Rainfall Forecasting Using Artificial Neural Networks

by

Ali Haidar

A thesis submitted in fulfilment
of the requirements for the degree of
Doctor of Philosophy

School of Engineering and Technology

Central Queensland University, Australia

September 2018

ABSTRACT

Rainfall is a complex meteorological process that affects the environment, human based activities, agriculture, transportation, and almost every aspect of life. The ability to determine the amount of rain that will fall is helpful for different agricultural industries. It enhances decision-making and management of farming processes, including plantation, irrigation, fertilization, and harvest. Different rainfall forecasting methods have been suggested and used to forecast rainfall over various durations. The existing forecasting models release forecasts over large grid areas, and the forecasted output is given to end-users as a probabilistic value which is considered uninformative because it does not provide accurate information about the type of expected rainfall. These models have also shown low prediction accuracy on some occasions.

In this thesis, new approaches are proposed for predicting monthly rainfall. The approaches are based on single artificial neural networks and ensembles of artificial neural networks. The first approach proposes an evolutionary algorithm to select the most significant features for single neural networks. The second approach extends the first approach by including network parameters in the selection process. The third approach introduces a fusion of multiple single neural networks and develops novel ensembles of neural networks for rainfall prediction. Several fusion methods are proposed to combine the single neural networks. The fourth approach uses ensemble components selection while building the rainfall prediction model. Two types of forecasts were targeted in this thesis: numerical and categorical. Numerical prediction is the process of predicting the actual amount of rainfall. In categorical prediction, rainfall is classified into categories, such as above average and below average.

The proposed approaches were applied in order to predict rainfall for areas of eastern Australia. The datasets were created by collecting and processing weather variables from multiple online resources. Several local and global weather attributes were used as possible predictors of rain. These predictors included temperature, solar attribute, and climate indices. A climate index represents a particular phenomenon over a selected area in oceans including Pacific and Indian oceans. Australian rainfall variability is correlated to these local and global weather attributes. Various statistical measurements, including Mean Absolute Error, Root Mean Square Error, Pearson Correlation, Skill Scores, classification error, and f-score, were used to assess the performance of the proposed approaches. The proposed approaches were compared to alternative techniques, and better performance was recorded with proposed neural network based approaches. In addition, the developed models using the proposed approaches were compared with the new forecasting system released by the Bureau of Meteorology, and better performance was obtained with the developed models used in this thesis.

The results and comparative analysis in this thesis show that the proposed neural network based ensemble approach is appropriate for monthly rainfall prediction. It was found that input features and neural network parameters should be carefully chosen when designing a neural network. In addition, correct selection of ensemble components increases the efficiency of the ensemble model. The proposed models can be useful for agricultural industries such as sugarcane, wheat, and cotton. Because various industries record rainfall and weather data, the proposed approaches can be utilized to build forecasting models for these industries, as adequate records of related atmospheric data are available on multiple online resources.

ACKNOWLEDGEMENTS

The research in this thesis could not have been completed without the support of many individuals.

I would like to express my sincere gratitude to my supervisor Prof. Brijesh Verma for his continuous support during my doctoral and related research, and also for his patience, motivation, and depth of knowledge. Without his support, it would not have been possible to conduct this research.

I want to thank Central Queensland University (CQU) for supporting this research, and gratefully acknowledge funding received from the CQU Research Division. Financial support provided by the Australian Government through the Research Training Scheme (RTS) was also very helpful.

I am grateful for the friendly support of the Research office staff and team members in the Centre for Intelligent Systems (CIS).

Special thanks are due to my father, “the first mentor”, my mother, Fatima Tawbi, “the first teacher”, and my brother and sisters for their help. I am also grateful to other family members and friends who have supported me along the way.

Finally, I would also like to express my gratitude for my wife Amani for her support and motivation since the first day of this research. For Vera and Rihanna, it is a joy to give both of you my special thanks at the end of this section. Words cannot express how grateful I am. Having you beside me throughout this research was a great support.

ACKNOWLEDGEMENT OF PROFESSIONAL SERVICES

Professional editor, Dr Scott Charlesworth, provided copyediting and proof-reading services, according to the guidelines laid out in the University-endorsed national 'Guidelines for Editing Research Theses'.

………………………………. ……………………………………………

(Original signature of Candidate) Date

DECLARATION OF AUTHORSHIP AND ORIGINALITY

I, the undersigned author, declare that all of the research and discussion presented in this thesis is original work performed by the author. None of the content of this thesis has been submitted or considered, either in whole or in part, at any tertiary institute or university for a degree or any other category of award. I also declare that any material presented in this thesis as performed by another person or institute has been referenced and listed in the reference section.

…………………………………….. …………………………………..

(Original signature of Candidate) Date

COPYRIGHT STATEMENT

………………………………. ……………………………….

(Original signature of Candidate) Date

LIST OF PUBLICATIONS

Journals

- A. Haidar and B. Verma, “A novel approach for optimizing climate features and network parameters in rainfall forecasting,” *Soft Computing*, pp. 1-12, 2017.
- A. Haidar and B. Verma, “Monthly rainfall forecasting using neural networks for sugarcane regions in Eastern Australia,” *Water Science and Technology: Water Supply*, *vol. 17*, pp. 907-920, 2016.

Book Chapters

- A. Haidar and B. Verma, "Monthly rainfall categorization based on optimized features and neural network," in *AI 2017: Advances in Artificial Intelligence* (Lecture Notes in Computer Science, no. 10400), W. Peng, D. Alahakoon, and X. Li, Eds. Cham, Springer, 2017, pp. 208-220.

Conferences

- A. Haidar, B. Verma, and T. Sinha, "A novel approach for optimizing ensemble components in rainfall prediction," *IEEE Congress on Evolutionary Computation (CEC)*, 2018, pp. 1817-1824.
- A. Haidar and B. Verma, "A hybrid genetic algorithm for climate input features and neural network parameters selection," *Annual Conference on Genetic and Evolutionary Computation*, 2017, pp. 281-282.
- A. Haidar and B. Verma, "Learning based fusion in ensembles for weather forecasting," *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2017, pp. 72-78.
- A. Haidar and B. Verma, "A genetic algorithm based feature selection approach for rainfall forecasting in sugarcane areas," *IEEE Symposium Series on Computational Intelligence* (SSCI), 2016, pp. 1-8.

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "A novel approach for optimizing climate features and network parameters in rainfall forecasting," Soft Computing, pp. 1-12, 2017.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-author provided guidance with concepts, experiments and analysis of the results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

………………………………………….. ……………………………………

(Original signature of Candidate) Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "Monthly rainfall forecasting using neural networks for sugarcane regions in Eastern Australia," Water Science and Technology: Water Supply, vol. 17, pp. 907-920, 2016.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 70%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-author provided guidance for analysing the results and preparing the paper. He also assisted by editing the paper. Contribution: 30%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

…………………………………………..   …………………………………

(Original signature of Candidate)   Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "Monthly rainfall categorization based on optimized features and neural network," in AI 2017: Advances in Artificial Intelligence (Lecture Notes in Computer Science, no. 10400), W. Peng, D. Alahakoon, and X. Li, Eds. Cham, Springer, 2017, pp. 208-220.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-author provided guidance with methodology/concepts, experiments and analysis of the results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

…………………………………………..   …………………………………

(Original signature of Candidate)   Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar, B. Verma, and T. Sinha, "A novel approach for optimizing ensemble components in rainfall prediction," IEEE Congress on Evolutionary Computation (CEC), 2018, pp. 1817-1824.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-authors provided guidance with methodology/concepts, experiments and analysis of results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

This paper has been listed in Toshi Sinha's thesis as she is third co-author in this paper but research in this paper was not claimed by her as her thesis work.

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

………………………………………….. ……………………………………

(Original signature of Candidate) Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "A hybrid genetic algorithm for climate input features and neural network parameters selection," Annual Conference on Genetic and Evolutionary Computation, 2017, pp. 281-282.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-author provided guidance with methodology/concepts, experiments and analysis of results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

………………………………………….. ………………………………….

(Original signature of Candidate) Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "Learning based fusion in ensembles for weather forecasting," 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), 2017, pp. 72-78.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-author provided guidance with methodology/concepts, experiments and analysis of results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

…………………………………………..　　　　…………………………………

(Original signature of Candidate)　　　　Date

DECLARATION OF CO-AUTHORSHIP AND CONTRIBUTION

Full bibliographic reference	A. Haidar and B. Verma, "A genetic algorithm based feature selection approach for rainfall forecasting in sugarcane areas," IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1-8.
Status	Published

Nature of Candidate's Contribution, including percentage of total

I implemented the methodology, conducted the experiments, analysed the results and written the paper. Contribution: 60%.

Nature of all Co-Authors' Contributions, including percentage of total

My co-authors provided guidance with methodology/concepts, experiments and analysis of results and rewritten/corrected the paper. Contribution: 40%.

Has this paper been submitted for an award by another research degree candidate (Co-Author), either at CQUniversity or elsewhere? (if yes, give full details)

No

Candidate's Declaration

I declare that the publication above meets the requirements to be included in the thesis as outlined in the Research Higher Degree Theses Policy and Procedure

………………………………………….. …………………………………

(Original signature of Candidate) Date

CONTENTS

LIST OF ABBREVIATIONS

Abbreviation	Definition
ACCESS	Australian Community Climate and Earth-System Simulator-Seasonal Prediction System
AFM	Average Fusion Model
ANFIS	Adaptive Network-Based Fuzzy Inference Systems
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
ARIMA	Autoregressive Integrated Moving Average
BP	Back Propagation
BPNN	Back Propagation Neural Network
BC	Before Christ
BOM	Bureau of Meteorology
CE	Correlation of Efficiency
CFPS-GA	Climate Features and Parameters Selection based Genetic Algorithm
CFPS-HGA	Climate Features and Neural Network Parameters Selection based Hybrid Genetic Algorithm
CFPS-TLO	Climate Features and Network Parameters Selection – Two Levels Optimization
CFS-GA	Climate Features Selection based Genetic Algorithm
Df	Degrees of freedom
DMI	Dipole Mode Index
DT	Decision Tree
EA	Evolutionary Algorithms
ECS_HGA	Ensemble Components Selection using a Hybrid Genetic Algorithm
ECS_PSO	Ensemble Components Selection using Particle Swarm Optimization
ECS_SGA	Ensemble Components Selection using a Standard Genetic Algorithm
EDI	Effective Drought Index
EI	Efficiency Index
ENSO	El-Nino Southern Oscillation
ERNN	Elman Recurrent Neural Network
ESRL	Earth System Research Laboratory

FFNN	Feed Forward Neural Network
FRNN	Fully Recurrent Neural Network
FTDNN	Focused Time Delay Neural Network
GA	Genetic Algorithm
GA-NN	Genetic Algorithm Neural Network
GASA-NN	Genetic Algorithm Simulated Annealing Neural Network
GCM	General Circulation Model
GFFNN	Generalized Feed Forward Neural Network
GUI	Graphical User Interface
GRNN	Generalized Regression Neural Network
HFM	Hopfield Model
IA	Index of Agreement
ICA	Independent Component Analysis
IITM	Indian Institute of Tropical Meteorology
IOD	Indian Ocean Dipole
IPO	Inter-decadal Pacific Oscillation
IPE	Ideal Point Error
ISMR	Indian Summer Monsoon Rainfall
ISMRI	Indian Summer Monsoon Rainfall Index
JMA	Japan Meteorological Agency
KNN	K-Nearest Neighbour
KNMI	Koninklijk Nederlands Meteorologisch Instituut (Royal Dutch Meteorological Institute)
LARS	Least Angle Regression
LENN-FM	Lowest Error-Neural Network based Fusion Model
LSM	Least Square Method
MAE	Mean Absolute Error
MAD	Mean Absolute Deviation
MAPE	Mean Absolute Percentage Error
MI	Mutual Information
MLFFNN	Multi-Layered Feed Forward Neural Network
MLP	Multi-Layered Perceptron

MLR	Multiple Linear Regression
MMT	Mean Max Temperature
MS	Mean Square
MSE	Mean Square Error
MVP	Mean Vapor Pressure
MRA	Multi-Regression Analysis
MRH	Mean Relative Humidity
MSLP	Mean Station Level Pressure
MSP	Mean Sea Level Pressure
MWBP	Mean Wet Bulb Pressure
NAO	North Atlantic Oscillation
NCI	National Computational Infrastructure
NFN	Neuro-Fuzzy Neuron
NMSE	Normalized Mean Square Error
NN-FM	Neural Network based Fusion Model
NNPSO-FM	Neural Network based PSO Fusion Model
NOAA	National Oceanic & Atmospheric Administration
NPI	North Pacific Index
NRMSE	Normalized Root Mean Square Error
PCA	Principle Component Analysis
PDO	Pacific Decadal Oscillation
PE	Prediction Error
PI	Persistency Index
POAMA	Predictive Ocean Atmosphere Model for Australia
PSO	Particle Swarm Optimization
RBFNN	Radial Basis Function Neural Network
RE	Relative Error
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
RO	Random Optimization
RSM	Regional Spectral Model
SAM	Southern Annular Mode

SARFIMA	Seasonal Auto Regressive Fractionally Integrated Moving Average
SD	Standard Deviation
SIDC	Solar Influences Data Analysis Center
SNN	Single Neural Network
SOI	Southern Oscillation Index
SPEI	Standardized Precipitation and Evapotranspiration Index
SS	Skill Score
SS	Sum of Squares
SST	Sea Surface Temperature
SSTA	Sea Surface Temperature Average
STR	Sub-tropical Ridge
SVM	Support Vector Machine
TDNN	Time Delay Neural Network
TLBO	Teaching Learning Based Optimisation
TPI	Tripole Index for Inter-Decadal Pacific Oscillation
TS-LARS	Time Series Least Angle Regression
UK	United Kingdom
WT	Wavelet Transform
WTA	Winner Takes All

LIST OF FIGURES

LIST OF TABLES

Chapter 1 Introduction

1.1 Background

Weather forecasting is a field in which the state of an atmospheric variable is determined over a specific location and for a selected duration. Different types of prediction models have been established and tested for forecasting weather variables such as temperature, humidity, and rainfall. Rainfall is a vital natural phenomenon that contributes to the maintenance of ecological balance all over the world. It is a result of complex meteorological interactions, and is a part of the vaporization process. Rainfall forecasting, a major type of weather forecasting, works by determining the state of rainfall over a specific area at a specific time. Rainfall forecasting is essential for agriculture, transportation, tourism, construction, and life itself. Therefore, predictions should be reasonably accurate, have economic impact, and be well communicated to users in order to be considered valuable [1].

Weather forecasting ensures the sustainable development of society and the economy. Therefore, there has been interest in forecasting since 650 BC when the Babylonians tried to predict weather based on observations of clouds (observed patterns) [2]. At that time, multiple philosophers proposed various forecasting theories. Over time, it was noticed that these theories were not adequate. Consequently, it was perceived that there was a need to understand the weather from a broader perspective. With the invention of new instruments, measurement of the atmosphere was undertaken. Various instruments, such as the telegraph and radiosonde, allowed better monitoring of weather conditions [3]. Nowadays, these instruments are used to record weather conditions. For modern rainfall forecasting, weather forecasts were produced before the invention of the computer, when Lewis Fry Richardson used arithmetic equations to predict weather after World War I (1922) [4, 5]. Consequently, scientists introduced new methods that were developed along with the vast spread of technology. Scientists now use different methods to apply forecasts. Some models require supercomputing facilities to release forecasts. Because of its relevance to human life and needs, weather forecasting is applied everywhere in the world.

Rainfall predictions can be beneficial for multiple disciplines in terms of decision-making, planning, and risk management. Predictions can be utilized to maximize profits and minimize losses. Agriculture receives the highest benefit compared to other sectors [6-8]. Rainfall forecasts enhance decision-making and management in electrical demand and tourism. Likewise, forecasting can be beneficial for the mining industry and construction. Wet conditions may lead to alterations in work schedules, so prior warning helps stakeholders and managers since work is done sequentially.

Climate variability affects different agricultural sectors in Australia, including wheat, cotton, beef, and sugar. The agricultural sector is impacted by short term weather and seasonal and annual climate variations. During a typical season, multiple agricultural processes are usually followed: plantation, irrigation, fertilization, and harvest. In addition, various operations are required to achieve a profitable season, such as transportation, maintenance, etc. Importantly, certain decisions should be made earlier rather than later in order to ensure that a season is profitable, and these decisions are affected by the rainfall amounts that are encountered through parts of the season. Hence, multiple forecasting methods are considered essential and helpful for the agricultural industry in general.

Outlook is a crucial consideration for industry decision-makers. The ability to determine the amount of precipitation in various areas of Australia can enhance the profitability of a season. Delivering accurate seasonal and annual rainfall forecasts in agricultural processes, such as plantation, fertilization, irrigation, and harvest, leads to larger yields and, therefore, higher profit. The ability to predict the time and the amount of precipitation allows growers to set schedules for planting, in addition to avoiding seed damage because of wet or dry weather after plantation. Various types of plants require water to grow. Usually, based on the seasonal amount of precipitation, growers decide whether to include irrigation. Precise rainfall prediction gives growers the opportunity to effectively manage irrigation, so that they can buy water, identify the most suitable machines, and sign contracts for maintenance, if forecasts reveal that the amount of rainfall throughout the season will be insufficient. Therefore, growers could avoid additional fees for higher water prices during the season, and schedule irrigation companies ahead of time so as not to be overwhelmed. On the other side, if forecasts predicted a wet season ahead, growers could avoid signing contracts with irrigation companies, and paying extra amounts for season insurance. This can be beneficial for irrigated and dryland agriculture. Fertilization is the process of adding chemicals and pesticides to fields, in addition to spraying fields at different stages during the season. Growers must be careful when adding fertilizers to fields. If there is heavy rain after adding these chemicals to fields, the chemicals would be washed in to rivers and cause pollution. This may have disastrous effects on river plants. Spraying time is crucial for growers. If there is heavy rainfall and wind, spraying may not lead to satisfactory consequences. Moreover, if there is a dry season, spraying may destroy the plants if there is no irrigation.

The ability to identify the type and amount of rainfall during the harvesting season contributes to managing this process. The success of the Australia harvest season depends on rainfall and the ability to forecast it. Furthermore, harvest is considered to be the most critical of the other agricultural processes. Growers could make a decision to start early if forecasts indicated a wet season, or to start first with paddocks prone to flooding. In addition, rainfall forecasts could encourage growers to sign maintenance contracts and set up a transport chain earlier, with a higher number of containers to move crops. On the other hand, if forecasts predict a dry season, growers could start harvesting the driest areas first so as to avoid loss of crops at the end of the season.

In addition to seasonal and annual forecasts, long decadal rainfall prediction enhances decision-making related to investment in irrigation infrastructure. Managers would benefit from data in improving long-term management of bagasse supplies. Furthermore, marketers could utilize this data to help make choices related to buying expensive items such as storage infrastructure. In addition, the stocking rates of an agricultural industry can be determined early with reasonable forecasts.

1.2 Problems and Motivation

As mentioned, rainfall forecasting contributes to water management and decision-making in different domains. Forecasts influence agricultural processes, industrial production, and transport management. Inaccurate rainfall forecasts have caused disastrous situations in multiple places both in Australia and around the world.

Precipitation is vital to water management. Climate forecasts provide warnings about natural disasters that are brought on sudden change in climatic conditions. In 2010, disastrous floods occurred in Brisbane, the capital city of Queensland. These floods were “dam-release” floods. Release of water from Wivenhoe dam (located on the Brisbane

River) was the principle cause of flooding along the mainstream and tributaries of the Brisbane River downstream of the dam over the period 11th-12th January 2011 [9]. Taking into consideration the effect of rainfall on water level in the Brisbane catchment, a different strategy could have been taken [9].

Rainfall forecasts help in managing the sugarcane industry [10]. Inaccurate outlooks can lead to disastrous effects. Wet conditions in the 1998 harvest season diminished industry wages by around 175 million dollars [11]. The loss was due to cane being left unharvested, reduced commercial cane sugar levels, and damage to paddocks from wet weather harvesting.

The current official rainfall forecasts running in Australia are released by the Bureau of Meteorology (BOM) and are based on General Circulation Models (GCMs). The Predictive Ocean Atmosphere Model for Australia (POAMA) is the current model that is used to release weekly, seasonal, and inter-annual rainfall outlooks. The POAMA is applied to forecast different variables, including temperature and rainfall [12, 13]. The POAMA does not release the actual amount of rainfall to fall on a specific location. Instead, it gives a probability value that represents the chance of exceeding a specific threshold, usually average. The POAMA prediction model consists of 33 ensemble members [14]. For seasonal rainfall predictions, forecasts are usually released for multiple months as a probability of exceeding a certain value (usually average), or as a chance of rainfall. Furthermore, POAMA forecasts are given to users over large spatial distributions ($\approx$ 250 km grids). These forecasts have shown low accuracy on some occasions, and supercomputing facilities are required to release forecasts.

Therefore, there is a need to generate new types of forecasts that would be location specific, with more information and details given for users about the type, amount, and timing of the expected rainfall patterns. Attempts have been made to establish new forecasting systems. While conducting this research, the BOM revealed a new forecasting model for the Australian continent: the Australian Community Climate and Earth-System Simulator-Seasonal Prediction System (ACCESS). It was said that ACCESS would replace POAMA in 2018 [15]. The first version was released in early 2018. It was stated that ACCESS is better at temperature prediction than rainfall [16]. The grid size in ACCESS was specified as 60 km.

In order to avoid disastrous situations and to provide more rainfall information and details for users, new studies have been launched in an attempt to find alternatives for the current forecasting models. Machine learning algorithms have been widely used in various applications. These algorithms are divided into multiple types: Artificial Neural Networks (ANNs), Support Vector Machines (SVMs), Decision Trees (DTs), K- Nearest Neighbours (KNNs), etc. ANNs are machine learning algorithms representing a computational technology built on the analogy of the human information processing system [17]. These algorithms have been successful in various classification and regression tasks, such as financial applications [18, 19], speech recognition, machine vision [20-22], engineering applications [23-26], energy demand [27], electric load consumption [28], bus traffic [29], agricultural applications [30-32], and medical applications [33-36].

Due to its applicability in mapping nonlinear relationships between data, ANNs have been incorporated in several weather prediction tasks including rainfall, temperature, climate indices, etc [37-39]. These models were utilized to estimate variables with different lead times. However, these models are usually built using the trial and error method or the grid search method [40-43]. The trial and error method selects the model characteristics based on the researcher preferences. Hereafter, the best ANN model is not always guaranteed. The grid search is considered expensive and time consuming, where all the available characteristics are investigated.

This study investigates the potential of ANNs and ensembles of ANNs to develop a new and more effective rainfall forecasting technique for Australian localities and a possible solution for climate variability. In addition, it proposes and investigates several methods for selecting the best model characteristics.

1.3 Research Questions

In this research, the ability of ANNs to predict rainfall is to be investigated. Particular interest is taken in the sugar industry in Queensland, Australia. The main research questions are as follows.

1. What climate indices can be used to predict rainfall, and what is the effect of these indices on forecasts which use an ANN model?

2. What is the best method for selecting possible predictors and ANN parameters?

3. What is the effect of using ensembles of ANNs on model performance?
4. What is the best method for selecting ensemble components to obtain highest accuracy?
5. Are ANN forecasts better than existing forecasting models? Could ANNs represent a possible alternative for existing complex forecasting models?

The aim of this research is to investigate the ability of ANNs to provide accurate rainfall forecasts. In this research, the aim is to search for the ANN models that best enhance rainfall forecasting, especially over the sugarcane areas of Queensland. The research is based on deploying an existing ANN for rainfall forecasting, and research on both climate indices and ANN models.

First, the focus is on finding the best ANN models that can be used to deliver accurate rainfall forecasts. We try to find a model that minimises the error between the observed and predicted values. Various methods are proposed to design ANNs. Individual ANNs and other models, which are diverse, are combined, and ensembles are developed and evaluated. Second, climate attributes are highlighted. Scientists have suggested different climate attributes as the main contributors in the formation of rain. We use these attributes as input, and then search for those which increase the accuracy of ANN models. Based on the results, models for forecasting rainfall for different lead times (seasonal) are formed.

1.4 Original Research Contributions

The original research contributions of this thesis are as follow.

- A review of existing techniques used in forecasting weather variables, including rainfall and designing ANNs for rainfall forecasting
- A collection and review of multiple weather indices for building the datasets
- A genetic algorithm-based feature selection approach for rainfall forecasting
- An approach for optimizing climate features and network parameters in rainfall forecasting
- A mutli-level optimisation approach for selecting ANN characteristics in rainfall forecasting
- An ensemble of ANNs approach for rainfall forecasting

- An ensemble components selection approach using a hybrid genetic algorithm for rainfall forecasting

1.5 Thesis Structure

Chapter 1 provides an overview about rainfall forecasting and its impact on various industries.

Chapter 2 contains a brief description of the core structure of ANNs and ensemble classifiers. In addition, related literatures are reviewed. Multiple applications that have been proposed to forecast weather attributes are studied and analysed. The input features and algorithms used to build the forecasting models are examined. Then the methods followed in building these models are reviewed and summarized.

Chapter 3 describes the weather variables used in conducting this research. Then, the processes followed to generate the datasets and case studies are detailed. Evaluation metrics are then defined.

Chapter 4 presents the first types of methods used in building the weather forecasting models. Two methods for selecting input features are presented and applied over two rainfall case studies.

Chapter 5 extends the work done in Chapter 4 to take the advantage of selecting network elements. It details two new ideas used for designing single ANNs. A hybrid evolutionary algorithm (EA) designed to enhance the searching criteria for best network models is also proposed.

