2018). *Implications of data sampling resolution on water use simulation*. *Global Environmental Change*, 50, 193–203.

[7] Adamowski, J., & Karapataki, C. (2010). *Comparison of multivariate regression and artificial neural networks for peak urban water-demand forecasting: Evaluation of different ANN learning algorithms*. *Journal of Hydrologic Engineering*, 15(10), 729–743.

[8] Donkor, E. A., Mazzuchi, T. A., Soyer, R., & Roberson, J. A. (2014). *Urban water demand forecasting: Review of methods and models*. *Journal of Water Resources Planning and Management*, 140(2), 146–159.

[9] Gato, S., Jayasuriya, N., & Roberts, P. (2007). *Forecasting residential water demand: Case study*. *Journal of Water Resources Planning and Management*, 133(4), 309–319.

[10] Herrera, M., Torgo, L., Izquierdo, J., & Pérez-García, R. (2010). *Predictive models for forecasting hourly urban water demand*. *Journal of Hydrology*, 387(1-2), 141–150.

[11] Zhou, S. L., McMahon, T. A., Walton, A., & Lewis, J. (2000). *Forecasting operational demand for an urban water supply zone*. *Journal of Hydrology*, 259(1-4), 189–202.

[12] Ghaffarianhoseini, A., et al. (2020). *Smart forecasting and optimization of urban water consumption: A systematic review*. *Sustainable Cities and Society*, 55, 102029.

[13] Wang, W., et al. (2023). *A hybrid LSTM-XGBoost model for short-term water demand forecasting*. *Water Resources Management*, 37, 813–831.

[14] Brentan, B. M., et al. (2017). *Short-term water demand forecasting using deep learning techniques*. *Procedia Engineering*, 186, 436–443.