Chapter 6 presents a new direction in building weather forecasting models based on ensembles. Various fusion approaches are proposed and evaluated.

Chapter 7 presents a new method for selecting ensemble components using EAs.

Chapter 8 summarizes the contributions made in this thesis and provides a base for extending the work done here in future research.

Chapter 2 Literature Review

This chapter gives a brief overview about ANNs, the main prediction model developed in this research. Then various forecasting methods that have been deployed to predict rainfall and other weather attributes around the world are reviewed. Various types of weather variables were targeted, including rainfall, temperature, humidity, solar, and wind. [41, 42, 44-46]. Two weather attributes are considered the most important for agriculture: temperature and precipitation. ANNs that have been deployed to predict various events, such as severe weather conditions, thunderstorms, rainfall runoff, and drought, are also reviewed [45, 47, 48]. These studies were collected from various countries, including Australia, India, Iran, China, United States of America, Canada, Indonesia, Ethiopia, Sri Lanka, and Malaysia. These methods are partitioned into three categories:

- Weather forecasting models using single machine learning based models;
- Weather forecasting models using ensembles of machine learning models; and
- Various forecasting models.

The methods followed to build the prediction models are reviewed and summarized. The utilization of single ANNs in forecasting rainfall and other weather attributes is examined. Next, forecasting weather variables using ensembles of ANNs is discussed and analysed. Finally, forecasting using other mechanisms is studied.

2.1 Overview of Artificial Neural Networks

ANNs are defined as “massively parallel distributed processing systems representing a new computational technology built on the analogy of the human information processing system” [17, 49]. They consist of multiple computational elements called neurons that process information by their dynamic state response to external inputs [50, 51]. These neurons imitate biological neurons [52]. Each neuron receives an input, performs calculations and returns an output [53]. Usually, biological neurons are distributed over multiple layers, where each set is responsible for certain sub-tasks of the original task.

Similarly, artificial neurons are distributed over multiple layers, each responsible for performing certain functionalities.

The basic structure of an ANN is formed of an input layer, hidden layer, and output layer, as shown in Figure 2.1. The first layer is the layer that receives input data. The second layer is the hidden layer that performs the computations and attempts to find relationships between data. Different number of hidden layers can be added to the same model. The third layer is the output layer that releases the forecasts. Each layer is comprised of a group of neurons. Neurons are linked to the next layer, where each link has a value (weight) that determines the relationship between connected neurons. The output of a neuron is multiplied by this weight while being transferred to the neuron of the next layer. Neurons not belonging to the input layer may receive multiple inputs, which are calculated based on an activation function in order to produce output.

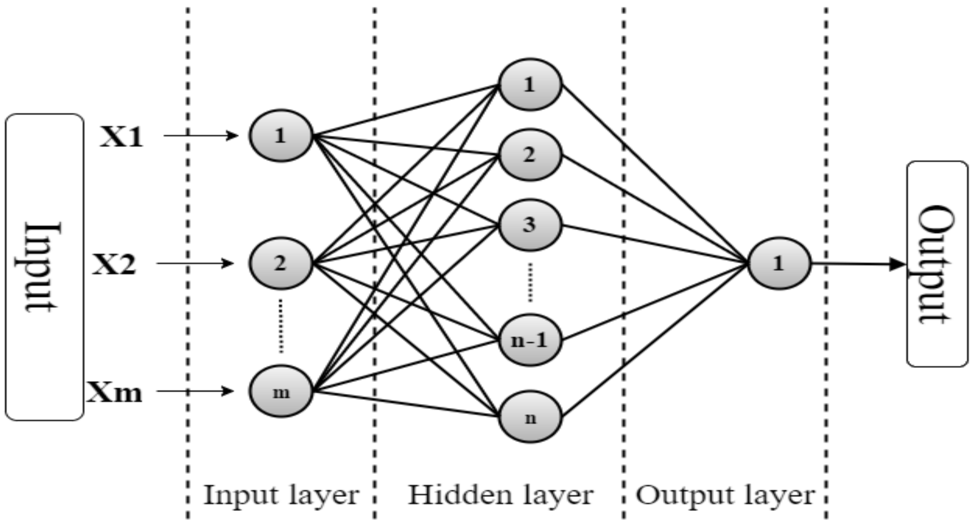

Figure 2.1 Three layered feed forward neural network.

The dataset is usually subdivided into three different sets: training, validation, and testing sets. The ANN is trained using the training dataset [49]. During this process, inputs are given to the model, outputs are calculated, and weights are modified to desired values [54]. The training data are used to train the network and minimize the error by consecutively updating the network weights. Usually, the weight of the interconnected nodes remains the same after finishing the training process [54]. One of the main problems that was noticed while training the ANN is over-fitting. Over-fitting occurs when the network starts memorizing trained data instead of learning to generalize from the trained data. This problem may occur because of the large number of hidden nodes in the network.

Conversely, if the number of hidden nodes is low, the network may not hold the ability to generalize [55]. The second set is the validation set, which is used to validate the reliability of the trained network, and to avoid over-fitting and being trapped in local minimum. The third set is called the test set, and is used to measure the accuracy of the developed model. It is basically a hind cast of data. If the performance during the testing is adequate, then the model can be used in predicting future data [17]. The testing dataset can be selected randomly or based on blocks. In a rainfall related problem, the last section of the dataset is typically used for measuring accuracy.

ANNs have different types of models and training algorithm. Each has different specifications and follows certain criteria through the training, validation, and testing phases. To design an ANN model, a decision about specific attributes/parameters should be taken. The generalization ability of ANNs varies based on multiple things, including network architecture, number of neurons, training algorithm, activation functions between layers, and initial weights. To design an ANN model, a decision about specific attributes/parameters has to be taken. These parameters are as follows.

- Number of input features
- Number of neurons
- Number of layers/hidden layers
- Activation functions
- Learning algorithm
- Connections between layers and neurons
- Learning rate

Based on these parameters, an ANN model is formed. An ANN has a great ability to learn by adjusting these parameters [56] .

Feed Forward Neural Networks (FFNNs) are regularly used in prediction applications [57]. Neurons between layers are connected in a feed-forward manner. Layers are connected using connection weights. The process of learning in a FFNN is achieved by consecutively updating connection weights so as to minimize error (back propagation).

2.2 Single Machine Learning based Models in Weather Prediction

Machine learning algorithms have been used around the world to forecast rainfall and weather attributes [37, 41, 58]. Luk, Ball, and Sharma conducted an experiment to predict rainfall 15 minutes ahead for the Parramatta river catchment, Sydney, Australia [59]. Three ANN models were generated: the Time Delay Neural Network (TDNN), the Multi-Layered Feed Forward Neural Network (MLFFNN), and the ELMAN recurrent ANN. A trial and error method was followed in order to select the best models. Rainfall amounts during 15 minute intervals were collected from 16 gauges between January 1991 and September 1996. The collected data contained 34 storm events. Normalized Mean Square Error (NMSE) was used to assess each model's performance. A NMSE equal to one corresponds to simply predicting the average. Various models were tested, and the eight best networks determined. The selection of these models was based on the minimum NMSE on validation phases. Highest accuracy was obtained with a TDNN having an architecture of 32-16-16. The overall accuracy of all the models varied between 0.63 and 0.67. Authors reported that the developed models were not able to predict peak rainfall rates during the storm event.

Chaudhuri and Chattopadhyay designed an ANN to predict two weather variables: maximum surface temperature, and maximum relative humidity. These two features assist in predicting thunderstorms in India 24 hours in advance [60]. Two models were proposed: single layer network, and a three layered. The activation functions were selected manually. The data consisted of four years for three months between 1994 and 1998. Lagged values of each attribute were used as an input feature. Authors did not use previous values from other months because of the change in co-relational patterns. To measure the accuracy, a prediction error (PE) was used. The smaller the value recorded by the PE, the better the forecasts. The performance of the hidden layer network was reported to be better than the single layer in terms of percentage errors in predictions.

Baawain, Nour, El-Din, and El-Din proposed an ANN based approach to forecast two El-Niño Southern Oscillation (ENSO) indicators: Southern Oscillation Index (SOI), and Nino 3.0 [50]. Two multi-layered perceptrons were designed to predict each indicator. Various types of attributes were used as inputs to predict the SOI and Nino 3.0 up to one year in advance. The network parameters were optimized using a systematic approach, in which the simplest network that would converge was searched. Then network parameters were tuned to lower the error. Multiple attempts were applied for each variable in each lead time (one to twelve months). Various numbers of neurons and activation functions were selected for each lead time. Accuracies were measured in terms of Pearson correlation (r). It was noticed that accuracy decreased when increasing lead times. Correlations of 0.8 and 0.9 were obtained with one-month lead time, while 0.7 and 0.8 r values were obtained with twelve-months lead time.

Chattopadhyay developed an ANN to forecast average rainfall during the summer monsoon in India [61]. A Multi-Layered Feed Forward Neural Network (MLFFNN) using a back propagation (BP) learning algorithm was deployed. The sigmoid function was used as the activation function. The target of the study was the average Indian summer monsoon rainfall. A MLFFNN (2 hidden layers) was developed so that the rainfall data of months of a specific year could be used to predict the average monsoon rainfall for the next year. Data were selected between 1950 and 1995, and were normalized before being added into the network. Prediction Error (PE) was calculated to measure the accuracy of the proposed model and comparison models. The proposed model was compared to a Persistence Forecasting (PF) model and a Multiple Linear Regression (MLR) on the test data where ANN recorded the highest accuracy at 0.15, while PF model recorded 0.17, and the lowest accuracy was obtained with the MLR model at 0.38.

Ayalew, Iler, and Reik designed an ANN to predict flooding of the Omo river in southern Ethiopia [62]. Multiple ANNs were designed to forecast for different lead times. Different combinations of network parameters and input features were assigned and tested. A trial and error method was followed to select the number of neurons. The best model was saved for each lead time (1 hr, 2 hrs, … 6 hrs). Lagged values of hourly run-off were used as input for the network models. Six storm events were used in the study. Four events were applied for training and two events for testing. Accuracy varied between 40.97 mm and 106.09 mm in terms of Root Mean Square Error (RMSE). It was noticed that accuracy decreased drastically with 4 hr, 5 hr, and 6 hr lead times.

Nasseri, Asghari, and Ebedini integrated a Genetic Algorithm (GA) into an ANN to predict rainfall for the Parramatta catchment [63]. Seven scenarios were developed to analyse the proposed approach: five with discrete values, and two with cumulative rainfall values. Sensitivity analyses were applied to select the best input features for the targeted weather station. GA was incorporated to determine the weights in the network architecture. Lagged values from 14 weather stations were used as input features. Better accuracy was found with models that use cumulative rainfall as target data. The authors claimed that this does not mean that the closest stations to the targeted station will be the ones with the highest effect on rainfall prediction at the selected station.

Karmakar, Kower, and Guharhakurta conducted a study to recognize and predict monsoon rainfall over a region in India [64]. Six non-rainfall attributes were used in the study: Mean Vapor Pressure (MVP), Mean Station Level Pressure (MSLP), Mean Relative Humidity (MRH), Mean Sea Level Pressure (MSP), Mean Wet Bulb Pressure (MWBP), and Mean Max Temperature (MMT). Two models were proposed: deterministic and probabilistic. Two types of studies were deployed; the first was to predict each sub-region (eight), and the second was to predict the region as a whole. For deterministic forecasts, a three-layered feed forward was used, while no indication about the structure of the network was displayed in the probabilistic forecasts. To measure accuracy, Mean Absolute Deviation (MAD), Standard Deviation (SD), and Pearson correlation (r) were used. In both sub-regions and the region, higher accuracy was reported with the probabilistic model. For the testing set, records showed 8.3 for the MAD and a correlation of 0.82 with the probabilistic model, while 9.9 was reported for MAD and 0.78 for r with the deterministic model.

Hung, Babel, Weesakul, and Tripathi proposed an approach that used a Generalized Feed Forward Neural Network (GFFNN) to predict short term rainfall up to six hours ahead in Bangkok, Thailand [65]. Data were collected for a duration of five years from 75 rainfall gauges. Multiple models were proposed, with one selected for forecasting. Seven models were developed as a part of the preliminary test. The models included two Multi-Layered Perceptron (MLP) models and five GFFNNs that varied in terms of input features, activation functions, and architecture. Input features consisted of rainfall data and other weather parameters: air pressure, wet temperature, humidity, and cloudiness. RMSE, Efficiency Index (EI), and Correlation Coefficient (r) were used to measure each model performance. The authors reported that using only lagged rainfall values resulted in poor accuracy. Analyses were applied by removing input features in order to understand the effectiveness of each input. The authors tried multiple combinations manually in order to figure out the most significant feature, which was found to be wet bulb temperature, followed by humidity. They then used the best model to create ANNs to make predictions for the 75 rain gauges. The mean RMSE at the 75 locations was reported as 0.87, 1.36, 1.72, 1.85, 1.88, and 1.93 for 1 hr, 2 hr, 3 hr, 4 hr, 5 hr, and 6 hr lead times respectively. Results showed that the larger the lead time the lower the prediction accuracy.

Sedki, Ouazar, and El Mazoudi proposed a method to predict rainfall run-off using an ANN and a GA [66]. The authors used a real coded GA to find the initial weights, then trained the network using BP algorithm. A trial and error method was used to select input features' vectors and number of neurons. The input vector consisted of rainfall and run-off values from the antecedent four days. The network architecture consisted of five neurons in the hidden layer (4-5-1). Tangent function was used to transfer the values between the input and hidden layers. The linear activation function was selected between the hidden and output layers. Three years of data (2000-2002) were used to train the network, and one year was selected for validation (2003). The proposed method was compared to a BP neural network in which better accuracy in terms of RMSE (0.162, 0.199) and R^2 (0.91,0.87) was recorded.

Wang and Sheng developed a model to forecast annual rainfall for Zhengzhou, China [67]. The proposed topology was selected as a Generalized Regression Neural Network (GRNN). Collected data consisted of 55 years between 1955 and 2009, and was divided into the following ranges: 70% for training, 15% for validation and 15% for testing. Data were then normalized before being added into the network. MATLAB was used to conduct experiments. Results over the testing years were compared to a Back Propagation Neural Network (BPNN) and a stepwise regression model. A Relative Error (RE) statistical measure was used to measure the accuracy of the models: 5.33% for GRNN, 30.99% for BPNN and 68.80% for step wise regression.

Castro et al. claimed that ANNs have proven that they have great potential in rainfall forecasting [68]. The authors developed an ANN based model to forecast rainfall over eight regions in Ceara, Brazil. The proposed model aimed to generate seasonal rainfall forecasts. The authors tried to model forecasts over specific seasons when rain occurs (4 months), but not over all of the months. A Neuro-Fuzzy Neuron (NFN) model that integrates the advantages of fuzzy logic and ANNs was deployed. The target value was the sum of rainfall in three consecutive months (seasonal forecasts). The collected input features consisted of monthly rainfall data, air temperature, vertical motion, divergence, vorticity, humidity, zonal and meridional wind, Nino 3.4, and dipole Sea Surface

Temperature (SST). The range of the data was between 1961 and 2010. The first 30 years were used for training, and the remaining for testing. The NFN model was compared to the Regional Spectral Model (RSM) which is based on GCMs by calculating RMSE and r values. In terms of correlation, authors showed that NFN was better in seven out of the selected eight locations where the highest correlation recorded was 0.76. For RMSE, lower values were obtained in five of the eight locations using the proposed NFN model. RMSE varied between 120 mm and 230 mm when compared to observed values in the eight selected locations.

Gan, Chen, Yang, and Yiu developed a study to predict the probability of rain falling during the wheat season harvest in Shangqiu, China [55]. The authors intended to forecast whether there would be rain during the period of harvesting. For training, 15 years of data were used, and another five years for testing. MATLAB was used to construct the network. The Levenberg-Marquardt algorithm was selected to train the ANN, while tansig and logsig were added as the activation functions. Different models were generated while changing the number of neurons. The results of the ANN were compared to a regression prediction equation, and accuracy was recorded as 100% and 67% respectively. No attempt was made to specify the amount of rainfall.

Moustris, Larissi, Nastos, and Paliatsos used ANNs to forecast rainfall variables for four consecutive months in four meteorological stations in Athens, Alexandroupolis, Thessaloniki, and Patras in Greece [69]. In order to predict the maximum, minimum, mean, and the cumulative precipitation totals for four consecutive months, 16 ANNs were developed. Different rainfall amounts were collected for each location: Athens (1891-2005), Alexandroupolis (1947-2003), Thessaloniki (1931-2003), and Patras (1901-1993). A trial and error method was followed to select the best network for each variable in each location. Seven input features were used in all the models. RMSE, R^2, and Index of Agreement (IA) were used to assess each network performance. Accuracies varied based on the amount of rainfall encountered in each location. The authors reported that peak values were not well predicted by the models. R^2 values for the 16 models ranged between 0.141 and 0.603.

Shukla, Tripathi, Pandey, and Das utilized Nino indices to predict the Indian Summer Monsoon Rainfall Index (ISMRI) [17]. Two models were proposed: a regression model, and a multi-layered FFNN. A total of 53 values were used to conduct the study. A trial and error method was followed to select the number of neurons. Three neurons were finally selected in the hidden layer. The sigmoid function was used as the activation function. Seven regression models were also developed to predict ISMRI. Lagged values of Nino indices were used as predictors. These lagged values were taken because of their correlation to ISMRI. The FFNN model was more accurate than the regression models in terms of RMSE and r values. The Pearson correlation value reported by the authors was 0.66 over the testing dataset (13 observations).

Nagahamulla, Ratnayake, and Ratnaweera examined the appropriateness of ANNs in forecasting seasonal monsoon rainfall in Sri Lanka [70]. Four months were targeted in the study: May, June, July and August. Multiple variables were proposed as possible predictors for each month, including climate indices. To determine the predictors of each month, correlation analysis between each available input and output were applied. Four ANNs were developed, each representing a month. For training, 20 years of data were taken, and ten years for testing. The authors started with the same architecture for the networks, then networks were pruned to remove unwanted nodes. RMSE was used to measure the accuracy of the models, and the highest accuracy was obtained in June (0.071 on normalized data). The authors did not compare the proposed networks to other methods. They reported that the four months in 2007 were not accurately predicted by any of the networks. They suggested optimizing network parameters to obtain higher accuracy in forecasting rainfall.

Campisi-Pinto, Adamowski, and Oron designed a BPNN to forecast monthly water demand in Syracuse, Italy [71]. To conduct the study, 86 measurements were used (January 2002 to December 2008). A set of five models were coupled with wavelet-denoising techniques: Haar and Daubechies of types db2, db3, db4, and db5. Another non-coupled ANN and a MLR model were developed for comparison purposes. A total of 196 network configurations were constructed to search for the best model. The network coupled with Haar produced the lowest RMSE and highest correlation in one-month ahead forecasts compared to other approaches.

Abhishek, Singh, Ghosh, and Anand developed a model that used ANN to forecast the maximum temperature for 365 days per year [72]. Input and target data were maximum temperature values only. MATLAB was used to conduct experiments. A FFNN with BP principles was the model used. The ANN was trained using the Levenberg-Marquardt algorithm. For training, 60% of collected data were used, 20% for validation, and the remaining 20% were used for testing. Model performance was measured in terms of Mean Squared Error (MSE), which is the squared difference between the observed and the predicted values. The authors aimed to study the effect of neurons on each hidden layer, the number of samples to add as input, and the number of hidden layers in the forecasts. Various models of different combinations were developed. The results showed that increasing the number of neurons/layers decreases the Mean Square Error (MSE). The authors indicated that increasing the number of samples reduced the MSE. In addition, they observed that a problem of over-fitting occurred when increasing the number of hidden layers.

Charaniya and Dudul applied an ANN approach to forecast Indian monsoon rainfall [73]. A Focused Time Delay Neural Network (FTDNN), an ANN that can memorize data based on a delay line in its structure, was selected as the network topology. Rainfall and Indian Ocean Dipole Index (IOD) were used as input features to predict one-month rainfall ahead of time. Data were normalized before being added into the network. A three-layered FTDNN with Gamma memory and conjugate gradient back propagation algorithm was used to produce forecasts (4-6-1 as architecture). Data were partitioned into: 60% for training, 15% validation, and 25% for testing. Multiple models that differed in regards to the number of hidden neurons in the hidden layers were developed, and accuracy based on the MSE was measured. The authors reported that a model with six hidden neurons was the best model (MSE of 0.0031 with training data, and 0.0105 with validation data). This model was used to perform forecasts on the test dataset. A Pearson correlation of 0.934, a Normalized Root Mean Square Error (NRMSE) of 0.149, and a MAE of 0.278 were reported. The generated network was not compared to any other model, and the data values collected were small compared to other studies. The authors reported a prediction accuracy of 93% during the testing period.

An ANN based approach was developed to predict rain intensity for four consecutive months in Athens, Greece [74]. The authors used 111 years of data to model and test the ANN models. Three ANN models were designed to predict mean monthly rain intensity, maximum monthly rain intensity, and minimum monthly rain intensity for four consecutive months. Seven input features were used as predictors. The three models recorded R^2 values of 0.443, 0.242, and 0.515 respectively. These networks were not compared to other techniques. The authors pointed to the need for further research in order to predict peak intensity rainfall values.

Singh and Borah used a FFNN to predict Indian Summer Monsoon Rainfall (ISMR) [56]. The monthly rainfall values for June, July, August, and September from 1871 to 2010 were collected. In addition, the sum of these months was taken (ISMR total). Five network models were developed to predict each collected variable. A different number of inputs were given to each network. The number of neurons was selected manually for the number of inputs plus one. Each network targeted a month, and the fifth network targeted ISMR. The authors stated that each network was trained multiple times, and outputs were averaged. RMSE, r, and Performance Parameter (PP) were used to measure the accuracy of the models. The networks were compared to an alternative approach, that uses BPNNs with different architectures and network parameters, in which lower errors have been recorded with the five proposed networks.

ANNs have been combined with GAs in addressing weather forecasting problems. Meng conducted an experiment to optimize BPNN weights using a GA [75]. The author aimed to combine the advantages of the two machine learning algorithms. Daily temperature values were targeted in the study. A total of 246 samples were used for training, and 31 samples for testing. Results showed an error of 0.001 with the improved genetic ANN.

Mekanik, Imteaz, Gato-Trinindad, and Elmahdi used climate indices and rainfall data to estimate long-term spring rainfall across Victoria, Australia [76]. A MLR model and an ANN model were used in the study. To train the network, the Levenberg-Marquardt training algorithm was used. Two types of activation functions were deployed; the tansig activation function was used in the hidden layers, while the purelin activation function was used in the output layer. The input data of the models was composed of the Nino 3.4, the SOI, and the Dipole Mode Index (DMI), which are identifiers of the ENSO and the IOD. Data were collected for three regions in Victoria, with each region having three weather stations. Datasets were divided into training (1900-1990), and validation (1991-2006). Three years of sample dataset (2007-2009) were used to measure the accuracy of both models. Data were normalized before being added into the two models. MSE, MAE, r, and the Willmott index of agreement (d) were used to measure the accuracy of the two models. The ANN model achieved lower MSE for most of the nine weather stations (eight out of nine) used in the study. Higher correlation coefficient values were obtained with the developed network compared to the MLR model. MAE was higher with the ANN. The designed ANN had better results in all locations based on the Willmott index of agreement (d).

Mekanik and Imteaz investigated the ability of an ENSO indicator (Nino 3.4) in forecasting spring rainfall for six areas in eastern Australia (three in Queensland, and three in Victoria) [40]. The input features vector consisted of three lagged values of the selected index, and the network output was spring rainfall. A dataset was selected between 1900 and 2009, and was divided into three datasets: training (1900-1990), validation (1991-2006) and testing (2007-2009). Levengberg-Marquadt was designated as the training algorithm. Accuracy in the training dataset varied between 0.25 and 0.79 in terms of Pearson correlation (r). Accuracy in the testing dataset varied between -0.97 and 1.00. Better generalization ability was found for areas in Queensland. The authors suggested using additional climate indices when forecasting rainfall for areas in Australia.

Jiang and Wu proposed a forecasting approach based on ANNs and EAs (GA and simulated annealing) [77]. The authors applied a hybrid algorithm that combines a GA and simulated annealing to find the initial weights of the ANN. Rather than using random initial weights, an algorithm was applied before running BP. Then they used a BP algorithm to search for the optimal set of weights. Data between 1991-2009 (25 years) were taken from 18 stations in the selected locations. In testing, 36 observations were used. Previous rainfall and run-off values over six months were used as input features. The network architecture was determined by a trial and error method to be 4-6-1. The proposed training algorithm was compared to alternative training methods based on the Genetic Algorithm Neural Network (GA-NN), BPNN, and Autoregressive Integrated Moving Average (ARIMA) models. Better accuracy was recorded with the Genetic Algorithm Simulated Annealing Neural Network (GASA-NN).

Jiang and Wu proposed a hybrid EA based on Particle Swarm Optimization (PSO) and GA in order to evolve ANN weights and architecture [78]. GA was used to help the PSO get out of local optimum. Hidden nodes were encoded as binary strings (1 meant a connection, 0 meant no connection). Weights were encoded as float strings between -1 and +1, and were randomly generated. The activation functions were manually assigned. GA parameters were applied if the termination condition was not found on PSO. GA and PSO were utilized to find the best optimal set of network parameters in a rainfall forecasting problem. The dataset ranged from 1991 to 2009. A total of 252 samples were collected, with 144 for training, 72 for validation, and 36 for testing. Six antecedent monthly rainfall figures were used in the input features vector. The proposed approach was compared to the BPNN, GA-NN, and Particle Swarm Optimization Neural Network (PSO-NN), and better performance was obtained with the hybrid evolutionary NN. RMSE recorded a value of 95.96 mm with BPNN, 67.98 mm with GA-NN, 68.93 mm with PSO-NN, and 67.92 mm with the hybrid PSOGA-NN that was used.

Rani, Srinivas, and Govardan used an ANN model to forecast monthly rainfall for Andhra Pradesh in India [79]. A Teaching Learning Based Optimisation (TLBO) ANN that simulates the teaching-learning phase in life was used in this study. The learning phase was modified with the used, in an attempt to enhance results. Monthly rainfall historical values were obtained from the Indian Institute of Tropical Meteorology (IITM). The dataset was composed of 1692 monthly observations from 1871 to 2011. Three different samples were used: a training sample, testing sample, and a hold out sample. An ANN that used BP algorithm for training was developed for comparison purposes. Better results in terms of RMSE were shown with the network trained using modified TLBO.

Sheela and Deepa proposed an ANN based approach to estimate wind speed [42]. Two network topologies were proposed: BPNN, and a Radial Basis Function Neural Network (RBFNN). Three parameters were taken as input features: temperature, wind direction, and past wind speed values (lagged values). Following a trial and error method, seven neurons were selected in the hidden layer. Levengberg-Marquardt was used as the training algorithm for the BPNN. Various sizes of dataset were given in order to analyse the ANN architecture. MAE and Pearson correlation (r) statistical measurements were calculated to measure the convergence of each topology. Better accuracy was recorded with the RBFNN (0.0013) as against the BPNN (0.0397) in terms of MAE.

Khedhiri developed three ANNs to forecast monthly rainfall for Prince Edward Island, Canada [80]. Two of the three ANN models input features were pre-processed with moving average and exponential smoothing (Holt-Winters) transformations. The three models were compared to a standard Seasonal Auto Regressive Fractionally Integrated Moving Average (SARFIMA). RMSE and r values were calculated to assess each model’s performance. The three models had better RMSE and r than SARFIMA, with Holt-Winters exponential smoothing having the highest accuracy. A RMSE of 0.5038 was obtained with the ANN developed using the original dataset. The two models with pre-processed inputs reported RMSE values of 0.1794 (moving average) and 0.1300 (exponential smoothing). The SARFIMA showed a RMSE of 1.8003.

Mislan, Haviluddin, Hardwinarto, Sumaryono, and Aipassa employed ANN to forecast rainfall for Tenggarong station, east Kalimantan, Indonesia [81]. Monthly rainfall data were collected between 1986 and 2009 for training and testing the models. For training, 75% of data were used, and the remaining for testing. Data were normalized before being added into the network. A BPNN was utilized. Levenberg-Marquardt was the algorithm selected for training the network and three activation functions were used: tansig, logsig. and purelin. Three architectures of BPNN were studied: 2-50-10-1 with 500 epochs, 2-50-20-1 with 1000 epochs, and 2-50-20-1 with 1500 epochs. To measure accuracy, MSE was calculated for each model. Compared to the other two models, the second model gave the better performance in both training and testing. In the test data, MSE was reported as 0.701 for 2-50-20-1 with 1000 epochs, 1.744 for 2-50-10-1 with 500 epochs, and 14.672 for 2-50-20-1 with 1500 epochs architecture. Authors did not compare the generated network against any other model. Instead they released forecasts for five consecutive years from 2009-2014.

Deo and Sahin conducted a study to forecast a relatively recent drought index known as the Standardized Precipitation and Evapotranspiration Index (SPEI) [82]. To predict SPEI, large-scale climate indices were utilized as the predictor variables. SPEI can assess drought impacts on multiple time-scales, and can be applied for separated regions. SPEI is used in Europe and China. It depends on rainfall, temperature, and evapotranspiration. A total of 18 site specific and climate input features were used as input to forecast SPEI. A total of 30 ANNs with various groupings of training algorithms, activation functions, and number of neurons in the hidden layer were formed. One model was selected as the best based on statistical evaluation. The authors affirmed the suitability of ANNs in predicting SPEI.

Wu, Long, and Liu combined PSO and GA to evolve a RBFNN [83]. The worst individuals in the GA iteration were manipulated using PSO mechanisms. The proposed RBF-HPSOGA approach was used to select the best components of the RBFNN. It was deployed over a rainfall prediction task in Liuzhou, China. Data ranged from 1949 to 2011. The available input features were used with no selection or optimization. The RBF-

HPSOGA approach was compared to alternative approaches: a single RBFNN, and a RBFNN evolved using pure GA (RBF-GA). Better accuracy, in terms of RMSE, r, and Mean Absolute Percentage Error (MAPE), was recorded with the authors’ model. The recorded accuracy in terms of RMSE was reported to be 67.72, 111.90, and 170.46 for RBF-HPSOGA, RBF-GA, and RBFNN respectively.

Mekanik, Imteaz, and Talei used multiple Adaptive Network-Based Fuzzy Inference Systems (ANFIS) models to predict spring rainfall (September-November) for nine locations in eastern Australia [84]. ENSO indications, IOD and Inter-Decadal Pacific Oscillation (IPO) were used as possible predictors when generating the datasets. Eight models were developed for each location based on single and combined climate indices. The models were trained using values between 1900 and 1999, while 10 years (2000-2009) were used for testing (10 values). The authors compared their approach to ANNs and the POAMA forecasting model. Better accuracy, in terms of RMSE, was found in five out of the nine locations using the ANFIS model. In addition, lower RMSE was obtained in five of the nine locations when compared to POAMA.

Jaedong and Jee-Hyong developed a weather prediction model based on SVMs to predict hazardous weather conditions, including monthly rainfall [85]. A top-down selection method was utilized to select weather attributes with the highest effect on classification. An overall accuracy of 79.61% was recorded with the developed models. These models are currently used as official forecasting models for hazardous weather prediction in Korea.

Kashiwao et al. developed a forecasting system to predict hourly rainfall for selected locations in Japan [43]. The proposed system utilized an ANN as the prediction model. The authors wanted to predict the heavy summer rainfall for each afternoon. Two main topologies were investigated when developing the proposed system: MLP, and RFBNN. MLP was trained using back propagation and a proposed random optimization (RO) technique. The Least Square Method (LSM) was used to train the RBFN topologies. The sigmoid function was selected as the activation function in MLP. Rainfall probability was initially targeted in the experiments, where six statistical measurements were used to assess each model’s performance. A 0.5 mm threshold was used to differentiate rain from no rain. A value greater than 0.5 mm represented a precipitation event. A value lower than 0.5 mm represented a non-precipitation event. Three cities were used in the first stage. Data from the morning of each day were used to predict precipitation for the afternoon of the same day. A total of 30 MLPs were trained with different initial weights, and 101 RBFNNs were trained while changing the standard deviation. Lower accuracies were obtained when comparing the forecasts to the official estimates released by the Japan Meteorological Agency (JMA). An additional 16 stations were used to verify the accuracy of the models. New input features based on lagged values were introduced. The output values were set to the total amount of precipitation in the remaining time of the day. For each location, 20 MLPs and 101 RBFNNs were trained. Better accuracies were obtained with the MLP models, but accuracies were still lower than the accuracy of forecasts released by the JMA.

Le, El-Askary, Allali, and Struppa conducted a study to predict the Palmar Z index, a drought index, in California, USA [86]. A total of 1452 monthly points between January 1895 and January 2016 were gathered. Of these, 1332 points were used for training, and 120 for testing. The Recurrent Neural Network (RNN) topology was employed as the forecasting model. Pearson correlation (r) was used to measure the accuracy of the models. The correlation varied between 0.641 and 0.434 from one to three months lead time.

Vathsala and Koolagudi proposed an approach for forecasting peninsular Indian summer monsoon rainfall [38]. A closed-itemset-generation-based association rule method was utilized for feature selection, and K-means clustering was used for dimensionality reduction. The processed data were then added to a MLP of seven neurons in the hidden layer so as to classify peninsular Indian summer monsoon rainfall: flood, excess, normal, deficit, and drought. The number of neurons was selected based on the average of input features and number of classes. The MLP was trained with different partitioning criteria: 10-fold, 5-fold, and 70%, 30% training testing datasets. The accuracies recorded were 95.59%, 94.59%, and 90.90% for 10-fold, 5-fold, and 70%, 30% respectively. The results were compared to alternative approaches based on the ability to identify rainfall conditions.

The following table represents a summary of the reviewed studies using single machine learning based models.

Table 2.1 Summary of the studies using single machine learning algorithms.

Year	Author	Location	Technique	Method	Forecasted attribute
2001	Luk, Ball, and Sharma [59]	Australia	TDNN MLFFNN ELMAN	Trial and error	Rainfall 15 min ahead
2005	Chaudhuri and Chattopadhyay [60]	India	ANN	Manual selection	Temperature Relative humidity
2005	Baawain, Nour, El-Din, and El-Din [50]	Global	MLP	Trial and error	SOI Nino3.0
2007	Chattopadhyay [61]	India	MLFFNN	Manual selection	Average rainfall in monsoon season
2007	Ayalew, Iler, and Reik [62]	Ethiopia	ANN	Trial and error	Run off (1 hr) Run off (2 hrs) Run off (3 hrs) Run off (4 hrs) Run off (5 hrs) Run off (6 hrs)
2008	Nasseri, Asghari, and Ebedini [63]	Australia	FFNN	Optimization of network weights. Sensitivity analysis to determine input features.	Rainfall events
2009	Karmakar, Kower, and Guharhakurta [64]	India	PNN FFMLNN	Manual selection	Average rainfall
2009	Hung, Babel, Weesakul, and Tripathi [65]	Thailand	GFNN	Trial and error	Hourly rainfall up to 6 hrs ahead
2009	Sedki, Ouazar, and El Mazoudi [66]	Morocco	ANN	Optimization of initial weights using GA	Daily run-off
2010	Wang and Sheng [67]	China	GFNN	Trial and error	Annual rainfall
2011	Castro et al. [68]	Brazil	NFN	Manual selection	Three months rainfall
2011	Gan, Chen, Yang, and Yiu [55]	China	BPN	Trial and error	Classification: rain/no rain
2011	Moustris, Larissi, Nastos, and Paliatsos [69]	Greece	ANN	Trial and error	Rainfall attributes
2011	Shukla, Tripathi, Pandey, and Das [17]	India	FFNN	Selection of input features based on correlation with target	ISMRI
2011	Nagahamulla, Ratnayake, and Ratnaweera [70]	Srilanka	ANN	Trial and error to select best subset of parameters and features.	Monthly rainfall
2012	Campisi-Pinto, Adamowski, and Oron [71]	Italy	ANN	Wavelet-denoising on time series	Monthly water demand
2012	Abhishek, Singh, Ghosh, and Anand [72]	Canada	FFNN	Multiple combinations of network parameters and size of input features	Maximum daily Temperature
2012	Charaniya and Dudul [73]	India	FTDNN	Trial and error	One month ahead rainfall

2013	Nastos, Moustis, Larissi, and Paliatsos [74]	Greece	MLP	Manual selection	mean monthly rain intensity maximum monthly rain intensity minimum monthly rain intensity
2013	Singh and Borah [56]	India	FFNN	Network parameters selected manually.	ISMR
2013	Meng [75]	China	BPNN	Optimization, using GA and ANN to forecast rainfall	Maximum daily temperature Minimum daily temperature
2013	Mekanik, Imteaz, Gato-Trinindad, and Elmahdi [76]	Australia	FFNN MLR	Trial and error	Long-term rainfall
2013	Mekanik and Imteaz [40]	Australia	MLP	Trial and error	Spring rainfall
2013	Jiang and Wu [77]	China	FFNN	Optimization	Monthly rainfall
2013	Jiang and Wu [78]	China	FFNN	Optimization PSO-GA-NN	Monthly rainfall
2014	Rani, Srinivas, and Govardan [79]	India	ANN	New algorithm for training the ANN weights named TLBO.	Monthly rainfall
2014	Sheela and Deepa [42]	India	RBFN BPNN	Trial and error	Wind speed
2014	Khedhiri [80]	Canada	BPNN	Times series processing	Monthly rainfall
2015	Mislan, Haviluddin, Hardwinarto, Sumaryono, and Aipassa [81]	Indonesia	BPNN	Three architectures	Monthly rainfall
2015	Deo and Sahin [82]	Australia	ANN	Trial and error	SPEI
2015	Wu, Long, and Liu [83]	China	ANN	PSO and GA optimization on network parameters	Monthly rainfall
2016	Mekanik, Imteaz, and Talei [84]	Australia	ANFIS	Multiple combinations of input features	Spring rainfall (3 months)
2016	Jaedong and Jee-Hyong [85]	Korea	SVM	Top down approach to select features	Hazardous weather
2017	Kashiwao et al. [43]	Japan	MLP RBFN	Trial and error	Hourly rainfall
2017	Le, El-Askary, Allali, and Struppa [86]	USA	RNN	Manual selection	Drought
2017	Vathsala and Koolagudi [38]	India	MLP	Closed-item-set-generation for features selection, and K-means clustering for dimensionality reduction.	Flood Excess Normal Deficit Drought

2.3 Overview of Ensembles

An ensemble is a combination of a finite number of classifiers combined to perform the same task [87]. Maqsood, Kahn, and Abraham defined ensemble of ANNs as “a learning paradigm where a collection of a finite number of ANNs is trained for the same task." [88]. Gadgay and Kulkarni defined an ensemble of ANNs as a “set of independently trained member models of SNNs whose outputs based on prediction are combined logically to get a single estimate of a desired output.” [89]. Ensemble techniques have been used in different types of regression and classification applications [18, 90-94].

The application of ensemble techniques is partitioned into two main steps: the first is to setup the ensemble members, and the second is to combine them. An ensemble component can be any algorithm which is trained on a certain problem (ANN, SVM, etc.). Various data fusion methods have been used to combine the output of different models to form the ensemble. The classifiers are then combined using various techniques including average, majority voting, and stacking. The basic ensemble of an ANN is a mechanism in which all of the outputs of the ANNs are averaged. Ensembles of diverse ANNs and other classifiers for weather forecasting will be developed and tested as a part of this research.

2.4 Ensembles of Neural Networks in Weather Prediction

Maqsood, Kahn, Abraham, Huang, and Abdalla used ANNs to predict temperature, wind speed, and relative humidity up to 24 hours ahead for autumn, winter, spring, and summer [95, 96]. To train the networks, hourly seasonal data were gathered. Hourly values for February 26, May 6, August 7, and November 10 were used to test the trained models. Multiple ANN models were deployed and tested: Multi-Layered Perceptron Neural Network (MLPNN), Elman Recurrent Neural Network (ERNN), RBFNN, and a Hopfield Model (HFM). To determine the architectures of the MLPNN, RBFNN, HFM, and ERNN, a trial and error approach was used. MLPNN had lower errors than HFM, but its learning-process was time-consuming and relied on the network parameters. ERNN was able to predict the dynamic behaviour of weather compared to MLPNN. The RBFNN obtained the best performance in terms of training time and accuracy. HFM showed higher values for RMSE, MAD, and MAPE. Finally, the outputs of the models were combined to form an ensemble of networks. The generated ensemble produced the lowest MAPE.

Maqsood, Kahn, and Abraham conducted another study in which ensembles of ANNs were used to forecast three weather variables (temperature, wind, humidity) 24 hours ahead of time [88]. The proposed ensemble consisted of multiple network models: MLPNN, Fully Recurrent Neural Network (FRNN), RBFNN, and HFM. Data collected were pre-processed by being divided into four seasons. Data were then normalized before being used. The training time varied between 5 to 30 minutes for the single networks. Time consuming ANNs led to better results. HFM was the less accurate, while RBFNN showed better performance. For the ensembles, two types of ensemble methods were proposed and applied: Weighted Average (WA), and Winner Takes All (WTA). Four statistical measurements were calculated: RMSE, MAD, MAPE, and r. WTA revealed an ensemble with higher performance than WA. The second method was more accurate when compared to single and other ensemble methods. WTA had better accuracy in terms of RMSE for all the seasons and weather variables, except for humidity during the summer season. In both studies, the authors used trial and error to select the best single and ensemble models.

Monira, Faisal, and Hirose developed ensemble models to predict one day ahead rainfall in Fukuako, Japan [97]. Three steps were followed to develop the ensembles. Least Angle Regression (LARS) is a variable selection technique, which ranks the candidate predictors according to their predictive content. Time Series Least Angle Regression (TS-LARS) was used to select lagged values of input features. Levengberg-Marquardt was used as the training algorithm of the single models (FFNNs). Independent Component Analysis (ICA) was used to determine the models which were independent of each other. Ensemble components were then ordered, based on Mutual Information (MI). Then, highly ranked networks were combined using weighted average. Data were selected between 1990 and 2010. Four months were used: June, July, August, and September. Nine variables were used as initial inputs of the single networks. TS-LARS selected four variables out of the nine features. Temperature was mentioned as the predictor with highest effect. The optimal number of lag values obtained was two. 100 networks were created initially, and then 10-15 were selected. The proposed method was compared to another ensemble model with no ICA and MI. RMSE, Correlation of Efficiency (CE), Persistency Index (PI), Bias, MAPE, and r statistical measurements were calculated, and better accuracy was found in terms of r (0.88 proposed, 0.70 compared). In addition, lower RMSE was obtained with the proposed method. The authors concluded that the smaller number of components in ensemble, the better the forecasts.

Jin, Huang, and Zhao developed an ensemble of ANNs to forecast monthly rainfall for April in Guangxi, China [98]. The single models were constructed based on a BPNN and a PSO algorithm. The weights and architecture of the models were designed based on PSO. Average fusion method was used to combine the ensemble components. The method was compared to the multiple regression method, and better results were recorded with the authors' method. They reported that forecasting anomalies were unsatisfactory.

Ensembles of ANNs were used to forecast rainfall for Colombo, Sri Lanka [99]. Three ANN topologies including BPNN, RBFNN, and GRNN were used to develop the ANNs ensemble, and a weighted average fusion method was applied to combine the models. The dataset consisted of 41 years from 1961 to 2001, with 25 years for training, eight for validation, and the rest for testing. Based on a trial and error method, an ensemble that consisted of eight BPNNs, two RBFNNs and a GRNN was selected as the final model to be used. For each trained model, a separate prediction was generated, and then results were combined based on weighted average. RMSE was used to measure the accuracy of the models where the ensemble showed the best performance (8.06) against the other GRNN (8.22), RBFNN (8.69), and BPNN (9.44) single models. Similar results were shown when measuring performance based on MAE and r. Although it generated better forecasts, the ensemble was unable to predict higher rainfall accurately. The ensemble was also unable to estimate higher rainfall occurrences (more than 100 mm).

Nagahamulla, Ratnayake, and Ratnaweera presented a genetic algorithm and k-means clustering algorithm to select the most suitable ANNs to form an ensemble in two locations in Sri Lanka, Colombo and Katugastota [100]. To generate the ensemble members, the authors developed a pool that contained GRNN topologies that were trained with varied data. These varied training datasets were formed through different pre-processing mechanisms and data sampling techniques. The goal of the GA was to find the best ANN models in the pool so that the total Mean Square Error (MSE) of the results was reduced. A GRNN was used to combine the selected ANN members of the ensemble for each method (k-means clustering and GA). To measure the accuracy, RMSE and MAE were recorded. The collected data set contained 41 years of daily observed data for 26 features. For training, 25 years were used, eight for validation, and remaining eight for testing. The pool consisted of 1023 GRNN trained with different datasets. The two proposed methods were compared to bagging and boosting. The best ensembles in the two locations with each algorithm (GA and k-means) were compared, and GA based ensemble had the lowest RMSE in both Colombo and Katugastotata (7.33, 6.25), followed by k-means clustering method with 7.38 in Colombo and 6.37 in Katugastotata.

Saba, Rehman, and AlGhamdi used a MLP and a RBFNN to estimate precipitation [101]. The authors proposed combining the two network architectures to overcome the limitations of each architecture. The outputs of each model were combined using the average fusion method. The proposed approach was compared to single MLP and RBFNN models and produced a higher accuracy. A trial and error method was used to select network parameters.

The reviewed work is summarized in Table 2.2.

Table 2.2 Summary of ensemble studies.

Year	Author	Location	Technique	Method	Fusion	Forecasted attribute
2003	Maqsood, Kahn, and Abraham [95]	Canada	MLPN FRNN RBFN HFM	Trial and error	Average	Temperature 24 hrs ahead Wind speed 24 hrs ahead Humidity 24 hrs ahead
2004	Maqsood, Kahn, and Abraham [88]	Canada	MLPN FRNN RBFN HFM	Trial and error to select number of components	Weighted average Winner takes all	Temperature 24 hrs ahead Wind speed 24 hrs ahead Humidity 24 hrs ahead
2011	Monira, Faisal, and Hirose [97]	Japan	FFNN	TS-LARS to select input features for single models. Independent Component Analysis (ICA) and Mutual Information (MI) to select models.	Average	Daily rainfall
2012	Jin, Huang, and Zhao [98]	China	BPNN	Architectures and weights of single models were optimized using PSO	Average	Monthly Rainfall (April)
2012	Nagahamulla, Ratnayake, and Ratnaweera [99]	Srilanka	BPNN RBFN GRNN	Trial and error to build single models.	Average	Rainfall
2014	Nagahamulla, Ratnayake, and Ratnaweera [100]	Srilanka	GRNN	Ensemble components selection GA and KNN	GRNN	Rainfall
2017	Saba, Rehman, and AlGhamdi [101]	Saudi Arabia	MLP RBFN	Trial and error	Average	Daily rainfall

2.5 Other Forecasting Methods

Kishtawal, Basu, Patadia, and Thapliyal utilized a GA to forecast rainfall in India [102]. Past rainfall values of three months (June, July, and August) were gathered to conduct the experiments. The authors aimed to find the optimal equation that represents the temporal variations of seasonal rainfall in India. An analytical expression was generated by the GA and then used to perform forecasting. Some 132 years of rainfall were used in the study between 1871 and 2003. To find the equation that best fit the data with the GA, 122 years were applied, and the remaining 10 years were used to validate. To measure the accuracy of the model, standard error and fitness strength were calculated. The fitness of the final equation was reported to be 0.705 for the training dataset (1871-1993), and 0.644 for the testing set (1993-2003).

Everingham, Clark, and Van Gorder developed a method that aims to forecast precipitation for the Australian sugar industry [103]. The authors focussed on early prediction of weather conditions (wet/dry) in the second half of the harvesting season (September-October-November). A model proposed by Clarke and Van Gorder [104], which is based on a mathematical equation that contains climate indices as attributes, was used in the study. Results were shown in terms of probabilistic values, where above or below median values were computed. The authors recommended that their findings be used as additional information for the industry.

Hawthorne, Wang, Shepen, and Robertson proposed an approach that combines bridging and calibration models to forecast rainfall [105]. These models are based on GCMs. To evaluate results, a skill score was developed and compared with climatology, which is the reference model in atmospheric science. Large grid areas were used in the comparison. The range of skill scores varied between -20 to +20% compared to climatology.

He, Guan, Zhang, and Simmons proposed a wavelet based MLR model to forecast monthly rainfall in South Australia [106]. Drought is relevant to large-scale climate variabilities, and may be to global warming as well. The study used historical rainfall data and large-scale climate indices as input to estimate South Australian precipitation on a monthly time-scale. In order to optimize input, Multi-Regression Analysis (MRA) based on Wavelet Transform (WT) were applied. WT is defined as a "useful mathematical tool that provides time-frequency representation of an analysed signal in time domain" [106]. Some 46 stations were used in the study. SOI, Southern Annular Mode (SAM), IOD, Pacific Decadal Oscillation (PDO), and Sub-tropical Ridge (STR) were used as possible predictors. The proposed wavelet multiple regression model was compared to a traditional multiple regression model, and improved accuracy was obtained.

Table 2.3 summarizes the forecasting approaches.

Table 2.3 Summary of other forecasting approaches.

Year	Author	Location	Method	Forecasted attribute
2003	Kishtawal, Basu, Patadia, and Thapliyal [102]	India	GA used to optimize an equation to predict rain	Rainfall probabilities
2008	Everingham, Clark, and Van Gorder [103]	Australia	Mathematical equation which contains climate indices	Rainfall prediction for harvest early
2013	Hawthorne, Wang, Shepen, and Robertson [105]	Australia	Calibration and Bridging models	Seasonal rainfall
2013	He, Guan, Zhang, and Simmons [106]	Australia	Multiple Linear Regression (MLR)	Monthly rainfall

2.6 Chapter Summary

In this chapter, several prediction techniques have been reviewed. The use of ANNs and other machine learning algorithms is increasing in weather forecasting applications. Based on the comprehensive review of related literature, the following conclusions can be drawn.

- FFNNs are the most commonly used topologies in weather related applications.
- There are no benchmarks datasets when using rainfall forecasting; all of the datasets used are location-based.
- In the traditional approach, trial and error based methods are followed to select the forecasting model elements.
- Much work on the applicability of ANNs has been carried out; however, there were limitations in the methods followed to design these algorithms.
- The accuracy of the prediction model is highly dependent on input features and network parameters.
- In most of the cases, the networks were incapable of estimating peak rainfall values.
- Increasing the lead time of the forecasts decreases the performance of the model.

Chapter 3 Data and Evaluation Metrics

This chapter describes the data used in deploying the proposed approaches. The statistical measurements used to assess and compare the forecasting models are listed and defined. In Section 3.1, rainfall and other weather attributes are explained. Multiple sources were utilized to collect the required weather variables. The sources of the collected data are listed. In addition, the steps followed to create datasets are outlined and described. The collection of input features was done throughout the time-line of this research project. In Section 3.2, several statistical measurements that were used in classification and regression tasks are explained. These measurements represented the baseline of the assessment of the proposed approaches.

3.1 Data

3.1.1 Output (Rainfall)

Multiple locations were selected to perform this research: Innisfail, Plane Creek, Bingera, and Maryborough in Queensland, and Yamba in New South Wales. The main reason for this selection is their closeness to sugarcane agricultural areas and infrastructure (mills). The selected locations are shown in Figure 3.1. All of the selected weather stations are located in eastern Australia. Innisfail was mostly used in deploying the proposed approaches. The main reason for selecting Innisfail is the high values of rain it receives throughout the year (3553.00 mm). The monthly averages for each selected location are shown in Figure 3.2. It is clearly noticeable that Innisfail had the highest amount of rainfall for all of the months. Details about rainfall values are shown in Table 3.1.

Seasonal and annual forecasts are the foci of this thesis. The main goal of the research is to find new rainfall forecasting systems, that are based on machine learning approaches,

Parts of this chapter appeared in: A. Haidar and B. Verma, "Monthly rainfall categorization based on optimized features and neural network," in *AI 2017: Advances in Artificial Intelligence* (Lecture Notes in Computer Science, no. 10400), W. Peng, D. Alahakoon, and X. Li, Eds. Cham, Springer, 2017, pp. 208-220.

from which informative and useful forecasts can be made. Such forecasts would provide very useful information for users. The types of forecasts to be generated in this research are summarized as follows.

- **Quantitative and Categorical**: The quantity of rainfall is the main focus. Some rainfall models use probabilistic values, which do not give an indication about the amount of rainfall. Although they can be useful, farmers require more specific information about the amount of rainfall in order to manage the season. The use of numerical modelling is illustrated in four out of the five proposed case studies.
- **Location**: Rainfall forecasts will be generated as location-specific, rather than using large grid areas to describe the forecasts.
- **Time**: Seasonal and annual forecasts are studied. Seasonal forecasting is the process of predicting rainfall for the next month(s). Data is used to predict the monthly amount/category of rainfall for the following months. Annual forecasting is the process of forecasting for the next year. Data is used to predict the monthly amount of rainfall up to one year ahead.

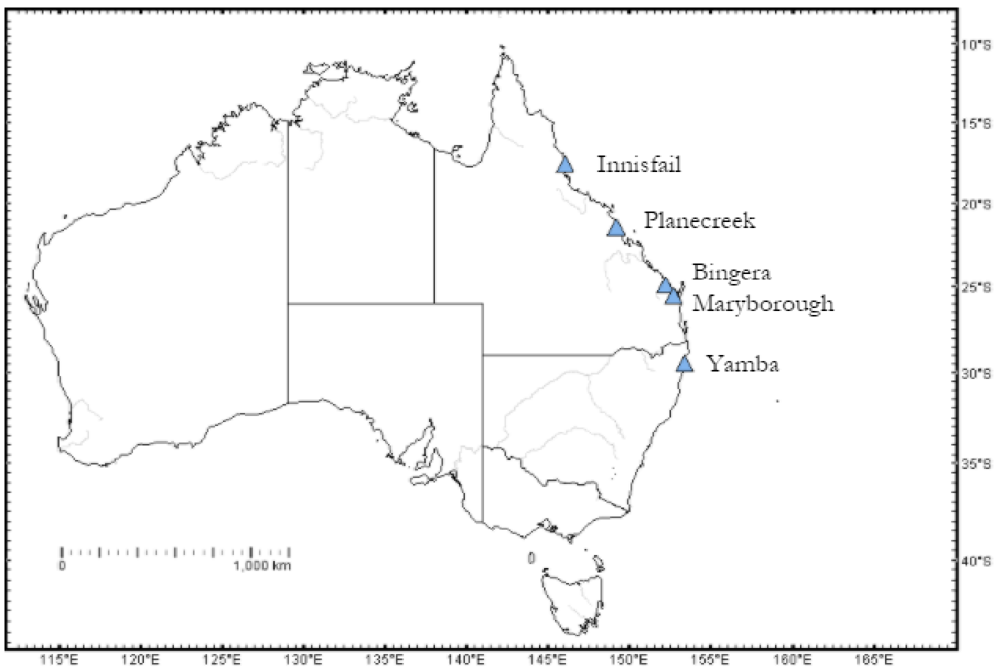

Figure 3.1 Selected locations used in this research.

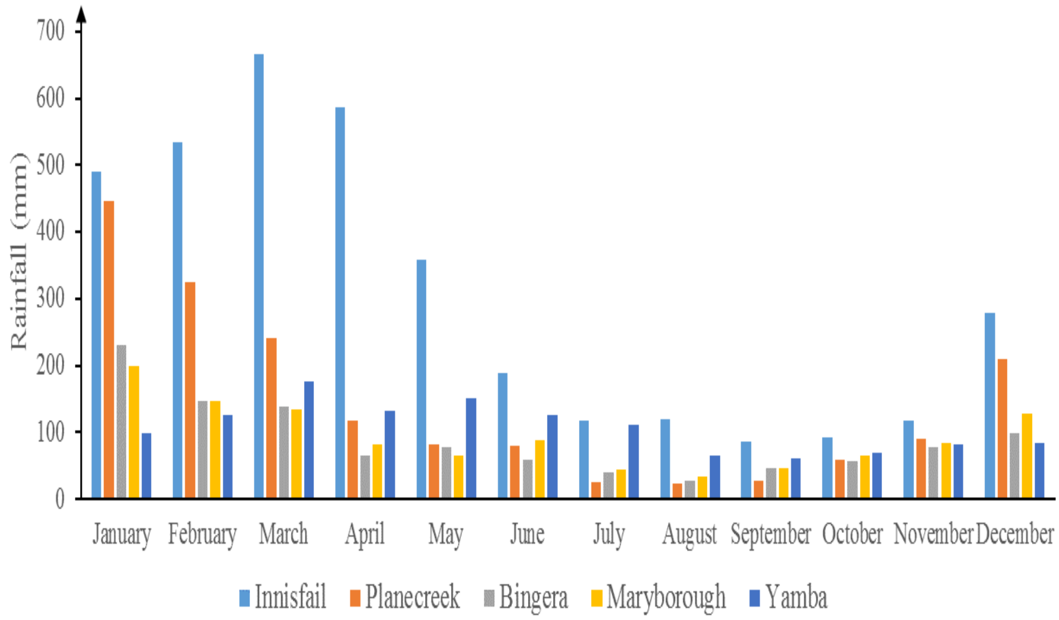

Figure 3.2 Monthly rainfall average values for each month in the selected locations.

Table 3.1 Details of the selected locations used in this research.

Location	Start	Finish	Annual Average (BOM 2015)
Innisfail	Jan. 1908	Dec. 2015	3553.0 mm
Plane Creek	Jan. 1909	Dec. 2015	1762.3 mm
Bingera	Jan. 1901	Dec. 2015	1024.5 mm
Maryborough	Jan. 1909	Dec. 2015	1138.2 mm
Yamba	Jan. 1899	Dec. 2015	1463.7 mm

3.1.2 Input Features (Possible Predictors)

3.1.2.1 Temperature values

Temperature values have been used as possible predictors of rainfall conditions in most rainfall forecasting models. Mean monthly maximum temperature and mean monthly minimum temperature are used as possible predictors of rain in this research.

3.1.2.2 Climate Indices

Past studies have linked Australian climate to various phenomena in the world. These phenomena are measured and represented as climate indices. Extensive research has been done to study and analyse the effect of these indices on rainfall variability across regions of Australia. Multiple studies have shown that climate indices are potential predictors of seasonal and annual rainfall [107, 108]. Climate indices are considered vital for any type of rainfall forecasting models.

ANNs require data as input in order to perform. Usually, the larger the dataset, the better the forecasts are. Data is an important aspect of all forecasting approaches. Different types of climate attributes can be applied and analysed in rainfall forecasting. ANN inputs can be chosen based on their contribution to formation of rain. Although some climate indices are considered rainfall influencers, such as the Madden Julian Oscillation (MJO) and winds, recorded values were considered insufficient. Therefore, these values were discarded when generating input features datasets. The numerical indices used as possible predictors when developing the forecasting models used in this research are explained below.

- ***A. El Niño Southern Oscillation (ENSO):*** Climate indices have different impacts on rainfall variability across Australia. ENSO is the foremost source of weather variability [109]. It is an important climate phenomenon that affects primarily the atmospheric conditions of the tropical Pacific region, including the local climate and weather in Australia [110]. ENSO affects eastern and north-eastern areas of the country [111]. Multiple indices are used as measurements for this phenomenon, including the SOI and Nino 3.0, the two most widely used indicators [50]. ENSO affects multiple countries in the world and the agricultural activities in those countries [107]. According to Fawcett, ENSO must be taken into consideration when developing forecasting systems for Australia [112].
- ***B. Southern Oscillation Index (SOI):*** The SOI measures the difference in surface air pressure between Darwin and Tahiti [50]. A high positive value SOI represents a La Niña event, while a high negative value represents an El Niño event. La Niña events are associated with cooling, while El Niño events are associated with warming. SOI values were the highest ever recorded in October and December 2010, and in February and March 2011. Floods occurred in specific locations in Australia between September 2010 and March 2011 [113]. Australia's wettest 24 months were recorded between April 2010 and March 2012 [113]. Attempts have been made to predict rainfall using the SOI [108]. Finally, the SOI has been

widely used as a possible predictor in various types of forecasting applications [76, 106, 114].

C. ***Nino Values (Nino 1.2, Nino 3.0, Nino 3.4, Nino 4.0):*** Nino values are used to monitor the tropical Pacific area. Each index represents sea surface temperature in a specific region: Nino 1.2 (0°-10°S, 90°W-80°W), Nino 3.0 (5°N-5°S,150°W-90°W), Nino 3.4 (5°N-5°S,170°W-120°W), and Nino 4.0 (5°N-5°S,160°E-150°W) [115]. The Nino 3.0 index characterizes the sea surface temperature anomalies averaged over the area constrained by 5°N to 5°S and 90°W to 15°W [50]. Nino 3.4 is the average sea surface temperature anomaly (170°W to 120°W) in the region bounded by 5°N to 5°S. A small temperature deviation in the Nino 3.4 region significantly increases or decreases the chance of rainfall in Australia. Climate anomalies are expected to be an effect of ENSO, which may lead to consequences such as high crop loss. This temperature deviation is frequently referred to as the value of the Nino 3.4 index [11]. Nino 3.4 values and additional Nino values (Nino 4.0, Nino 1.2) that can be used as a measurement for the ENSO phenomena were collected from the KNMI climate explorer and added to the dataset.

D. ***Inter-decadal Pacific Oscillation (IPO):*** The IPO is an oceanographic and meteorological phenomenon that is seen around the Pacific basin. It affects rainfall and temperature over different countries, including Australia [116]. IPO covers the whole Pacific basin and is linked to decadal weather variability over some of its parts. Salinger, Renwick, and Mullan claimed that the IPO controls decadal rainfall trends [117]. Studies have shown that the IPO influences Australian rainfall [118]. The IPO measures temperature and pressure variations in the Pacific Ocean. IPO records were added to datasets. IPO data were collected from the UK Met Office as monthly values.

E. ***Indian Ocean Dipole (IOD):*** The IOD is an ocean-atmosphere phenomenon that recent studies have shown contributes to rainfall variability across Australia [111]. It is in the equatorial Indian Ocean. The DMI is a measure of the IOD, defined as the difference in Sea Surface Temperature (SST) between the tropical western Indian Ocean (50°E-70°E and 10°S-10°N) and the tropical south-eastern Indian Ocean (90°E-110°E and 10°S-0°N) [111]. DMI based on HadISST1 was used in this study. This climate index was collected as daily values from the KNMI climate explorer, and converted to monthly representations for each month.

F. ***Tripole Index for Inter-Decadal Pacific Oscillation (TPI):*** The TPI is the difference between the Sea Surface Temperature Average (SSTA) averaged over the central equatorial Pacific, and the average of SSTA in the northwest and southwest Pacific [118]. It is a robust description for the IPO. The regions used to calculate the TPI are: region 1 (25°N–45°N, 140°E–145°W), region 2 (10°S–10°N, 170°E–90°W), and region 3 (50°S–15°S, 150°E–160°W).

G. ***North Pacific Index (NPI):*** The NPI is an area-weighted sea level pressure over the region 30°S-65°N, 160°E-140°W. The NPI is used to measure decadal variations connected to ENSO events. There have been links between NPI and precipitation in the southern Pacific region, where wetter conditions are associated with positive NPI values [119].

H. ***Pacific Decadal Oscillation (PDO):*** PDO is defined as “the leading empirical orthogonal function for North Pacific sea surface temperature monthly averaged anomalies” [120]. This index is considered to be long-term as it can stay in the same state for 20-30 years. PDO is recognized as an aggregation of multiple physical procedures [121]. Links have been made to its effect on ENSO phases (El-Niño, La-Niña) [122], and since ENSO is considered the first driver for climate variability, PDO was included as a possible forecaster.

I. ***North Atlantic Oscillation Index (NAO):*** The NAO is a climate phenomenon in the North Atlantic Ocean. It is originally described as the normalized pressure difference between Iceland and Azores High. It was then extended by using data from Gibraltar and composite sites in southwestern Iceland. It is considered to be a significant element in the global climate system [123]. This index is the farthest away from Australia in comparison with previously mentioned indices.

3.1.2.3 Sunspots

Monthly mean sunspots are the sum of the daily observed sunspots over the length of a month. The aim of using these values is to incorporate a sunlight variable, since sun is considered to be a key driver in a rainfall ecosystem.

3.1.3 Data Collection

In this research, different types of weather attributes that would affect rainfall were studied. The datasets consisted of rainfall, temperature values, climate indices, and sunspots. These variables were mainly collected from five different sources.

1. Bureau of Meteorology (BOM): The BOM is an executive agency of the Australian Government in charge of providing climate administration to Australia and encompassing ranges [124].
2. Royal Netherlands Meteorological Institute Climate Explorer (KNMI): KNMI is a web application that analysis climate data statistically [125].
3. Solar Influences Data Analysis Center (SIDC): The SIDC is the solar physics research division of the Royal Observatory of Belgium [126].
4. Earth System Research Laboratory (ESRL): The ESRL is a lab in the National Oceanic and Atmospheric Administration (NOAA) [127].
5. Climate of the 20th century website (C20C) [128].

This research focuses on monthly values for each attribute to be studied. Rainfall data are freely available on the BOM website as daily or monthly values from a large number of weather stations across Australia. Maximum and minimum monthly temperature values are made available on the same website as daily and monthly values. Monthly rainfall values, monthly minimum temperature, and monthly maximum temperature values were collected from the BOM. The source of each weather variable is shown in Table 3.2.

Table 3.2 Source of each weather variable.

Number	Climate Attribute	Source
1	Rainfall	BOM
2	Mean maximum temperature	BOM
3	Mean minimum temperature	BOM
4	SOI	BOM
5	Nino 1.2	KNMI
6	Nino 3.0	KNMI
7	Nino 3.4	KNMI
8	Nino 4.0	KNMI
9	IPO	C20C
10	IOD	KNMI
11	TPI	ESRL
12	NPI	KNMI
13	PDO	KNMI
14	NAO	KNMI
15	Sunspots	SIDC

3.1.4 Data Pre-processing

Pre-processing is an approach used to set up datasets for the ANN. An ANN requires full series values in order to forecast. Downloaded data were first reviewed to investigate missing values and unusual occurrences. Then the collected variables were manipulated where needed.

A. ***Data Infill:*** For rainfall and temperature, some data values were missing. The selected weather stations contained missing values for different durations. Datasets taken from the online resources contained missing values that ranged between one single value to years. Different techniques, as outlined below, were used in an attempt to recover these values.

- ***Replacemen*t*:*** Missing values were replaced with values from the nearest available weather station. An average was used when these were not available.
- ***Averaging:*** Data collected from weather stations for a specific location may contain missing information. The reasons for this

include machine faults, machine modifications, etc. Averaging was used with small range missing values. When additional data values were collected from the closest weather stations, the average of these values was calculated as a replacement for the missing value. Combining was used when this approach was not available.

- ***Combining:*** Combination is a method of joining data from two or more weather stations in order to represent a dataset for one specific location. Some weather stations were closed because of the establishment of a new stations. In order to have a longer record, the latest values of one dataset were merged with the start of another dataset so as to form one combined dataset to represent a specific location.

B. ***Lagging:*** Lagging is the process of adding previous values for a specific month. For the proposed datasets, values were shifted so that the network model could be useful for forecasting. Values from previous months were incorporated as input predictors for the following months.

C. ***Feature Generation:*** In some cases, new features were created to increase the identification of antecedent weather conditions. Features were created to determine the type of rain (extreme/normal) in antecedent months. For a month at time t, two new binary features were used to identify the occurrence of heavy rain at months at time $t-1$ and $t-12$. Those features were represented as $Plag$ and were used in some of the proposed case studies.

D. ***Dataset Normalization:*** The ANN performs better when using small ranges in training. However, the ranges of the climate attributes were totally different. Therefore, each column in the dataset was normalized to a smaller range between zero and one. For each feature, the upper and lower bounds were calculated. Then the difference between each value and the minimum was divided by the difference between upper and lower boundaries as shown in the following equation.

$$u_i' = \frac{u_i - u_{min}}{u_{max} - u_{min}} \tag{3.1}$$

Where u_i is a value in the set being normalized, u_{max} and u_{min} are the maximum and minimum values (upper and lower bounds) of the weather variable, u' is the new ranged value, $1 < i \leq n$, and n is the number of elements in a dataset.

E. ***Dataset Manipulation:*** Two main methods were followed in forecasting. The first one is all months and the second is one month. Due to seasonality in rainfall datasets, a second way of training the networks was introduced.

- ***All months:*** The first method uses all of the historical data as one dataset, and inputs this set into the ANN.
- ***One month:*** A single month optimization is a technique for inputting data into the ANN, where instead of adding all the monthly records as a single set, each month's data (rainfall and input features) are added to the ANN as a dataset.

F. ***Dataset Division:*** For each dataset created in this research, 75% were used for training, 15% were used for validation, and the remaining 10% were used for testing.

3.1.5 Datasets Generation (Case Studies)

The input features were collected consecutively. Throughout the time-line of this research, new input features were collected and used as possible predictors. In addition, the number of locations used in forecasting varied between studies. The collected variables were manipulated so that the generated input features would have the same start and finish time as rainfall. The created datasets are summarized in Table 3.3. In most of the studies, only one location was used. The summary of case studies is shown in Table 3.4. The input features used in each dataset are shown in Table 3.5. Dataset II in Case study 1 was used for comparison purposes in Chapter 5.

Table 3.3 Summary of the generated datasets.

Dataset	Location	Input Features	Partition	Forecasts
Dataset I	Bingera	10	One month	Yearly
Dataset II	Innisfail	11	One month	Yearly
Dataset III	Innisfail	28	One month	Monthly
Dataset IV	Planecreek	28	One month	Monthly
Dataset V	Bingera	28	One month	Monthly
Dataset VI	Maryborough	28	One month	Monthly
Dataset VII	Yamba	28	One month	Monthly
Dataset VIII	Innisfail	11	All months	Yearly

Table 3.4 Summary of each case study.

Case study	Datasets	Type	Chapter
Case study 1	Dataset I, Dataset II	Numerical	Chapter 4 Section 1
Case study 2	Datasets III, IV, V, VI, and VII	Categorical	Chapter 4 Section 2
Case study 3	Dataset II	Numerical	Chapter 5
Case study 4	Dataset VIII	Numerical	Chapter 6
Case study 5	Dataset III	Numerical	Chapter 7

Table 3.5 Summary of features used in each dataset.

	Location	**Bingera**	**Innisfail**	**Innisfail**	**Planecreek**	**Maryborough**	**Bingera**	**Yamba**	**Innisfail**
Nb	**Dataset ID**	**I**	**II**	**III**	**IV**	**V**	**VI**	**VII**	**VIII**
1	MaxT	✓	✓	✓	✓	✓	✓	✓	✓
2	MinT	✓	✓	✓	✓	✓	✓	✓	✓
3	SOI	✓	✓	✓	✓	✓	✓	✓	✓
4	Nino 1.2	✓	✓	✓	✓	✓	✓	✓	✓
5	Nino 3.0	✓	✓	✓	✓	✓	✓	✓	✓
6	Nino 3.4	✓	✓	✓	✓	✓	✓	✓	✓
7	Nino 4.0	✓	✓	✓	✓	✓	✓	✓	✓
8	DMI	✓	✓	✓	✓	✓	✓	✓	✓
9	IPO	✓	✓	✓	✓	✓	✓	✓	✓
10	Sunspots	✓	✓	✓	✓	✓	✓	✓	✓
11	TPI	-	✓	✓	✓	✓	✓	✓	✓
12	NPI	-	-	✓	✓	✓	✓	✓	-
13	NAO	-	-	✓	✓	✓	✓	✓	-
14	PDO	-	-	✓	✓	✓	✓	✓	-
15	$rain_{t-1}$	-	-	✓	✓	✓	✓	✓	-
16	$rain_{t-2}$	-	-	✓	✓	✓	✓	✓	-
17	$rain_{t-3}$	-	-	✓	✓	✓	✓	✓	-
18	$rain_{t-4}$	-	-	✓	✓	✓	✓	✓	-
19	$rain_{t-5}$	-	-	✓	✓	✓	✓	✓	-
20	$rain_{t-6}$	-	-	✓	✓	✓	✓	✓	-
21	$rain_{t-7}$	-	-	✓	✓	✓	✓	✓	-
22	$rain_{t-8}$	-	-	✓	✓	✓	✓	✓	-
23	$rain_{t-9}$	-	-	✓	✓	✓	✓	✓	-
24	$rain_{t-10}$	-	-	✓	✓	✓	✓	✓	-
25	$rain_{t-11}$	-	-	✓	✓	✓	✓	✓	-
26	$rain_{t-12}$	-	-	✓	✓	✓	✓	✓	-
27	$plag_{t-1}$	-	-	✓	✓	✓	✓	✓	-
28	$plag_{t-12}$	-	-	✓	✓	✓	✓	✓	-

3.2 Evaluation Metrics

Numerical and categorical weather forecasts were the focus of this research. For agricultural domains, both probabilistic and deterministic forecasts can enhance season profitability. Probabilistic forecasts usually determine the probability of having rainfall. In deterministic forecasts, forecasters try to predict the actual amount of rainfall over a specific time and region. Therefore, two types of statistical measurements were used.

3.2.1 Numerical Prediction Evaluation Metrics

Determination of monthly rainfall values is central to this study. To assess the accuracy of the developed models, several statistical measurements that have been widely used in regression problems were calculated.

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Pearson Correlation (r)
- Determination of Coefficient (R^2).

In addition, in order to calculate the skill of each model, two measurements were used.

- Skill Score (SS)
- Ideal Point Error (IPE)

Mean Absolute Error (MAE): MAE is used to measure the difference between actual values and forecasts. MAE measures the average magnitude of the error of a set of forecasts without considering their direction. It is shown in the following equation:

$$MAE = \frac{1}{n}\sum_{i}^{n}|y_i' - y_i| \tag{3.2}$$

y_i is the actual value, y_i' is the predicted value, and n is the number of elements in the dataset, $1 < i \leq n$.

Root Mean Square Error (RMSE): RMSE is also a statistical technique that is used to measure the difference between observed and forecasted rainfall in terms of quantity. RMSE is the square root of the average square difference between actual rainfall and forecasts. MAE and RMSE are represented as positive real numbers, where the closer to zero the better the predictions. It is shown in the following equation.

$$RMSE = \sqrt{\frac{\sum_i^n (y_i' - y_i)^2}{n}} \tag{3.3}$$

y_i is the actual value, y_i' is the predicted value, and n is the number of elements in the dataset, $1 < i \leq n$.

Pearson Correlation (r): r is a statistical measure that indicates the extent to which two or more variables fluctuate together. It measures linear interdependence between actual values and outlooks. Correlation usually varies between -1 and +1, where zero means no correlation, +1 means high positive correlation and -1 high negative correlation. It is shown in the following equation.

$$r = \frac{n(\sum_i^n y'y) - (\sum_i^n y')(\sum_i^n y)}{\sqrt{(n\sum_i^n y^2 - (\sum_i^n y)^2)\left((n\sum_i^n y'^2 - (\sum_i^n y')^2)\right)}} \tag{3.4}$$

y_i is the actual value, y_i' is the predicted value, and n is the number of elements in the dataset, $1 < i \leq n$.

Determination of Coefficient (R^2): R^2 is the square of r, and its values range between 0 and 1. The closer the calculated value to 1, the better the performance of a model. It is shown in the following equation.

$$R^2 = \left(\frac{n(\sum_i^n y'y) - (\sum_i^n y')(\sum_i^n y)}{\sqrt{(n\sum_i^n y^2 - (\sum_i^n y)^2)\left((n\sum_i^n y'^2 - (\sum_i^n y')^2)\right)}}\right)^{\frac{1}{2}} \tag{3.5}$$

Skill Score (SS): In order to determine the skills of the proposed model against other approaches, the SS was calculated. SS returns the percentage improvement against reference models [129]. SS is applied to calculate the skill of the proposed approach against a reference model. A value lower than zero indicates that the model's performance is lower than the reference model. A value higher than zero shows the better skill of the proposed approach against the reference model. Zero means the same performance as reference model. SS measurement is shown in this equation.

$$skill\ score = 100 * \frac{(SM_{ref} - SM_{prop})}{(SM_{perf} - SM_{ref})} \tag{3.6}$$

SM is a statistical measurement that can be MAE, RMSE, r, R^2, or IPE. SM_{ref} represents the performance of the reference model, SM_{perf} is the perfect model (0 in rainfall), and SM_{prop} shows the examined model. For example, in order to measure the SS using RMSE statistical measurement, the following equation applies.

$$SS = \frac{RMSE - RMSE_{ref}}{RMSE_{perf} - RMSE_{ref}} * 100\% \tag{3.7}$$

$RMSE$ is the performance of the proposed model, $RMSE_{ref}$ is the root mean square value of the reference model, and $RMSE_{perf}$ is the perfect forecast of 0 (actual rainfall values).

Ideal Point Error (IPE): IPE is a statistical measurement that can be used to assess the overall performance of multiple approaches [130]. IPE combines multiple statistical measurements computed over the same dataset to release a single assessment value. Three out of the four measurements constituted the IPE in this research. The main reason behind not using r values is because of its similarity to R^2. IPE is shown in the following formula.

$$IPE_i = \left(0.33\left(\left(\frac{MAE_i}{max_1^t(MAE)}\right)^2 + \left(\frac{RMSE_i}{max_1^t(RMSE)}\right)^2 + \left(\frac{R_i^2 - 1}{min_1^t(R^2) - 1}\right)^2\right)\right)^{\frac{1}{2}} \tag{3.8}$$

i represents a forecasting model, $1 \leq i \leq t$. t is the total number of models to be compared.

3.2.2 Classification Evaluation Metrics

To assess the performance of prediction models in classification tasks, accuracy over the testing dataset was measured. In addition, mean F-score was calculated to test significance of the models. F-score ranges between 0 and 1, where 1 means that classification is perfect. It was calculated using the following formulas:

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives} \tag{3.9}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives} \tag{3.10}$$

$$Fscore = \frac{2 * Precision * Recall}{Precision + Recall} \tag{3.11}$$

$$meanFscore = \frac{1}{n}\sum_{i=1}^{n} Fscore_i \qquad (3.12)$$

n is the number of classes in the classification task.

3.3 Chapter Summary

This chapter summarises the data and evaluation metrics used in this research. The first part of this chapter described the data used in this research. The sources used to collect weather variables related to this research were listed. Then, a description about each variable was given. The processes followed to create the datasets were discussed. In the second part of this chapter, the statistical measurements that were used to assess the performance of the proposed approaches and comparison approaches were defined.

Chapter 4 Climate Input Features Selection

This chapter presents two climate features selection techniques for designing rainfall forecasting models using single ANNs. The two approaches utilize EAs to select climate input features for numerical and categorical prediction. Initially, the whole available dataset is used to predict or classify rainfall for the next duration. Following the proposed methods, portions of the available input features are selected when designing the single ANN.

The chapter is divided into four sections. Section 4.1 introduces feature selection, and its applicability for rainfall forecasting. Section 4.2 discusses a new method for selecting climate features using GAs for numerical weather forecasts. This method was applied to a numerical prediction problem. Section 4.3 details another climate feature selection approach based on PSO. The second approach was applied to a classification problem in which forecasts were released for the next month. Conclusions are drawn in the final Section 4.4.

4.1 Introduction

Precise details about rainfall over a specific area at a specific time can be valuable for various types of industries. Accurate forecasts enhance decision-making and management in almost all aspects of human life.

Different types of forecasting models have been developed and employed to predict precipitation for different durations. Various models have shown their applicability for rainfall forecasting around the world, and different climate features have been used in

Parts of this chapter appeared in A. Haidar and B. Verma, "Monthly rainfall categorization based on optimized features and neural network," in *AI 2017: Advances in Artificial Intelligence* (Lecture Notes in Computer Science, no. 10400), W. Peng, D. Alahakoon, and X. Li, Eds. Cham, Springer, 2017, pp. 208-220.

A. Haidar and B. Verma, "A Genetic Algorithm based Feature Selection Approach for Rainfall Forecasting in Sugarcane Areas," IEEE Symposium Series on Computational Intelligence (SSCI), 2016, pp. 1-8.

weather forecasting problems. Appropriate feature data are essential if forecasting systems are to perform well. ANN accuracy varies in accordance with the selection of multiple elements, including network topology, number of layers, number of neurons in hidden layer, input features, etc.

Different types of climate attributes and indices have been applied in rainfall forecasting. Various number of weather parameters are typically collected and added as input to ANN models. At some locations only rainfall amounts were available [69], while in other locations authors applied some or all of 26 features to predict weather attributes such as sea level pressure, wind height, and direction [99]. Nagahamulla, Ratnayake, and Ratnaweera used Nino 3.4, wind index, Ocean-Land Temperature Contrast (OLTC), and rainfall to predict monsoon rainfall over India [70]. He, Guan, Zhang, and Simmons used SOI, PDO, SAM, and IOD to forecast the South Australian monthly rainfall anomaly [96]. Deo and Sahin used 13 different attributes and a learning approach to predict monthly values for the Effective Drought Index (EDI) in Eastern Australia [131]. These attributes were categorized into site-specific and climate variables. The dataset was composed of year, month, latitude, longitude elevation, monthly mean rainfall, monthly mean temperature, monthly maximum temperature, monthly mean air temperature, SOI, PDO, SAM, and IOD [131].

Climate attributes have been applied not only for rainfall prediction, but also for agricultural yields prediction. Everingham, Muchow, Stone, and Coomans used SOI to forecast sugarcane yields for Northern Australia [132]. Everingham, Clark, and van Gorder have also used SOI to generate a long lead rainfall prediction model that aids sugarcane industry decision-making in several locations across Eastern Australia [103]. The addition of these climate attributes into different studies was considered when setting up the dataset in this study.

Using all possible features may not facilitate best performance. In addition, inclusion of some climate features to assist prediction may lower performance. Therefore, there is a need to use feature selection in order to choose the best input features for predicting with highest accuracy. EAs, including GA and PSO, can be applied to select a subset from all possible variables with the aim of achieving best performance. These algorithms have been deployed in various forecasting applications. Kishtawal, Basu, Patadia, and Thapliyal applied a GA to find the best parameters for an equation used for forecasting summer rainfall in India [102]. Nasseri, Asghari, and Ebedini used a GA to select ANN weights to predict hourly rainfall for a selected location in Sydney, Australia [63].

Data are essential, if an ANN is to perform optimally. In some locations, recorded data are rare, and this affects the capability of evolving models. For monthly rainfall forecasting, most of the logged data from stations around the world started at the beginning of 19th century. Therefore, using each month's dataset to forecast may be insufficient. Also, because of faults and machine errors, information could be found to be incomplete. However, it has been shown that ANN based models can even perform with such obstacles [133]. Different climate attributes, including previous local and global variables have been generally collected to illustrate rainfall uncertainty. Following the butterfly effect, a variability in any part of the world can have a consequence on another part of the world [134]. Because of this phenomenon, uncertainty in weather conditions on land and oceans can have an associated effect on rainfall amounts across Australia. Extensive research has tried to link these climate indices to rainfall variability in Australia. For this reason, some of the global measurements of weather conditions were gathered and studied.

Rainfall forecasting models can be deployed to forecast for different periods including hourly, daily, monthly, annually, etc. These models reveal different types of information, including actual amount of rainfall, chance of rainfall, or probability of rainfall based on a certain threshold (exceeding a value). Each of these models has its own effectiveness in terms of prediction.

All of the available input features can be used as possible predictors. These variables may affect certain locations, but have no effect on other locations. At a deeper level, these variables may disturb rainfall trend in some months in the same location, but have no effect in other months. Furthermore, the combination of two or more indices may lower the accuracy of the network. Therefore, in this research, feature selection is applied when developing a rainfall forecasting model. Hence, this chapter investigates the use of two novel selection approaches for forecasting and classifying monthly rainfall respectively. The first study in Section 4.2 employs a GA and an ANN to predict monthly rainfall for twelve months lead time. The second study in Section 4.3 utilizes a PSO algorithm and an ANN to classify monthly rainfall for one month lead time.

4.2 A Genetic Algorithm based Feature Selection Approach for Rainfall Forecasting in Sugarcane Areas

Different types of climate indices and attributes are usually applied to model rainfall forecasting systems. In this study, we present a novel GA based feature selection approach to determine which climate indices and attributes are most significant for numerical rainfall forecasting in a specified location. The most significant features are features that return the highest accuracy for rainfall forecasting through ANNs. The approach is evaluated on real-world data that contain different weather forecasting features. A set that contains maximum temperature values and SOI proved to be the best combination for the selected location among the other models with a Root Mean Square Error (RMSE) of 0.027 in November. An Average RMSE of 0.0638 for the GA based forecasts was recorded. The proposed model was compared to other models, and the proposed model obtained higher accuracy in forecasting monthly rainfall.

4.2.1 Proposed Approach

This study proposes an approach for predicting rainfall in locally specified areas. A GA was utilized to select the best input features for generating outlooks, while an ANN was created to assess the features and to produce the forecasts. The proposed approach is shown in Figure 4.1.

The GA is a meta heuristic algorithm that belongs to the class of EAs [135]. It has been widely used in the areas of selection and optimisation. GA consists of multiple elements called chromosomes. These chromosomes are grouped into a population. GA mimics the natural evolution process in which chromosomes in a population interact together in a different number of iterations. [136]. Each chromosome has a scalar value that represents its performance against the problem being applied. This scalar value is determined by the fitness function. Each chromosome's fitness function is calculated to determine the best chromosome in the iteration. Then, these chromosomes interact together to form a new population. Ultimately, the chromosomes that have the best fitness function will be traversed directly into the next iteration without being manipulated. This process is called elitism. Some chromosomes are then selected to be used as parents of the next population. Different selection criteria are applied to select parents. Then, crossover and mutation are deployed over the selected parents to create a new population. Different types of crossover and mutation methods have been proposed. These methods are usually

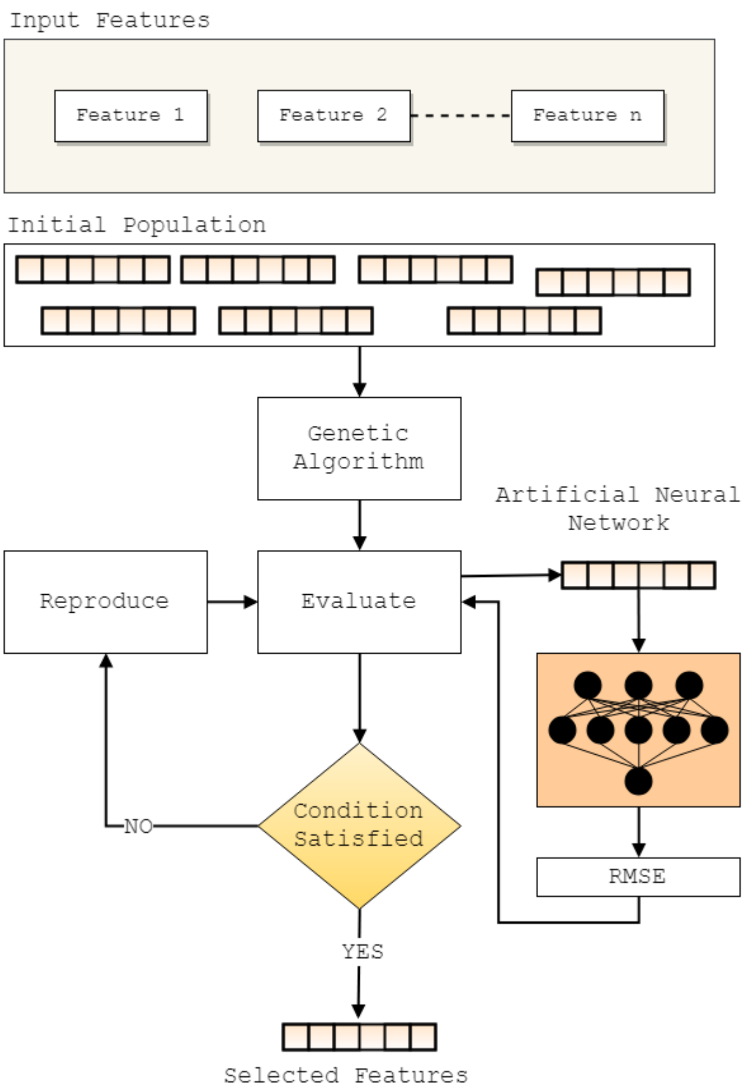

Figure 4.1 Proposed model.

determined when setting up the GA. Crossover is a process of combining two individuals so as to create a new offspring. Mutation is the mechanism of tweaking elements of a chromosome after crossover has occurred. Each of these processes has its own probability, which is based on the number of chromosomes in a population without the chromosomes selected for elitism. The main purpose of GA is to find the best optimal solution through consecutive iterations. Several options control the evolution process, including finding the optimal solution, reaching the maximum number of iterations, or reaching the maximum number of iterations without an update on the best recorded performance.

A GA was designed and incorporated into the proposed approach to select the best features from a given dataset that would enhance the performance of the forecasting model. Feature selection is the mechanism of identifying a subset from the whole dataset that generates the best optimal solution [137]. Feature selection is a mapping of a set of features into a smaller set so as to produce higher results for a specific problem. An initial population is supplied to the GA, then population members are subjected to evolutionary processes. The GA for the proposed model is represented in Figure 4.2. It has been added into the study to investigate its ability in selecting the optimal combination of climate attributes to ensure highest accuracy in forecasting monthly rainfall.

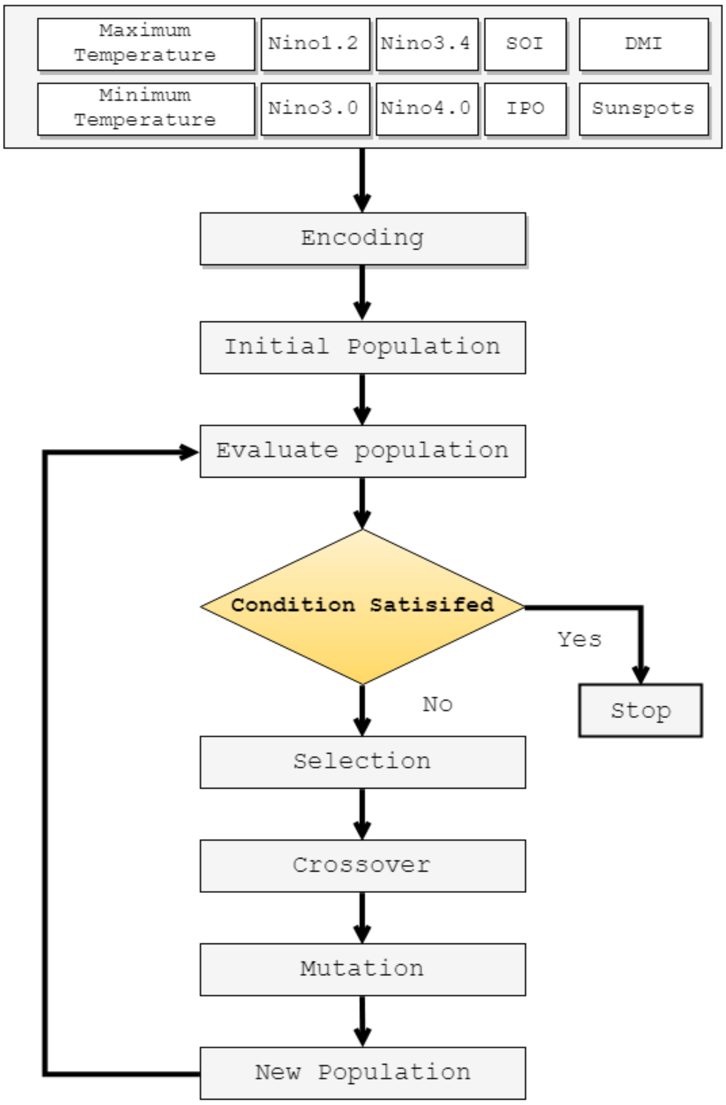

Figure 4.2 Genetic Algorithm.

A FFNN was combined with a GA and incorporated into the proposed model. It is formed of interconnected processing elements called neurons that process information by learning. The interconnected nodes are organised into layers. Three types of layers are in the structure of the proposed ANN: input layer, hidden layer, and output layer. The feed-forward architecture used in this study is shown in Figure 4.3. The input layer receives features, performs calculations, and sends results to the hidden layer where information manipulation takes place. The hidden layer processes the information and sends to the output layer that returns the network results. The ANN learns to generalize data through the training process, in which weights and connections between layers are modified in order to obtain desired values. With weather forecasting problems, the ANN minimizes the error between actual and predicted values so as to increase the performance. Particular learning algorithms that hold different mechanisms are usually used with an ANN. The ANN can be taught the dynamics of the system which would lead to improvements in overall approximation accuracy of the outlooks [134]. The output of the ANN can be completely different when a small part of its parameters is changed.

Data is essential for several forecasting models, including ANN. Discrete weather attributes are usually collected in an attempt to set up a model for prediction using ANNs. Climate indices and attributes represent a specific situation on land or in the oceans. The formation of rainfall in a specific location can be related to climatic events in different parts of the globe. With the vast spread of technology throughout the 20th century, the ability to record and save climate variables became much easier. This abundance of data revealed a new and challenging task, which is the selection of climate features that are highly conducive to being incorporated when generating accurate conjectures for a specific location. In order to specify the climate attributes to be added within a model, close geographic indices to the targeted area can be chosen. But, a variation at one corner of the globe may produce a tornado in another place that is geographically far away (butterfly effect) [134]. Because of this issue, there is still a need to look at global climate indices. Increasing the dataset size increases processing time. Typically, an ANN performs better with larger dataset, but appending some features may return a lower performance. The high dimensionality in data may affect the performance [137]. Hence, climate attributes should be chosen carefully so as to ensure accurate rainfall forecasts. To select the optimal input data (e.g., features) that would result in the highest accuracy, a trial and error based manual method can be used. In trial and error, different input data are formed, based on user preferences, and added to the network. In addition, diverse computational techniques, such as PSO, K-mean clustering, Principle Component Analysis (PCA), and GAs, can be utilized to select the optimal subset. For this research, the GA will search for the optimal subset.

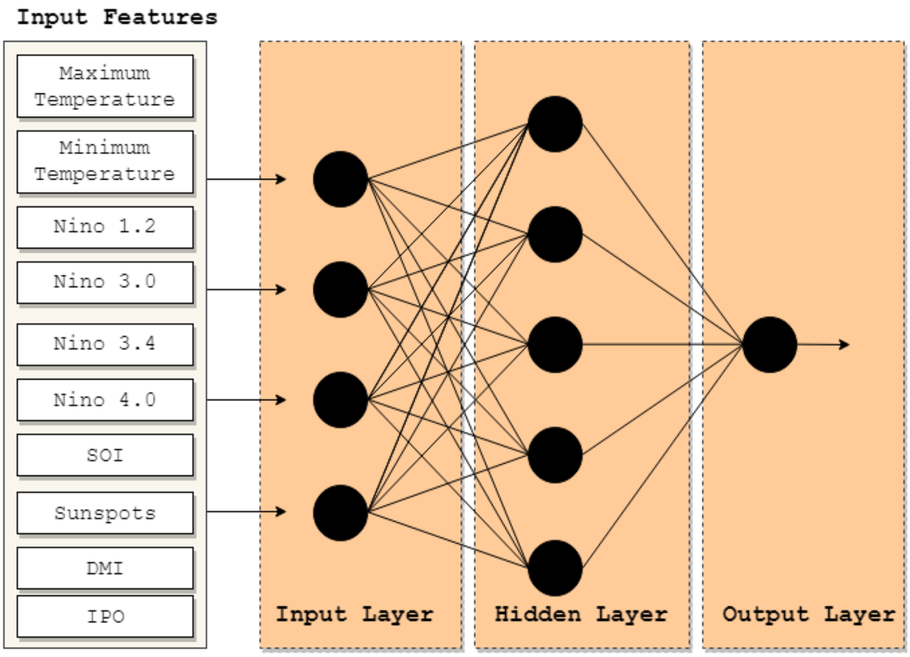

Figure 4.3 Three layered feed forward neural network architecture used in the study.

4.2.2 Data (Case Study 1)

The study area selected was Bingera, a town located in Queensland, Australia. Bingera has an annual rainfall average of 1024 mms. Multiple weather stations are located in the area. Local and global climatic attributes were collected to set up the dataset for Bingera. A total of ten input features which were gathered and used in the study are shown in Table 3.5. About 115 years were used in the study. Description and pre-processing of input features is shown in Chapter 3.

4.2.3 Experiments and Results

4.2.3.1 Experimental Setup

A GA models natural evolution. Its population consists of a number of elements called chromosomes. A chromosome is composed of genes (climate features) that represent a possible solution for the problem that the GA is trying to solve. Each month has its own predictors. Therefore for each month, the same GA was deployed. The following steps were followed to setup the proposed GA.

Encoding

Encoding is a mechanism whereby intended features are mapped into chromosomes. As shown in Table 3.5, ten climate features were collected to set up this experiment. The chromosome type was selected as a binary string, so each possible gene can have a value of either 0 or 1 only. If the gene is included in the chromosome, 1 will be found in its index, otherwise 0. Since there are ten input features, chromosomes consisted of ten genes (each gene equals 1 index). The chromosome highly depends on the sequence of features being selected. The index of each climate feature is shown in Figure 4.4.

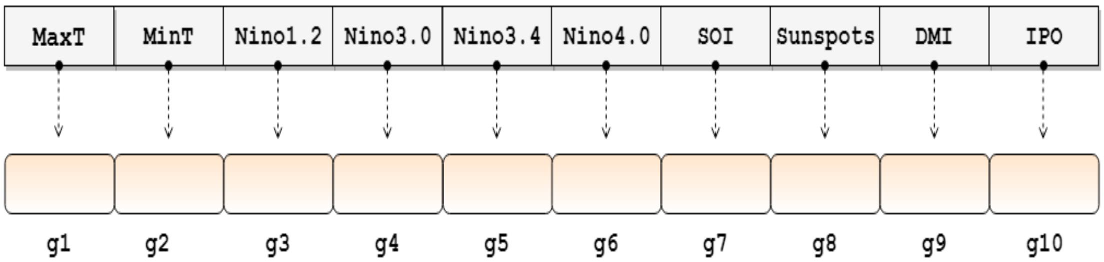

Figure 4.4 Encoding of climate features into binary chromosome.

Initial Population

To ensure that multiple combinations are formulated through generations, the initial population should contain some chromosomes that are formed of two or more genes. Hence, a random initial population was generated.

Evolution of Chromosomes

A. ***Selection***: Selection is the process of choosing individuals that are used in reproduction. This study has been set to four. Four chromosomes out of the generated population were used for generating the next population.

B. ***Crossover:*** Crossover is an operator for the GA on which performance is highly dependent. Two chromosomes are combined to generate a new offspring that can achieve better results. Through crossover, diverse combinations can be created which will assist in finding the best solution.

C. ***Mutation***: Mutation includes random transforming of values of genes in the chromosome. The combination of different genes may generate better forecasts. Mutation probability refers to the chance that a gene in the chromosome can be flipped. Through this technique, removing a feature from the dataset may enhance the performance (transforming 1 into 0). On the other hand, adding a new feature to a given dataset may generate a better chromosome (transforming 0 into 1).

D. ***Fitness Function:*** Fitness Function is used in each generation by the GA to evaluate the performance of the chromosomes. It returns a scalar value that determines the effectiveness of the chromosome (collection of features). A fitness function consisting of a FFNN with three layers was proposed. Based on the size of the dataset, the number of hidden neurons in the hidden layers was selected. The number of neurons should be proportional to the data size so as to avoid low accuracy. A gradient descent based BP algorithm was considered and used as the training algorithm for the proposed network. Tansig activation functions were used between input to hidden and hidden to output layers. The data were partitioned into training, validation and testing as mentioned in Chapter 3 (75% training, 15% validation and 10% testing).

To measure the fitness of each chromosome, RMSE was calculated. RMSE is a positive number; the closer the number to zero, the better the performance. In other words, the closer the number to zero, the better are the forecasts. Fitness function return value was selected as the RMSE measurement of the developed ANN. The aim of the study was to get the best features for forecasting rainfall. Setting the initial weights randomly each time in order to measure the fitness function would have generated different results for the best chromosome. Therefore, ANN connection weights and bias connections were generated from the same set of random values.

4.2.3.2 Experimental Results

Experiments were conducted using MATLAB. Monthly rainfall values were selected as the target, while the remaining ten features were used as predicators. For the selected area, monthly rainfalls varied between different seasons and months. Rainfall ranges fluctuate between those months. Therefore, each month may have different climate attributes that affect the formation of rain. The SOI, which is a climate attribute (gene), may be paramount for forecasting January rainfall, but not for July. Therefore, the dataset was divided into 12 months to find the optimal subset for each month. Ten climate features were used to predict rainfall. In total, 1024 distinct datasets can be generated for forecasting. To find the best combination for each month, the proposed approach was utilized.

As discussed in Section 4.2.3, various steps were followed to setup the GA. The initial population contained 50 chromosomes that were randomly created, ensuring that all genes are not zero. The number of iterations was set to 20. Hence, a maximum value of 1000 networks could be tested (not including the first 50 for fitness). The aim was to increase the initial population so that the algorithm could start identifying the best genes in earlier generations, while avoiding the local minimum. Trapping in the local minimum could be encountered when the algorithm begins to check closely similar features. Two points crossover when two chromosomes are combined by selecting genes from the second parent and are added to the same location (index) as the parents. Uniform Mutation was selected with a probability of 0.3. Table 4.1 summarizes the GA specifications.

Table 4.1 Genetic algorithm parameters.

Parameter	**Value**
Population	50
Population Type	Bit String
Generations	20
Crossover	Crossover Two Points
Mutation	Uniform Mutation
Fitness function	FFNN
Mutation Probability	0.3
Selection	Selection tournament (size 4)

The stopping criterion selected was the number of generations (20). The fitness function was selected to be a FFNN. Some 13 neurons were added to the hidden layer. The number of epochs for the network was 500. Table 4.2 summarizes the ANN specifications.

Table 4.2 Neural network specifications.

Attribute	Value
Type	Feed Forward Neural Network
Layers	Three layers
Neurons in hidden layer	13
Activation function	Tansig
Training Algorithm	Gradient descent
Epochs	500
Training ratio	85%
Validation ratio	15%

Each month's dataset was used by the GA to find the best subset of the whole dataset. Table 4.3 shows the best selected subset for each set. The largest subset selected (in terms of number of features) was in February with nine features. Error ranges varied between 0.027 and 0.181 in all of the months. Maximum temperature, Nino 3.0, and SOI were selected in seven of the 12 months (Table 4.4).

Table 4.3 Results obtained using the proposed approach.

Month	No. of Attributes	RMSE	MaxT	MinT	Nino 1.2	Nino 3.0	Nino 3.4	Nino 4.0	SOI	Sunspots	DMI	IPO
January	2	0.1813	-	-	-	✓	-	-	✓	-	-	
February	9	0.0890	✓	✓	✓	✓	✓	✓	-	✓	✓	✓
March	4	0.0863	-	-	✓	✓	-	-	-	✓	✓	-
April	3	0.0387	✓	✓	-	-	✓	-	-	-	-	-
May	4	0.0341	✓	-	-	✓	-	-	-	✓	✓	-
June	3	0.0459	✓	-	-	-	-	✓	✓	-	-	-
July	3	0.0300	✓	-	-	✓	-	✓	-	-	-	-
August	8	0.0344	-	✓	✓	-	✓	✓	✓	✓	✓	✓
September	3	0.0346	-	-	-	-		✓	✓	✓	-	-
October	5	0.0342	-	✓	-	✓	-	-	✓	✓	-	✓
November	2	0.0270	✓	-	-	-	-	-	✓	-	-	-
December	5	0.1311	✓	✓	✓	✓	-	-	✓	-	-	-

Table 4.4 Number of times each feature was in best subset.

Feature	Gene index in chromosome	Number of times used in selected subsets
MaxT	1	**7**
MinT	2	5
Nino 1.2	3	4
Nino 3.0	4	**7**
Nino 3.4	5	3
Nino 4.0	6	5
SOI	7	**7**
Sunspots	8	6
DMI	9	4
IPO	10	3

Nino 3.4 and IPO were used in three of the 12 subsets. Some months needed no time to capture the best solution while, in other months, multiple iterations were required to enhance the optimization. These months hold low average monthly values compared to others. In February, the GA fluctuated through most of the iterations, while in July and August results were found from early iterations. Using 50 individuals in the initial population allowed the network to optimize earlier.

4.2.4 Comparative Analysis

To compare the proposed GA based forecasting model, another 12 ANNs were developed that have the same specifications that the network in fitness function had. For comparison purposes, the same set of network parameters was used. A FFNN with three layers and 13 neurons in the hidden layer was designed.

A gradient descent training algorithm was designated to train the networks. The only difference was the number of inputs. For each month, the whole dataset was trained and tested. Networks that used all of the features as input features were compared with networks that used the GA as a feature selection tool (as in the proposed approach). In order to measure the accuracy, RMSE over the testing set was calculated. The results of the two models are shown in Table 4.5. RMSE for all feature-based networks varied between 0.039 in November and 0.2354 in February. The RMSE of the network that used all of the features exceeded the GA based approach in all of the months. The highest difference in performance was recorded in both February and December, with a difference of 133.58 and 73.56 mms respectively. The average error of GA based forecasts was reported as 0.0638 (58.35 mms). This proves that using appropriate climate attributes to forecast could produce higher performance than using all of the available climate attributes.

Table 4.5 Comparison of results in terms of RMSE.

Month	**RMSE (All Features)**	**RMSE (Proposed Approach)**
January	0.2301	0.1813
February	0.2354	0.0890
March	0.1603	0.0863
April	0.0711	0.0387
May	0.0526	0.0341
June	0.0904	0.0459
July	0.0591	0.0300
August	0.0598	0.0344
September	0.0506	0.0346
October	0.0474	0.0342
November	0.0393	0.0270
December	0.2115	0.1311
Average	0.1089	0.0638

4.2.5 Summary

In this study, a GA based approach was used to select the best features for monthly rainfall forecasting in eastern Australia. Many experiments were conducted and, over time, improved results were obtained. The results revealed that, for most of the months, several climate attributes – maximum temperature values, Nino3.0, and SOI – were essential. There was no need to combine all the features, since some of them resulted in lower performance when added to the classifier. The results showed that a set consisting of maximum temperature values and SOI achieved the highest accuracy (for the November forecasts). In the next study (Section 4.3), this research was extended by including data from more locations near agricultural areas and new climate variables in order to perform a classification task.

4.3 Monthly Rainfall Categorization based on Optimized Features and Neural Network

In this study, a feature selection approach was used to classify monthly rainfall. Rainfall is classified into categories based on the amount of rainfall. Five distinct locations were selected to perform the study: Innisfail, Plane Creek, Bingera, and Maryborough in Queensland, Australia, and Yamba in New South Wales, Australia. Multiple local and global climate indices have been linked to formation of rain. Hence, different local and global climate indices were assessed as possible predictors of rain. A PSO algorithm was incorporated to select the best features for each month in each location. Using this approach, an average accuracy of 87.65% was recorded over the five selected locations. The developed models were compared to other ANN models where all features were used as input features. An average difference of 25.00%, 23.89%, 24.02%, 20.00%, 20.59% was recorded for Innisfail, Plane Creek, Bingera, Maryborough, and Yamba respectively. Analysis of statistical results suggested that ANNs is a promising alternative approach for rainfall categorization over multiple weather zones and over Australia. In addition, it was found that selection of input features should be carefully considered when designing rainfall forecasting models.

4.3.1 Proposed Approach

The proposed approach is based on selecting climate features for predicting rainfall categories. The selection is based on a PSO algorithm, and the classifications are based on an ANN. The proposed approach is shown in Figure 4.5.

PSO is population-based algorithm that mimics social behaviours such as bird flocks and fish schooling [78]. PSO was firstly introduced by Kennedy and Eberhart in 1995 [138]. It shares the same characteristics as GA in that it uses population with multiple elements [136]. Each population consists of multiple elements called particles, where each particle has two key components: position and velocity. Each particle in the search space denotes a possible solution [23]. Similarly, PSO particles are evaluated using a fitness function. Each particle position in the search space is influenced by its previous position ($pbest$) and global best particle position ($gbest$). Each particle in a population acts as a possible solution [139]. Based on the findings, the elements of the population are tweaked (direct mutation) [135]. Particles are then updated over generations to search for the optimum solution [140]. The final solution is typically achieved after running the PSO for a

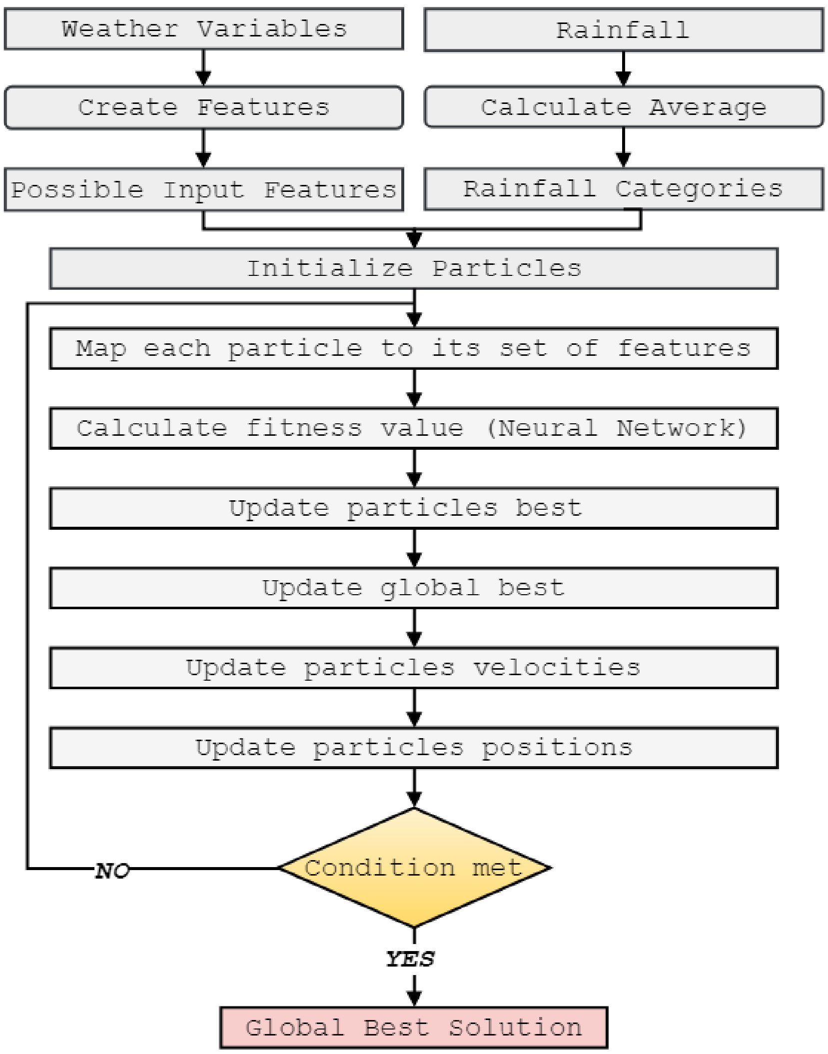

Figure 4.5 Proposed method.

specified number of iterations. A particle's velocity and position are updated in the PSO search space using the following formulas.

$$vel_j^{t+1} = wvel_j^t + c_1 r_1 \left(pbest_j^t - pos_j^t\right) + c_2 r_2 (gbest^t - pos_j^t) \quad (4.1)$$

$$pos_j^{t+1} = pos_j^t + vel_j^{t+1} \quad (4.2)$$

j characterizes a particle in the search space, c_1 and c_2 represent the acceleration coefficients, r_1 and r_2 are two random numbers, w is the inertial coefficient, vel_j^t represents the velocity of the j^{th} particle at iteration t, pos_j^t represents position of the j^{th} particle at iteration t, $pbest_j^t$ represents the j^{th} particle best position, $gbest^t$ represents the global best at iteration t, and t is the iteration number.

Three steps are usually followed in each PSO iteration. First, the fitness function for each particle is calculated. Second, the local best of each particle and the global best of the search space are then updated. Finally, the velocity and the position of each particle in the iteration are updated. This process continues until reaching the termination condition.

Rainfall classification is described as rainfall categorization into different ranges. In this study, rainfall amounts were categorized into values lower or higher than monthly average. Official rainfall forecasts are generated based on large spatial areas (≥ 250 km grid areas with POAMA), while in the following method we target monthly rainfall for a specific location.

Rainfall values are mapped into categories based on each dataset average. These values were mapped into binary classes to represent the ANN target as shown in the following equation.

$$C(x) = \begin{cases} [0\ 1], & x < z \\ [1\ 0], & x \geq z \end{cases} \tag{4.3}$$

where $z = \frac{\sum_{i=1}^{n} x_i}{n}$, n is the sum of instances in the dataset (training/validation), and C is the mapping function of numerical values to classes.

ANN input plays a key role in determining the overall accuracy, especially in rainfall prediction problems. All of the climate indices listed in Section 2 are intended to be used as possible predictors for classifying monthly rainfall over selected locations. Developed models learn from previous historical features, and use these indices to classify rainfall categories. Rainfall lagged values, which are the previous amounts of rainfall up to one year, were incorporated in an attempt to enhance rainfall categorization. Some 12 features were created for each target value from previous monthly rainfall observations, starting from observation in the preceding month up to one year.

In addition, the first lagged monthly rainfall value ($rain_{t-1}$) and same month value in previous year ($rain_{t-12}$) are considered to be the most significant in comparison with the other ten for a rainfall value at time t ($rain_t$). Hence, two new features are proposed following those two lagged values ($plag_{t-1}$, $plag_{t-12}$). The monthly rainfall amount average for all the years was calculated, then binary values (0,1) were used to identify the occurrence of higher than average rainfall or not.

Since recent values of monthly rainfall ($rain_{t-1}$) were considered as possible input features, and one-month ahead forecasts were released. This means that in the beginning of each month, forecasts were made. Extensive research was done to identify rainfall influencers over Australia, in general, and over eastern Australia, in particular. Climate indices are typically considered to be significant over multiple locations at different periods of the year. Therefore, a climate attribute may have high correlation in determining rainfall category for one month, and have no linkage to categorization for another month in the same location. Hence, a PSO algorithm was utilized in order to select the best rainfall category predictors for each month in each specific location. Each particle is represented as a collection of input features, and the fitness function was selected to an ANN.

In previous literature, climate indices were compared to rainfall amounts and correlations were created in order to understand which has an effect on precipitation. We use climate variables at time t to categorize rainfall at time t+1, so that if reasonable accuracy was recorded, the model can be deployed for rainfall categorization. The following equation represents rainfall classification with all of the available features.

$$y_t = f\begin{pmatrix} rain_{t-1}, rain_{t-2}, \dots, rain_{t-12}, plag_{t-1}, plag_{t-12} \\ , input_{t-12}^{1}, input_{t-12}^{2}, input_{t-12}^{3}, \dots, input_{t-12}^{n} \end{pmatrix} \quad (4.4)$$

y_t is rain range at time t, $rain_{t-z}$ is a lagged value for rainfall at time t where $1 \leq z \leq 12$, plag_{t-1} and plag_{t-12} are the proposed new features, $input_t^i$ is a climate variable value at time t, $1 \leq i \leq n$, n is the number of available climate variables, f is the trained ANN model.

4.3.2 Data (Case Study 2)

Five locations were used to conduct this study. The dataset of each location was modified so that it had the same start and finish time. Dataset durations are shown in Table 3.1, where last column represents annual average for all recorded years. Then, each collected dataset for each location was divided into 12 datasets, each representing one month. Rainfall ranges were created based on monthly overall averages over the training/validation dataset. Two ranges were initiated: below or higher than average ([0 1], [1 0]).

4.3.3 Experiments and Results

MATLAB was used to create all of the runs. PSO algorithm parameters are shown in Table 4.6. Population size was selected to ten particles. Stopping conditions were either 50 iterations or 100% accuracy over the testing dataset. Particles were randomly initialized in each monthly run. Each particle consisted of a binary number with 28 digits. Each index represented an input feature, where 1 meant inclusion of the feature into input dataset, and 0 meant exclusion of a feature into dataset.

Table 4.6 Particle swarm optimization algorithm used parameters.

Parameter	Value
Population	10
Iterations	50
Fitness function	Three-layered feed forward network
Scalar value	Accuracy error over test dataset

The selected objective function for the PSO algorithm consisted of a three-layered FFNN with scaled-conjugate BP as the training algorithm, 13 neurons in the hidden layer, hyperbolic tangent as the activation function over the hidden layer, and softmax as activation function over output layer. The trained model that showed the lowest error in each month in each location was saved. The scalar value of the objective function was the accuracy error. Hence, particles traverse the search space to find the combination of input features that would result in the lowest error over the testing dataset.

Each dataset was divided into 70% training, 15% validation, and 15% for testing. Usually to measure classification accuracy, cross validation is applied. With cross validation, the dataset is partitioned into subsets in which each subset is used once as a testing dataset. Because of the nature of the problem specified here, and the linkage between dataset instances, the latest 15% of each dataset's instances were taken for measuring the accuracy of the proposed model over each month. The testing datasets' sizes varied at each location since different historical values were recorded.

The proposed approach was applied 60 times (once for each month in each location). The calculated accuracies for each month are shown in Table 4.7. Categorization accuracy varied between 70.59% as the lowest (December, Yamba) and 100% accuracy as the highest in multiple months (July, Innisfail; August, Innisfail; September, Innisfail; January, Bingera; May, Maryborough). Innisfail had the highest annual average and the highest accuracy average with 91.15% through all months, followed by Bingera, Plane Creek, Yamba, and Maryborough. An overall average of 87.89% accuracy was recorded over the 60 months in the selected locations. Similar performance was obtained for the Plane Creek months, even though different features were used.

The datasets consisted of 28 possible input features: rainfall lagged values, two proposed new features, maximum and minimum temperatures, and climate indices. The selected features using PSO for each month varied between seven and 19. The highest feature used was Nino 3.0, which was selected in 66.67% of the months (40 times), which is similar to the study discussed in Section 4.2. The highest features used were: Innisfail—Nino 3.0 and plag-12 (were used for 9 months); Plane Creek—Nino 3.0 (for 10 months); Binger—Nino 3.0 and rain-3 (9 months); Maryboroug—plag-1, DMI, and sunspots (9 months); and Yamba—DMI and rain-12 (9 months). Nino 3.0 was used as possible predictor in all of the February classifications. The two created features that were proposed were selected in 58.33% and 56.67% of all the runs. The variations in the selected features for each month imply that feature selection is essential for forecasting monthly rainfall ranges.

Table 4.7 Categorization accuracy obtained for each month in each selected location using the proposed approach.

Location	**Innisfail**	**Plane Creek**	**Bingera**	**Maryborough**	**Yamba**
January	87.5	86.67	**100**	86.67	88.24
February	93.75	86.67	94.12	86.67	82.35
March	93.75	93.33	88.24	86.67	88.24
April	93.75	86.67	88.24	93.33	82.35
May	87.5	86.67	94.12	**100**	82.35
June	87.5	86.67	76.47	80	94.12
July	**100**	86.67	94.12	93.33	94.12
August	**100**	86.67	88.24	80	88.24
September	**100**	86.67	94.12	86.67	88.24
October	75	86.67	82.35	73.33	88.24
November	87.5	86.67	82.35	80	94.12
December	87.5	93.33	76.47	80	70.59
Average	91.15	87.78	88.24	85.56	86.77

4.3.4 Comparative Analysis

The developed approach was compared to another ANN that uses all weather variables as possible features. No feature selection was incorporated into the second approach. The same specifications were given to the ANN as the selected objective function in the deployed PSO algorithm. For each month, the network was trained multiple times and the best accuracy was recorded. Categorization accuracies of the comparison model are shown in Table 4.8. Using all features resulted in accuracies that varied between 43.75% as the lowest in June in Innisfail, and 88.24% as the highest in September in Bingera.

The differences between the two approaches and their accuracies are shown in Table 4.9. The proposed approach that applied feature selection obtained better accuracy in all months. An average difference of 25%, 23.89%, 24.02%, 20%, and 20.59% were obtained for Innisfail, Plane Creek, Bingera, Maryborough, and Yamba respectively. This demonstrates the ability of feature selection in increasing rainfall classification accuracy. It was noticed that the highest difference between accuracies was recorded in locations with lowest annual averages. However, the use of features should be carefully considered since there is no guarantee that their inclusion will always enhance performance. Nonetheless, ANNs are considered to be promising for revealing rainfall classifications (categorizations) if they are well-designed.

Table 4.8 Categorization accuracy obtained using a comparison approach (all features).

Location	**Innisfail**	**Plane Creek**	**Bingera**	**Maryborough**	**Yamba**
January	62.50	73.33	88.24	66.67	58.82
February	68.75	60.00	64.71	53.33	58.82
March	68.75	73.33	52.94	53.33	58.82
April	75.00	60.00	64.71	73.33	76.47
May	81.25	66.67	76.47	73.33	58.82
June	43.75	60.00	52.94	73.33	52.94
July	62.50	53.33	64.71	86.67	88.24
August	68.75	60.00	64.71	60.00	70.59
September	75.00	66.67	88.24	66.67	82.35
October	62.50	66.67	52.94	53.33	64.71
November	62.50	66.67	47.06	60.00	70.59
December	62.50	60.00	52.94	66.67	52.94
Average	66.15	63.89	64.22	65.56	66.18

Table 4.9 Categorization accuracy difference between the two approaches.

Location	**Innisfail**	**Plane Creek**	**Bingera**	**Maryborough**	**Yamba**
January	25.00	13.33	11.77	20.00	29.41
February	25.00	26.67	29.41	33.33	23.53
March	25.00	20.00	35.29	33.33	29.41
April	18.75	26.67	23.53	20.00	5.88
May	6.25	20.00	17.65	26.67	23.53
June	43.75	26.67	23.53	6.67	41.18
July	37.50	33.33	29.41	6.67	5.88
August	31.25	26.67	23.53	20.00	17.65
September	25.00	20.00	5.88	20.00	5.88
October	12.50	20.00	29.41	20.00	23.53
November	25.00	20.00	35.29	20.00	23.53
December	25.00	33.33	23.53	13.33	17.65
Average	25.00	23.89	24.02	20.00	20.59

To test the significance of the proposed approach, the mean F-score was calculated for each generated model in each month. F-scores range between 0 and 1, where 1 means that the classification is perfect. The average f score for the proposed approach are shown in Table 4.10. The dataset with the highest amount of rainfall produced the best f-score of 0.889.

Table 4.10 F-score measure for each optimization.

Location	**Innisfail**	**Plane Creek**	**Bingera**	**Maryborough**	**Yamba**
January	0.795	0.712	1	0.861	0.871
February	0.937	0.866	0.933	0.861	0.813
March	0.935	0.928	0.879	0.861	0.882
April	0.816	0.85	0.717	0.921	0.813
May	0.833	0.792	0.91	1	0.773
June	0.867	0.712	0.673	0.72	0.94
July	1	0.83	0.91	0.907	0.91
August	1	0.861	0.882	0.785	0.871
September	1	0.792	0.883	0.792	0.798
October	0.733	0.83	0.821	0.732	0.871
November	0.873	0.861	0.821	0.785	0.933
December	0.873	0.921	0.764	0.762	0.702
Average	0.889	0.830	0.849	0.832	0.848

4.3.5 Summary

An ANN based approach was investigated to classify rainfall patterns. A PSO algorithm was incorporated into select features that would expose the highest accuracy for monthly rainfall. The results are promising, and represent an alternative for current official rainfall forecasts. The approach is beneficial in terms of selecting accurately the timing and location of rainfall categories (below/ higher than average). An overall accuracy of 87.89% was obtained for five locations in different weather zones along eastern Australia. The proposed model showed higher accuracy when compared to another ANN where all features were used as predictors.

4.4 Chapter Summary

ANNs have been widely applied in weather forecasting problems. In the studies described in this chapter, ANNs were used to approximate rainfall for local regions in eastern Australia. Feature selection methods were proposed to select climate input features that would reveal the best performance when used in an ANN. It was concluded that for each month in each selected location, feature selection is mandatory. The inclusion of feature selection when developing this type of prediction models is essential if reasonable performance is to be obtained from forecasting models.

Multiple weather stations are located over the Australian continent. In addition, many climate indices have been linked to climate variability for multiple locations. Using feature selection in the proposed approach enhances the applicability of the forecasts. Therefore, these approaches can be applied for developing local weather systems for every weather station in Australia. Data input features are a key part of delivering an accurate prediction model. Throughout the studies described in this chapter, the network parameters were manually selected and kept fixed. These have a significant effect on a model's accuracy. Therefore, in the next chapter, network parameters will be investigated and compared in order to evaluate their impact on model performance.

Chapter 5 Climate Features and Network Parameters Selection

In the previous chapter, it was demonstrated that feature selection using EAs enhances the accuracy of a forecasting model. Various possible predictors were given, within which subsets were selected. This agrees with previous research in climate science that has shown climate indices affect rainfall variability over a specific location through different parts of the year. Hence, for each location and climatic zone, the input features should be optimized.

In this chapter, two new approaches for selecting ANN components, including input features and neural network parameters, are described. Optimization was extended to include the network parameters. In addition, a hybrid GA that incorporates PSO was proposed to enhance the optimization process. Section 5.1 is a brief introduction to the use of climate features and network parameters. Section 5.2 provides a general overview of the selection process. Section 5.3 proposes a new approach for selecting both input features and neural network parameters. Section 5.4 provides details of a new study that extends the previous selection approach by integrating a multi-level optimization strategy. Finally, a summary is given in Section 5.5.

5.1 Introduction

Model design requires the selection of multiple elements that affect overall performance. Machine learning algorithms have been widely used in different real-life applications, including classification and regression [20, 27, 41]. Various hyperparameters are usually examined while training an algorithm. In addition, data used to train the machine learning algorithm are carefully analysed (see Chapter 4).

The performance of an ANN is dependent on multiple key characteristics including: input features and neural network parameters. The neural network parameters include number of neurons, training algorithm, activation functions, and initial weights. Usually, to obtain

Parts of this chapter appeared in : A. Haidar and B. Verma, "A novel approach for optimizing climate features and network parameters in rainfall forecasting," Soft Computing, pp. 1-12, 2017.

the best input features, feature selection is applied over the input dataset (see Chapter 4). Feature selection or reduction is the process of finding the best subset of input features that results in better accuracy than using all of the available inputs [137, 141]. Redundant and/or irrelevant features are removed through the process of feature selection [141]. Feature selection is usually applied by allocating a set of fixed hyper-parameters to the network. There is no guarantee that using a different set of parameters would select another subset of input features with better performance.

On the other hand, neural networks parameters are optimized to find the best combination that will lead to the highest accuracy of the model. Parameters selection is dependent on the type of problem [42]. No feature selection is usually applied to a dataset when performing parameters selection or optimization. Because input features data are fixed, another subset with different network parameters might have led to a better solution. Therefore, feature selection and parameters selection are mandatory in order to obtain the best network model.

Features and network parameters have been widely analysed to obtain the best forecasting model. Using all the available attributes may disturb the ability of specific input features in determining unusual events [85]. To select the best features and/or network parameters, a trial and error method can be used, in which a subset of the available features and another subset of the possible parameters are given to the network to be trained. Multiple networks are usually established, and the network with best performance over the testing dataset is selected. At the same time, it is known that another network with different parameters which have not been tested could have resulted in a better performance. An alternative approach is to apply a grid search in which all of the possible combinations are examined. The two previously mentioned methods are considered to be unusable when there are a large number of network characteristics or features to be optimized. Abhishek, Singh, Ghosh, and Anand developed multiple ANN models to forecast weather attributes [72]. They analysed the performance of the models based on subsets of network parameters, and concluded that the larger the number of neurons in the hidden layer the higher the performance. Wang and Sheng developed a generalized regression neural network to predict rainfall for a selected location in China [67]. The proposed model was compared to BPNN and step wise regression, and higher accuracy was recorded. The network parameters were manually selected with no optimization. Training the proposed or comparison model with another subset of parameters could have shown better accuracy. Khedhiri developed three ANNs to predict rainfall patterns for Prince Edward Island, Canada [80]. Data were pre-processed before being fed into two of the ANNs. The models were compared to alternative traditional time series forecasting models, and better accuracy was recorded with the developed ANNs. An ANN with data pre-processed using Halt-Winter exponential smoothing had the lowest error in terms of RMSE. The network parameters were selected manually, while the input features dataset consisted of only one variable (rainfall). Saba, Rehman, and AlGhamdi used a MLP and a RBFNN to estimate precipitation [101]. They proposed combining the two network architectures so as to overcome the limitations of each. The outputs of each model were combined using the average fusion method. The proposed approach was compared to single MLP and RBFNN, and resulted in higher accuracy. A trial and error method was used to select network parameters. Mekanik, Imteaz, and Talei developed an ANFIS model to forecast seasonal rainfall values for various locations in Victoria, Australia [84]. The seasonal rainfall was for three months: September, October, and November. The proposed models were generated based on different datasets representing climate indices for antecedent months. Eight different models were created for each location, and accuracy was recorded. Combinations of input features were limited to eight combinations. As usual, another combination of features could have produced a better performance than the selected ones.

In his survey, Darji, Dharbi, and Prajapati analysed 25 studies about rainfall forecasting for different durations [57]. More than 50% of these studies used hyperbolic tangent activation function when developing network layers. Again, other activation functions could have resulted in better forecasting accuracy. Devi, Arulmozhivarman, Venkatesh, and Agarnal proposed multiple network architectures for daily rainfall prediction [142]. Multiple ANN topologies and architectures with different parameters were proposed and compared. They concluded that Levenberg-Marquardt was the most effective in training the ANNs. Similarly, a trial and error method was used to develop models. Vathsala and Koolagudi proposed an approach to forecast peninsular Indian summer monsoon rainfall [38]. The closed-itemset-generation-based association rule method was utilized for feature selection, and K-means clustering for dimensionality reduction. The processed data were then added to a MLP of seven neurons in the hidden layer to classify five classes of peninsular Indian summer monsoon rainfall: flood, excess, normal, deficit, and drought. The number of neurons was selected based on the average of input features and number of classes.

An alternative approach is to select model components using EAs such as GA and PSO. These algorithms have been combined with ANNs in different engineering applications [26, 143]. EAs have been mainly used in selection and optimization problems [140, 144-147]. Hameed, Bye, and Osen used GA and PSO to optimize crane design parameters [148]. Muralitharan, Sakthivel, and Vishnuvartan utilized a GA and a PSO to optimize ANN weights for energy demand prediction [27]. Yamasaki, Honma, and Aizawa applied PSO optimization to select convolutional neural network hyperparameters [149]. Nasseri, Asghari, and Ebedini integrated a GA into an ANN to predict short-term rainfall for the Parramatta River catchment, Australia [63]. Meng proposed a GA to optimize BPNN weights [75]. Ding and Dong incorporated GA and ANN in addition to other models to forecast Liujiang river runoff [150]. Jiang and Wu developed a hybrid approach to forecast monthly rainfall for Guilin, China [78]. GA operators were integrated into PSO to increase the prediction accuracy of the ANN. Some network parameters were used to evolve the ANN architecture, while other parameters were kept fixed as the activation functions between layers. Six precedent rainfall months represented the input vector. No input feature selection was applied on input data. The proposed approach was compared to BPNN, GA-NN, and the Particle Swarm Optimization Neural Network (PSO-NN). Better results in terms of RMSE were obtained with the proposed hybrid model over two years of testing data. Jiang and Wu introduced another hybrid algorithm (GASA-NN) that combined the GA and the simulated annealing algorithm to find the first initial weights of the ANN [77]. This optimization was followed by using the BP algorithm to search for the optimal trained network. The network architecture including number of neurons was selected by a trial and error method to 4-6-1. The proposed algorithm GASA-NN was compared to alternative training methods GA-NN, BPNN, and ARIMA model where lower error was recorded with the proposed training approach. Wu, Long, and Liu combined PSO into a GA to evolve a RBFNN [83]. The worst individuals in the genetic algorithm iteration were manipulated using PSO mechanisms. The proposed model (RBF-HPSOGA) was compared to RBFNN and a RBFNN evolved using pure GA (RBF-GA), and better accuracy in terms of RMSE, r, and MAPE was achieved with the proposed model. Hence, climate features and network parameters selection should be carefully considered in order to ensure better forecasting quality.

Multiple applications were applied to select ANN parameters. In addition, multiple studies were utilized to select best input features for prediction. When selecting input features, network parameters were assigned automatically. When selecting network parameters, all features were used as input features of the model. There is no guarantee that selection of parameters based on handcrafted network parameters will result in best performance. On the other hand, there is no guarantee that using all input features while applying network parameters selection will reveal best performance. The aim of the following studies is to select both the best input features and the neural network parameters for forecasting monthly rainfall based on a proposed hybrid GA.

5.2 General Overview

To develop a classification or prediction model, the main goal is to find the set of characteristics that when combined result in the highest accuracy. These characteristics may not belong to the same type, and could be correspondingly different. Let us assume that we have a set S that affects the model's overall performance $f(S)$.

$$f(S) = \begin{cases} S^1 = S_1^1, S_2^1, S_3^1, \dots, S_{n_1}^1 \\ S^2 = S_1^2, S_2^2, S_3^2, \dots, S_{n_2}^1 \\ \quad\quad . \\ \quad\quad . \\ \quad\quad . \\ S^m = S_1^m, S_2^m, S_3^m, \dots, S_{n_m}^m \end{cases} \tag{5.1}$$

S^i is a subset of characteristics on which model performance is dependent, m is the total number of sets the model relies on, n_i is the total number of subsets in each set, $1 \le i \le m$. The main target is to determine the best collection of subsets S that contains subsets from each characteristic set.

$$S = \left(S_i^1, S_j^2, \dots, S_k^m\right) \tag{5.2}$$

$(i, j, \dots, k)$ are positive integers that vary between (1, 1, …,1) and $(n_1, n_2, \dots, n_m)$ respectively, so that when S is given to the model, the highest accuracy is generated.

Different methods are followed to determine the combination that produces the best performance. In random search, a subset from each set is selected randomly. In other words, $(i, j, \ldots., k)$ is chosen randomly, then combined to estimate the function f. Using grid search, each subset in a set is combined with all the other subsets from other sets to find the model with highest accuracy. Selection and optimization algorithms, including EAs, are proposed to determine a subset from each set of characteristics that results in the best performance. With EAs, instead of trying all the possible combinations, the algorithm searches for the best combination (chromosome in GA and particle in PSO) to produce the optimal trained model.

In ANNs, two main types of characteristics affect the performance of the model: input features (S^1) and neural network parameters (S^2). The two types are distinct from each other, but are dependent. For each subset of input features $E \subseteq S^1$, there is a different set of parameters $O \subseteq S^2$ that reveal the best accuracy for that subset. Hence, selecting the best subset of input features based on a fixed subset of network parameters does not always guarantee that the best model will be obtained. There could be an alternative network parameters subset O' that could have obtained a higher accuracy. On the other hand, for each set of network parameters O, there is a subset of features $E \subseteq S^1$ that produces the best accuracy for that subset. Hence, selecting the best subset of network parameters based on a fixed subset of input features does not always guarantee that the best model is produced. There could be an alternative input features subset E' that could have resulted in better accuracy. Therefore, the two subsets of input features (F) and network parameters (P) are targeted.

$$F \subseteq S^1 \ and \ \ P \subseteq S^2 \tag{5.3}$$

When F and P are combined, the lowest error e of the ANN model β is achieved.

$$e = \ \beta(F, P) \tag{5.4}$$

5.3 A Novel Approach for Optimizing Climate Features and Network Parameters in Rainfall Forecasting

The climate input features and parameters of ANN highly affect the overall performance of the prediction model. Therefore, an appropriate approach for the selection of features and parameters is needed. A novel approach is proposed to select the input features and neural network parameters. A hybrid GA that combines natural reproduction and PSO characteristics was developed to select the best input features and network parameters for each month. The proposed method was used to forecast long-term rainfall values in

Innisfail, Queensland. The developed model was compared with an alternative climate and network parameters feature selection model, a climate feature selection model, and climatology, and a better accuracy was recorded with the proposed model. The skill score against the three alternative climate models was 17.41%, 21.68 % and 32.12 % respectively. The aggregated time series of the proposed model showed a Root Mean Square Error (RMSE) of 141.67 mm for a location with 3553.00 mm annual average.

5.3.1 Proposed Method

Each ANN topology has its own characteristics. ANN performance is highly dependent on its parameters. With FFNN, data flow in one direction from input toward output. The network performance varies based on various parameters, such as number of layers, number of hidden neurons, activation functions, training algorithm, initial weights, etc. Each combination of those parameters shows different performance after training. Hence, when designing an ANN, the parameters should be carefully considered.

In addition, the type and length of dataset affects the overall performance of the network. Having a large number of features does not always guarantee best accuracy. Some features may not be efficient and may decrease performance. The combination of multiple input features may increase or decrease accuracy. Through input feature selection, the available predictors are mapped into a subset that contains the best features and that results in the highest accuracy. This was shown in Chapter 4. Therefore, the best features should be considered when developing rainfall models.

To find the best combination of input features (F) and neural network parameters (P), an evolutionary based approach for selecting input features and network elements was proposed. The proposed hybrid algorithm is shown in Figure 5.1. The steps followed in generating the best chromosome with the highest accuracy are shown in Figure 5.2.

Usually, GA selects the best chromosome from the last iteration. The GA was modified in this study in an attempt to save the best chromosome generated in each iteration. Typically, the GA keeps no record about previous population, but because of the dynamic initial random weights generation used in this study, and the need to save trained neural networks, the best network in each iteration was saved so that it could be compared with the other generated networks in other iterations as shown in Figure 5.1.

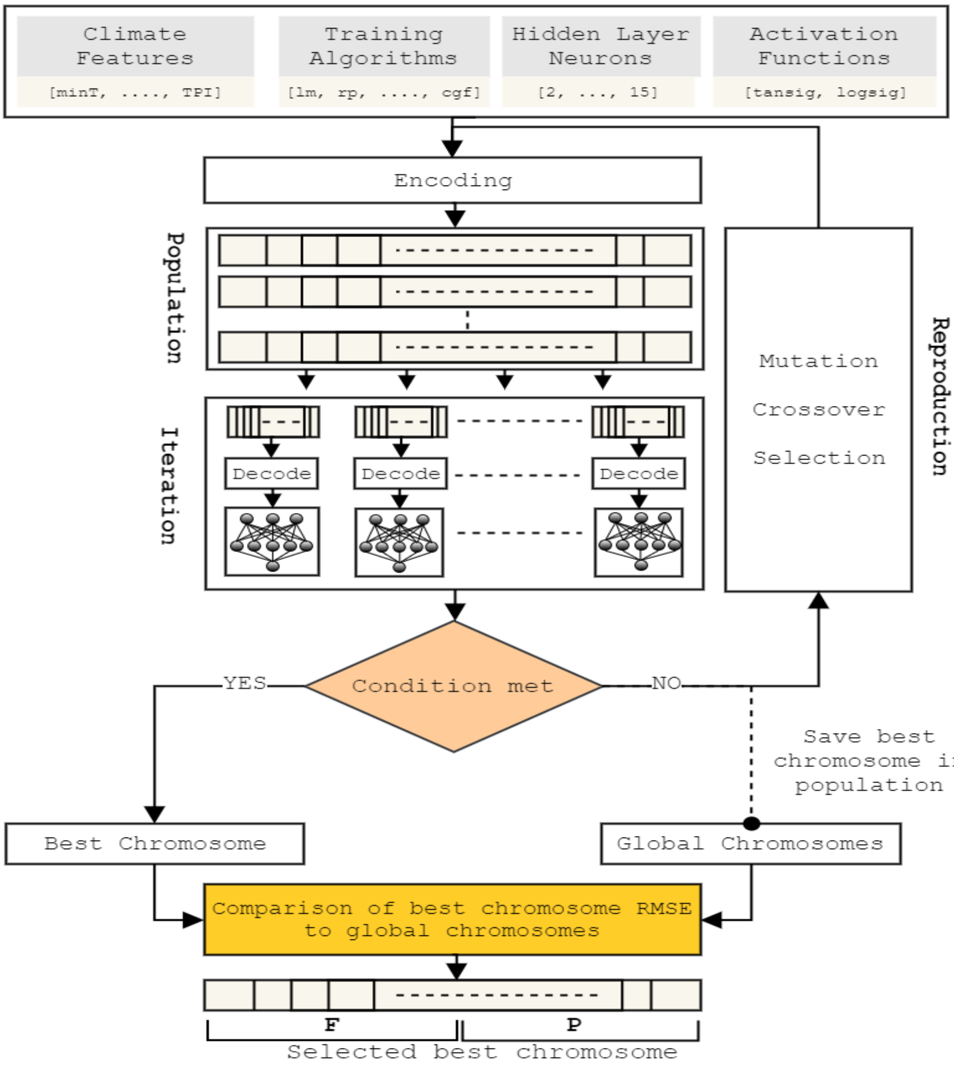

Figure 5.1 Proposed approach

The proposed hybrid GA based approach consists of many steps as shown in Figure 5.2. The first step generates population in which each chromosome is a collection of binary numbers representing climate input features, learning algorithms, hidden neurons, and activation functions. The second step is to train an ANN for each chromosome using selected parameters (1 selected, 0 not selected). The third step is to find and store the best chromosome in the population based on RMSE for each network. The fourth step is to conduct selection, crossover, and mutation, and get a new population. The above steps are repeated until the condition is met (max iteration or RMSE = 0). Once final iteration is completed, the selected chromosome is compared with global stored chromosomes from previous iterations. The best chromosome which produces the best network is selected. The accuracy is calculated based on the final selected chromosome and network model. The novelty of this approach is that in each iteration dynamic initial random

Algorithm: Hybrid Genetic Algorithm

Input: input features and network parameters

Output: selected best chromosome (subsets F and P).

```
Start
Generate Initial population.
Do
For each chromosome in iteration
    Decode the chromosome to input features and network parameters.
    Use selected features to make dataset.
    Take 10% of the dataset and put aside.
    Specify the feed forward network parameters based on decoded genes.
    Train and validate the network.
    Calculate RMSE over testing dataset.
End
If Condition not met
    Compare chromosome performance in this iteration.
    Save best chromosome and its network (recorded global chromosomes).
    Selection.
    Crossover.
    Mutation.
    Repeat
Else Condition met
    Compare best chromosome in last iteration performance to all saved
    chromosomes.
    Select best chromosome and best trained network.
End
```

Figure 5.2 Steps followed in generating best chromosome.

weights are generated, ANNs are trained, and best chromosome and its corresponding network is stored in a global chromosome register.

This novel idea gives the GA a wider searching space since different random initial weights and biases, if selected as the best to survive, are assigned to the same network in the next population. The above characteristic is taken from PSO where the global best is usually saved through iterations and particles flow towards the best solution. In proposed hybrid GA, the best chromosome is saved but other chromosomes do not follow it. When saving global chromosomes, the best network with the highest performance is guaranteed. This gives the proposed GA a wider searching criterion since different random initial weights and bias are given to the same network if selected as the best to survive.

In PSO, the global best is usually saved through iterations and particles flow towards the best global solution. In this updated GA, the best chromosome is saved but other chromosomes do not follow it. As shown in Figure 5.1, in each generation the best chromosome is saved. When the condition is met, the selected chromosome in the final population will be compared with “parents” and “grandparents” to select the best network that contains the best parameters and input features. The main reason behind not selecting the initial random weights and bias is that the randomly generated weights and bias may not be the optimal selection. Following this technique, there is a 50% chance that the selected chromosome may have lower performance in the next iteration. On the other hand, it may reveal better performance. When saving global chromosomes, the best network with the highest performance will be guaranteed.

ANN performance varies based on data. On the other hand, data output is affected by network design. Therefore, the “perfect” combination of both data and parameters should result in the highest accuracy. With that outcome in mind, when selecting input features, network parameters should also be taken into consideration. Furthermore, there is a need to know the effect of climate features on different months and locations across Australia. Therefore, each chromosome would contain information about features and network parameters.

5.3.2 Data (Case Study 3)

The selected location used to perform the study was Innisfail (17.52°S, 146.03°E). In total, 11 input features made up the input features dataset. Monthly rainfall amounts varied between the months because of seasonality. The monthly averages are shown in Figure 5.3. Further details are shown in Table 3.4 and Table 3.5.

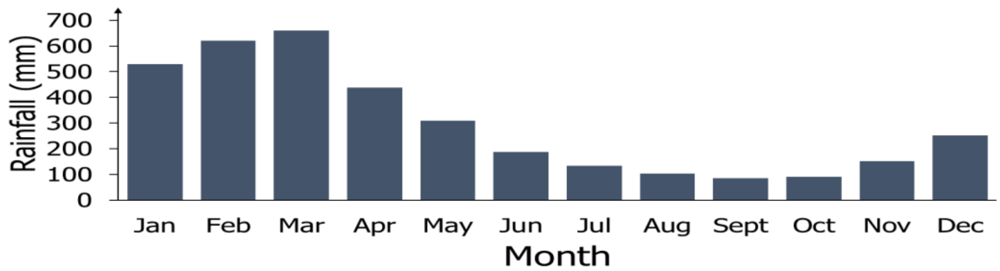

Figure 5.3 Monthly rainfall averages for Innisfail (1908-2015).

5.3.3 Experiments and Results

The collected climate indices have different effects on rainfall variability over different locations and time-frames in Australia. Therefore, a climate index may have a significant contribution to make when determining the amount of rainfall in a specific month, and have no such contribution to make as regards another month. In an attempt to determine the input features and best parameters for each month, the dataset was divided into 12 datasets representing each month. The main reason for this was to let the network learn to predict the value for the next year based on the weather conditions of the current year. Following this technique, models can be used to forecast monthly rainfall if reasonable performance was recorded.

The dataset was adapted so that the climate features collected at time t are applied to predict the monthly values at time t+1 as shown in the following equation:

$$y_{t+1} = f(input_t^1, input_t^2, input_t^3 \; \ldots\ldots, input_t^m) \qquad (5.5)$$

where y is rainfall generated by the trained ANN, f denotes trained ANN and $input_t^i$ represents a climate feature at time t where $1 \leq i \leq m$. m is the number of input features selected by the network m ≥ 1.

5.3.3.1 Input Features and Network Parameters Encoding

To conduct experiments, MATLAB was used. The selected chromosome size was 20. Some 11 binary digits were taken to represent input features, with one binary number each. The number 1 means that the feature at this index is included. Four binary numbers were used to represent the number of neurons. The highest number of neurons was selected to 15 since the dataset was small. Three binary numbers were taken in the chromosome to represent training algorithms. Eight training algorithms were selected as possible candidates to forecast monthly rainfall. Two activation functions, hyperbolic tangent sigmoid (tansig) and log-sigmoid (logsig), were used. The binary number 1 was allocated to identify the activation function between input-hidden layer and hidden-output layer. Thus, 0 represented tansig, while 1 denoted logsig. A summary of the network parameters that were used is shown in Table 5.1. Chromosome encoding is shown in Figure 5.4.

Table 5.1 Summary of the used network parameters.

<table>
<tr><th>Network parameters</th><th>Number</th><th colspan="15">Values</th></tr>
<tr><td rowspan="8">Training algorithms</td><td rowspan="8">8</td><td colspan="15">Bayesian Regularization</td></tr>
<tr><td colspan="15">BFGS Quasi-Newton BP</td></tr>
<tr><td colspan="15">Fletcher-Powell conjugate gradient BP</td></tr>
<tr><td colspan="15">Gradient descent BP</td></tr>
<tr><td colspan="15">Levenberg-Marquardt BP</td></tr>
<tr><td colspan="15">Powell- Beale conjugate gradient BP</td></tr>
<tr><td colspan="15">Resilient BP</td></tr>
<tr><td colspan="15">Scaled conjugate gradient BP</td></tr>
<tr><td>Neurons</td><td>15</td><td>1</td><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td><td>8</td><td>9</td><td>10</td><td>11</td><td>12</td><td>13</td><td>14</td><td>15</td></tr>
<tr><td>activation functions</td><td>2</td><td colspan="7">Hyperbolic tangent sigmoid</td><td colspan="8">Log-sigmoid</td></tr>
</table>

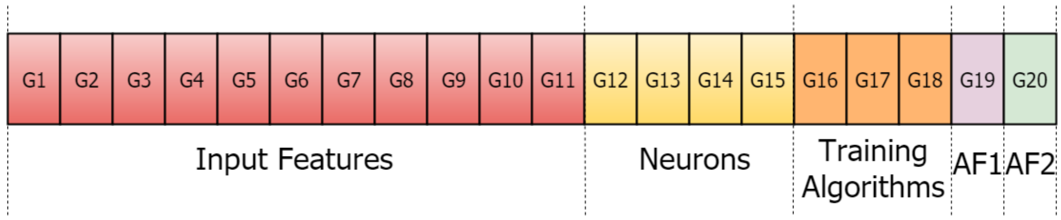

Figure 5.4 Chromosome encoding.

5.3.3.2 Climate Features and Neural Network Parameters Selection based Hybrid Genetic Algorithm (CFPS-HGA)

The selected fitness function consisted of a three-layered FFNN (input-hidden-output). Throughout each month in the proposed GA, 75% of the selected dataset for each month was used for training 15% for validation, and 10% for testing.

In order to select the best combination of ANN input features and elements, a hybrid GA was used. The hybrid GA parameters are shown in Table 5.2. The proposed GA identified the model that generated the highest accuracy in each iteration. The hybrid GA attempts to select the best chromosome that encodes the best network parameters and input features. Initial population was randomly created, and population size was set at ten. Crossover was set to “scattered,” and mutation to “uniform.”

The hybrid GA was applied 12 times, with each time corresponding to one month. MAE, RMSE, and r values were calculated by the GA in order to measure the accuracy of each selected chromosome. The MAE, RMSE, and correlation values are shown in Table 5.3.

Table 5.2 Hybrid genetic algorithm parameters.

Parameter	Value
Population	10
Chromosome Length	20
Chromosome Type	Binary
Iterations	50
Elitism	2
Crossover	Scattered
Mutation	Uniform
Mutation probability	0.2

Table 5.3 MAE, RMSE, and r values generated by each optimized network based hybrid genetic algorithm (CFPS-HGA).

Month	MAE (mm)	RMSE (mm)	r
January	151.14	199.73	0.62
February	235.98	276.78	0.61
March	169.67	194.63	0.84
April	125.91	150.34	0.60
May	100.40	119.46	0.70
June	70.60	100.40	0.70
July	74.63	84.30	0.65
August	37.58	50.20	0.82
September	43.76	51.28	0.84
October	100.94	118.12	0.47
November	66.58	84.30	0.89
December	70.34	81.07	0.91

The annual average of the chosen location is 3553.0 mm. March has the highest monthly rainfall average, as shown in Figure 5.3 (≅662 mm). The maximum amount of precipitation in one month, 2684.6 mm, was in January 1981. MAE for each optimized network varied between the highest, 235.98 mm in February, and the lowest, 37.58 mm in August. RMSE varied between 276.78 mm in February and 50.20 mm in August. December showed the highest correlation with 0.91, while the lowest correlation obtained was in October with 0.47.

The nominated features for each month are shown in Table 5.4. Different input features were chosen for each month. The highest subset of features selected in optimization was six in both January and April. The Nino 3.4 climate index was selected in seven out of 12 months. Nino 1.2 and DMI were chosen in six months, and minimum temperature was selected five times. This means that these features are highly effective for the chosen location and weather zone. Two ENSO indices (Nino 1.2 and Nino 3.4) were the most highly used features. This coincides with previous investigations that considered ENSO as the first cause of rainfall variability over Australia [109].

Table 5.4 Selected features in each month.

Month	Count	Features
January	6	MinT, Nino 1.2, Nino 3.0, Sunspots, IPO, TPI
February	5	MinT, Nino 3.0, Nino 3.4, Nino 4.0, Sunspots
March	4	MinT, Nino 1.2, Nino 3.4, TPI
April	6	MaxT, Nino 1.2, Nino 3.0, Nino 3.4, DMI, IPO
May	2	Nino 1.2, DMI
June	2	DMI, IPO
July	5	MinT, SOI, Nino 1.2, Nino 3.4, Sunspots
August	4	MinT, MaxT, SOI, Nino 4.0
September	5	MaxT, Nino 1.2, Nino 3.4, DMI, TPI
October	5	MaxT, Nino 3.4, Nino 4.0, DMI, Sunspots
November	1	Nino 4.0
December	4	Nino 3.0. Nino 4.0, DMI, IPO

Minimum temperature was used in the three months along with highest rainfall averages (January, February, March), while maximum temperature was selected as the predictor for three months with the lowest rainfall averages (August, September, October). This demonstrated the effectiveness of temperature values in forecasting rainfall amounts. The range of neurons varied between six and 15 along the optimized networks. The available number of neurons was low because the dataset is small. A total of 83% of the best chromosomes contained neurons from the upper available values.

The selected training algorithms are shown in Table 5.5. Scaled conjugate gradient BP was used in five months. Four out of the eight available training algorithms were not chosen in any month. Both scaled conjugate BP and Levenberg-Marquardt BP were selected in 75% of the months.

Table 5.5 Selected algorithms based on each month

Number	Training algorithm	Months
1	Levenberg-Marquardt BP	March, May, August, November
2	Resilient BP	None
3	Bayesian regularization	February, October
4	Scaled conjugate gradient BP	January, April, June, July, September
5	Gradient descent BP	None
6	BFGS Quasi-Newton BP	None
7	Powell- Beale conjugate gradient BP	December
8	Fletcher-Powell conjugate gradient BP	None

This reveals the effectiveness of those training algorithms in rainfall forecasting. There was no correlation between the type of monthly rainfall and training algorithm selection since the hybrid GA adopted the same training algorithm for months with different rainfall averages. Tansig was selected seven times in the input-hidden layer and five times in the hidden-output. Log sigmoid was specified five times in the input-hidden layer and seven times in hidden-output. Therefore, both activation functions should be considered when anticipating rainfall.

5.3.4 Comparative Analysis

The generated forecasts for each month were compared against three approaches: Climate Features and Parameters Selection based Genetic Algorithm (CFPS-GA), Climate Features Selection based Genetic Algorithm (CFS-GA), and climatology. The only difference between the proposed approach and CFPS-GA approach was saving the best chromosome in each population. The same specifications were given to the GA as shown in Table 5.2, and the same network type (FFNN) for the fitness function. Network parameters and features were selected without saving the best chromosome in each population.

The CFS-GA approach has shown its suitability in precipitation forecasting [151]. An ANN based GA was applied to predict rainfall based only on climate features selection. Furthermore, the alternative approach had no memory which recorded the best chromosome in each population. The fitness function contained a three-layered FFNN with eight neurons in the hidden layer, and used Levenberg-Marquardt BP as the training algorithm.

The GA was applied 12 times, once for each month in each approach. MAE, RMSE, and r values for the comparison approaches are shown in Table 5.6 and Table 5.7 respectively. The proposed approach resulted in better accuracy than CFPS-GA in terms of RMSE over 11 months. A higher RMSE was recorded in June. The highest difference recorded was in March, with 95.91 mm. The average difference was 28.46 mm for each month.

Table 5.6 MAE, RMSE, and r values for climate features selection approach based GA (CFS-GA).

Month	**MAE**	**RMSE**	**r**
January	188.82	222.69	0.48
February	288.09	321.65	0.64
March	258.74	309.71	0.37
April	147.42	182.27	0.42
May	125.78	145.09	0.46
June	89.02	116.78	0.44
July	87.11	98.77	0.48
August	63.65	81.23	0.57
September	57.65	69.30	0.66
October	85.64	126.16	0.43
November	89.93	127.06	0.78
December	81.87	116.86	0.74

The proposed approach that is based on selecting both input features and network parameters using hybrid GA produced better accuracy in terms of MAE, RMSE, and r values in all the months compared to climate features selection. The highest difference in terms of RMSE was the same as in CFPS-GA, where a 155.07 mm difference was recorded in March. This demonstrates the ability of the proposed approach in performing well with high rainfall values. The average difference was 33.91 mm for each month.

Table 5.7 MAE, RMSE, and r values for parameters and climate features selection approach based GA (CFPS-GA).

Month	MAE (mm)	RMSE (mm)	r
January	182.71	219.07	0.52
February	248.23	288.42	0.53
March	263.39	290.54	0.85
April	153.66	180.44	0.14
May	131.82	149.24	0.49
June	75.75	89.03	0.74
July	89.08	97.78	0.55
August	41.21	51.95	0.84
September	63.27	82.67	0.55
October	99.57	121.39	0.44
November	128.17	156.34	0.53
December	116.93	125.28	0.59

In order to measure the skill of the proposed approach, SS values were generated for each month against three reference models (climate features and parameters selection, climate features selection, and climatology). Climatology is a reference model with which comparison is made when evolving rainfall models [105]. The training/validation datasets were combined and months were averaged. Then forecasts were released based on each month's average. The SSs for the proposed approach against alternative models are shown in Table 5.8 for each month. The SS varies based on the amount of rainfall and location. The CFPS-HGA showed better forecast skill than the CFPS-GA in all months, except June. The CFPS-HGA showed better SSs against climate features selection and climatology in all of the months. The average SS in comparison with alternative models and climatology was 17.41%, 21.68% and 32.12% respectively. The proposed approach performed better than alternative approaches in March (33.01%, 37.16%, and 41.27%), which is the month that receives the largest rainfall amount annually.

Table 5.8 Proposed model monthly skill scores compared with alternative approaches.

Month	SS against CFPS-GA	SS against CFS-GA	SS against climatology
January	8.83	10.31	21.53
February	4.03	13.95	18.91
March	33.01	37.16	41.27
April	16.68	17.52	25.02
May	19.95	17.66	26.49
June	-12.78	14.03	22.11
July	13.79	14.65	26.14
August	3.36	38.19	42.40
September	37.98	26.01	42.12
October	2.69	6.37	17.41
November	46.08	33.66	54.56
December	35.29	30.62	47.54
Average	17.41	21.68	32.12

Boxplots for SSs against each model are shown in Figure 5.5 Skill scores boxplots of the proposed approach against alternative approaches.. The red line represents the median of SSs. The highest range of SSs was obtained against CFPS-GA. Some 25% of SSs in comparison with CFPS-GA were higher than 34.72%, 25% of SSs in comparison with CFS-GA ranged above 32.90%, while 25% of SSs against comparison with climatology ranged above 41.00%.

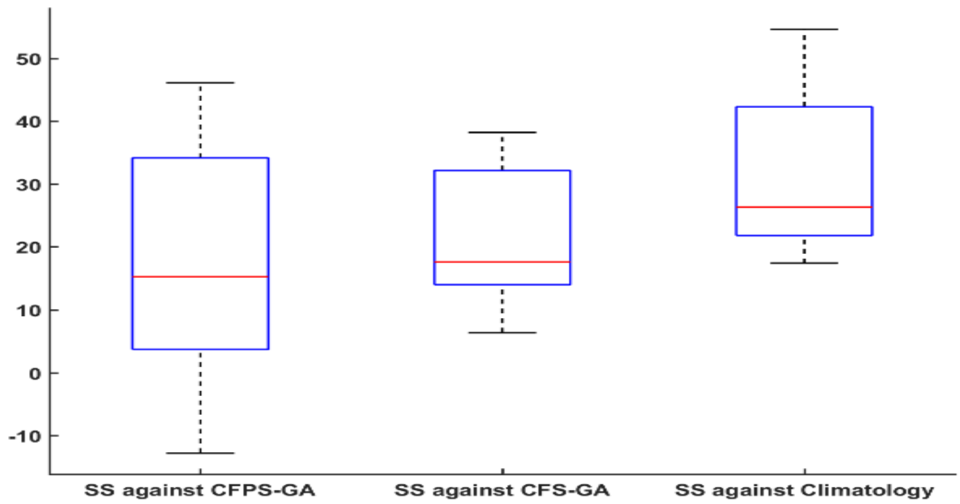

Figure 5.5 Skill scores boxplots of the proposed approach against alternative approaches.

In order to measure the performance of the whole method, the outputs of each month (testing dataset of each optimized network) were aggregated to create the time series between January 2005 and December 2015 (11 years). MAE, RMSE, and r values of the developed models and climatology are shown in Table 5.9. The three ANN methods had better performance as against climatology. This demonstrates the capability of machine learning approaches in monthly rainfall forecasting. The highest accuracy was recorded with CFPS-HGA (103.81 mm, and 141.67 mm in terms of MAE and RMSE). In addition, higher correlation values were also obtained. Differences of 58.65, 36.55, and 29.67 mm in terms of RMSE was obtained when compared to climatology, CFS-GA, and CFPS-GA respectively. Although the proposed approach requires more memory, and takes a longer time, its accuracy surpasses that of the other models.

Table 5.9 MAE, RMSE, and r values for climatology, CFS-GA, CFPS-GA, and CFPS-HGA.

Model	MAE (mm)	RMSE (mm)	r
Climatology	147.21	200.32	0.69
CFS-GA [151]	130.30	178.22	0.76
CFPS-GA	132.81	171.34	0.78
CFPS-HGA (proposed approach)	103.81	141.67	0.86

In Figure 5.6, climatology generated forecasts are compared to actual rainfall throughout the testing dataset. Figure 5.7, Figure 5.8, and Figure 5.9 denote the combined dataset generated by the 12 developed ANNs as compared to actual rainfall for each selection method (CFS-GA, CFPS-GA, CFPS-HGA). The continuous line defines actual rainfall, while the dotted line shows outlooks. The generated forecasts in Figure 5.6 show repetitive values for each year. There was no skill in forecasting peak or non-rainy values. It was noticed that the climate feature selection approach had the capability to learn the pattern, but was not able to forecast the amount of rain accurately (Figure 5.7). In Figure 5.8, rainfall values were underestimated at some locations. In Figure 5.9, the networks had the skill to accurately predict precipitation quantity at some locations (as in March 2012: actual, 1228.5 mm; forecasted, 1148.20 mm). This ability in forecasting peak rainfall is helpful for different industries, including sugarcane. Furthermore, some non-rainy periods were also predicted, and in 2013 and 2015 no values were higher than 700 mm. This is also helpful as farmers could take into consideration rainfall uncertainty at an earlier time. Even if the amounts are not exact, it provides information about weather uncertainty for sugarcane farmers.

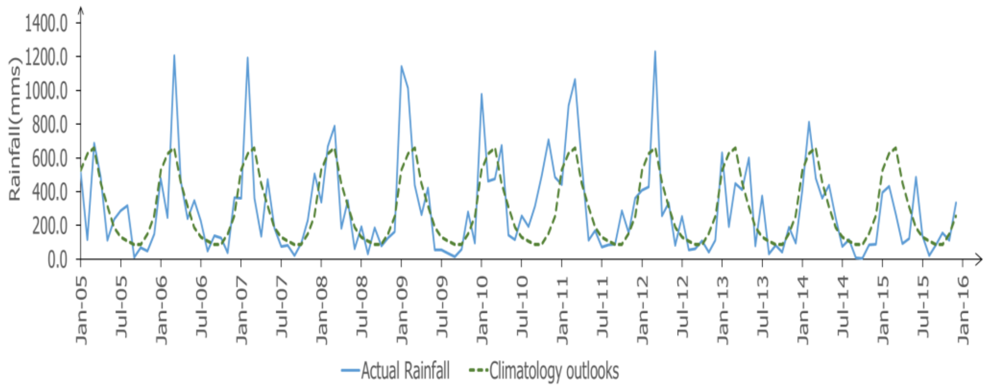

Figure 5.6 Rainfall values as compared to climatology.

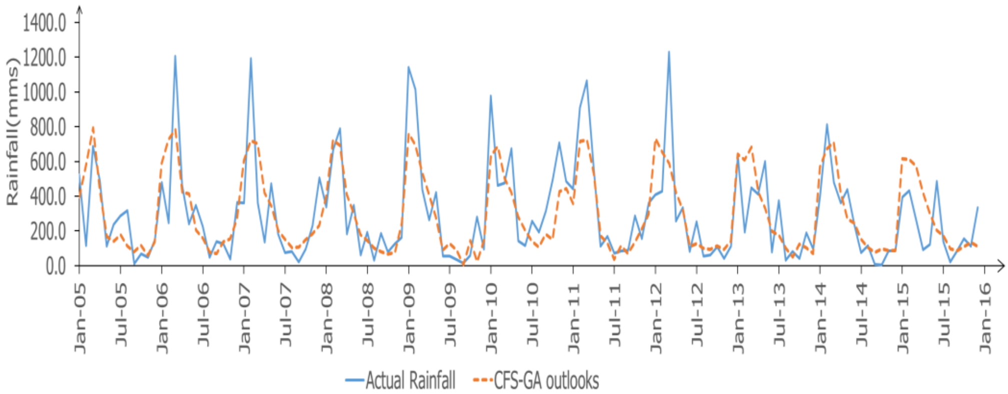

Figure 5.7 Rainfall values as compared to climate features selection based GA outlooks.

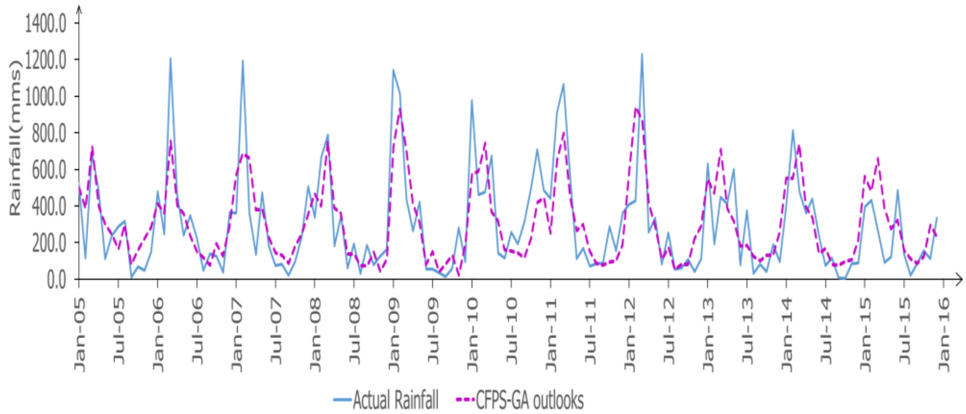

Figure 5.8 Rainfall values as compared to climate features and network parameters selection based GA outlooks.

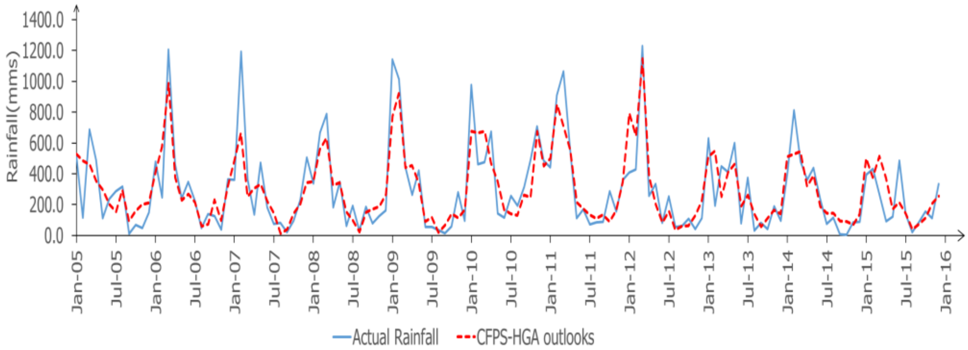

Figure 5.9 Rainfall values as compared to climate features and network parameters selection based hybrid GA outlooks.

5.3.5 Summary

A novel GA based approach for selecting climate input features and network parameters was examined in this study. The proposed model, which is based on input and parameters selection using a hybrid GA, produced the highest SS when compared to climatology and alternative selection methods (32.12%, 21.68% & 17.41%). Climate features and network parameters selection using the proposed hybrid GA showed better performance with 141.67 mm RMSE for a location with an annual average rainfall of 3553.0 mm. However, climatology, climate features selection based GA, and climate andparameters selection based GA resulted in 200.32 mm, 178.22 mm, and 171.34 mm respectively. The recorded correlation was 0.86. This study showed that the proposed approach is promising for rainfall forecasting, and can be presented as an alternative model.

5.4 A Mutli-Level Optimisation Method for Selecting Neural Network Characteristics in Rainfall Forecasting

Selecting model characteristics is essential in order to obtain the best performance in prediction. Different sets of characteristics affect the overall accuracy of a specific machine learning algorithm. EAs have been widely used to select model characteristics. In this section, a new approach for selecting model characteristics based on a novel multi-level optimization method is examined. The method introduces two hybrid GAs to select input features and ANN parameters. The proposed approach was evaluated on real-world data for monthly rainfall prediction. MAE, RMSE, r, R^2, IPE, and SS statistical measurements were used to assess the performance of the proposed approach against alternative forecasting and optimization methods. Some 12 ANN models were generated using the proposed method, representing each month in the year. The aggregated time series of forecasts from each model (January 2005, and December 2015) resulted in 65.30 mm, 96.92 mm, and 0.938 MAE, RMSE, and r respectively. The SS of forecasts generated by the models between January 2005 and December 2015 was calculated against climatology, MLR, climate feature selection using a standard GA, climate and network parameters selection using a standard GA, and climate and network parameters selection using a hybrid GA respectively. Higher SS values were recorded against each method with 51.62% for climatology, 45.80% for climate feature selection using a standard GA, 43.44 % for climate and network parameters selection using a standard GA, and 31.59% for climate and network parameters selection using a hybrid GA. Finally, eight years (January 2005-December 2012) were compared with the Australian Community Climate and Earth System Simulator (ACCESS). A better SS was reported in comparison with ACCESS by the generated models. This demonstrates the capacity of this approach when designing alternative forecasting systems.

In our previous work, we applied input features selection using a GA [151]. The main purpose of the study was to select best climate features to predict rainfall values for each month in a selected location in Queensland, Australia (Bingera). The proposed approach had its fitness function as a FFNN with a set of fixed neural network parameters. Although results were better than using all of the dataset, another portion of the available network parameters might have shown better performance. In addition, we applied feature selection using PSO to determine whether the rainfall range would be higher or lower than average in each month [152]. The proposed approach was used over five different locations along the eastern side of Australia: Innisfail, Plane Creek, Bingera, Maryborough, and Yamba. Reasonable accuracy was recorded with the proposed selection method, but the network parameters were fixed. A different set of hyper parameters in the objective function might have obtained a higher classification accuracy. In an attempt to overcome the issue of determining the elements of the model, a hybrid GA was proposed to select both input features and network parameters for predicting monthly rainfall values in Innisfail, Queensland, Australia [153]. The proposed approach was compared to alternative mechanisms, and better accuracy was recorded in terms of MAE, RMSE, and r values. Two types of the model's characteristics were selected on the same level (input features and network parameters). Mutation and crossover operators combined different subsets from different characteristics of the neural network while generating the best trained model.

This study focussed on model characteristics optimization as a key driver in obtaining optimal accuracy. The original contributions of this study are summarized below.

- A new approach for selecting a model's characteristics through a multi-level optimization strategy.
- A mechanism to optimize neural network parameters and features through hierarchical selections using a hybrid GA.
- A method for automatically building weather forecasting models for locally specified regions.
- Evaluation of the proposed approach for monthly rainfall forecasting data in eastern Australia and comparison to alternative forecasting methods.

Network parameters can be selected at the same time as input features. This is a valid approach because two different types of characteristics of the model are being examined on the same level. This method has been evaluated when, at the same time, a hybrid GA was utilized to select input features and neural network parameters [153]. The level of abstractness was given so that subsets from $S^1$and S^2 could intervene with each other to determine the best combination for the proposed model. The encoded chromosome consisted of genes that represented both targeted subsets (P and F).

5.4.1 Proposed Method

It is proposed here that the characteristic groups of a model should be selected with each on the same level, which can be called Climate Features and Network Parameters Selection – Two Levels Optimization (CFPS-TLO). The optimization process should be applied over multiple levels because of the internal connections between each group's members. Selecting values to form the subset F (P) should be applied only from values of S^1(S^2), which is the set that contains all possible input features (network parameters). The optimization process for each characteristic will be recursively embedded on each other, so that from each subset in S^1 elements will be selected without directly intervening in the selection of other elements from the other subset S^2. The mutation and crossover (position and velocity in PSO) are applied over each characteristic set element only. The proposed method is shown in Figure 5.10.

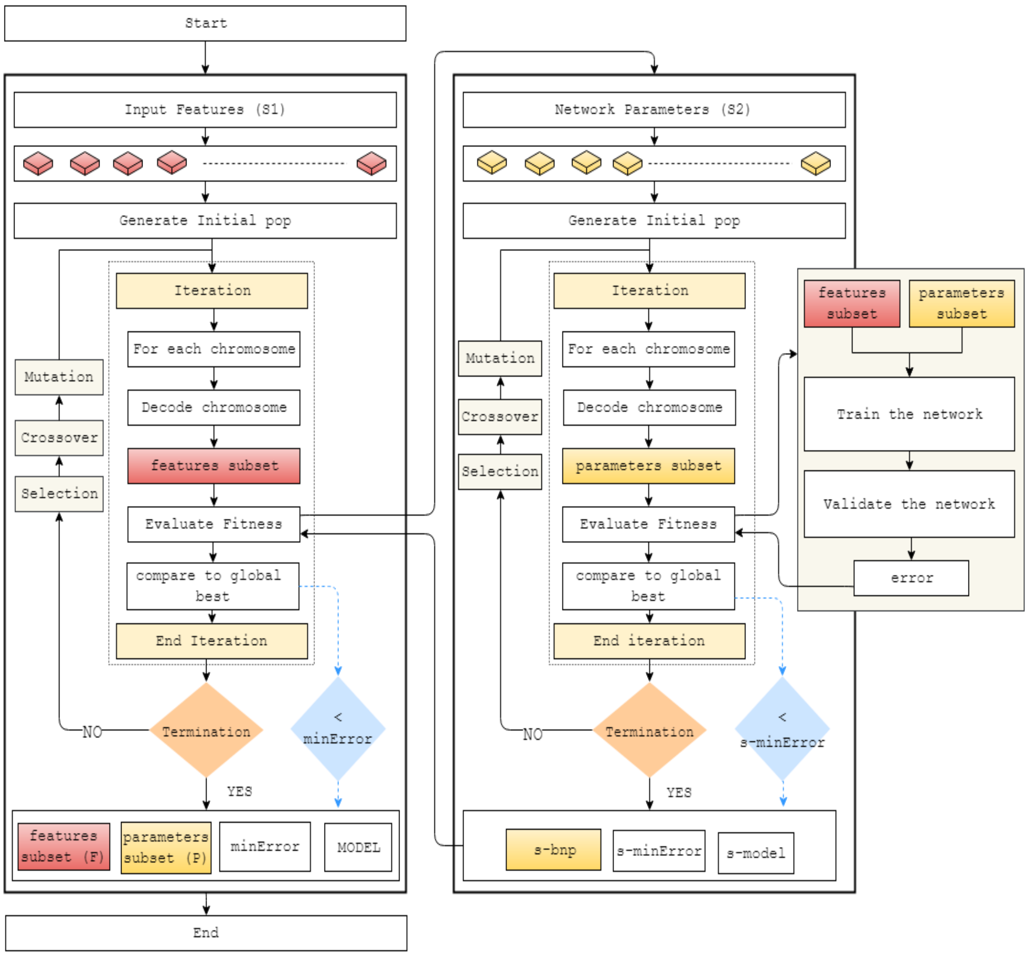

Figure 5.10 Proposed method: Climate Features and Network Parameters Selection – Two Levels Optimization (CFPS-TLO).

To find the best ANN model, the same hybrid GA proposed in the previous Section 5.3.1 was incorporated, but this algorithm was used over two levels of optimization. Knowing that, different optimization algorithms can be selected and used as the standard GA and PSO. The first optimization is applied to select the best input features subset. Inside that, another optimization is applied to select the network parameters that would produce the highest accuracy. For each chromosome in the first optimization, a new optimization technique is used to select the best network parameters. In other words, the fitness function of the first optimization technique is another optimization process.

The hybrid GA over the two levels combines natural reproduction with a characteristic of PSO. With PSO, particles compare their performance to the global best solution and replace it if better accuracy was obtained. With the GA, the best chromosome that reveals the highest performance in the last iteration is returned as the selected best solution. When running the ANN with random initial weights, the chromosome that was selected as the best chromosome in the previous iteration may reveal a lower accuracy when trained again in the new iteration.

Therefore, the best performance of the optimization problem may be lost. To overcome this, the ability to save the best global solution applied in PSO was integrated into the GA so that the best solution is always guaranteed. The chromosomes will be compared to a global best solution that will be returned when the termination condition of the GA is achieved. Each chromosome scalar value (error) is compared to the global best solution, and those which hold better prediction accuracy replace it. Further details about the algorithm are shown in Figure 5.11 [153].

The optimization algorithm is applied over two levels. In the first level, climate input features are represented in the search space. In the second level, neural network parameters are represented in the search space. The steps followed in optimizing the input features and the network parameters are shown in Figure 5.11. The best input features subset (F), best network parameters (P), lowest error in optimization (minError), and the model with highest accuracy (MODEL) are initialized first. The first hybrid GA starts with the initial population being randomly created. The randomly created chromosomes represent subsets of the input features dataset. For each chromosome, the fitness function is evaluated. The fitness function runs an optimization process to select the best network parameters that fit the selected input features subset (chromosome). The aim of this optimization is to select the best network parameters (s-bnp), min error (s-minError), and best model (s- model). Each chromosome in the iteration is decoded to determine the network parameters, combine it with the features subset, then after training an ANN returns the error of prediction. The error of prediction will be compared against error (s-min) in the global best of the selected optimization, and will replace it if better accuracy was obtained. When the condition is met, (s-bnp, s-min, s-model) values are returned.

This s-min error represents the scalar value of the chromosome fitness function in the first level. It is compared to the global best, and replaces it if a lower error was found. This process continues until the condition is met in the first optimization, where a global best that contains F, P, minError, and MODEL is returned.

Algorithm: Climate Features and Network Parameters Selection – Two Levels Optimization

Input: input features set (S^1), network parameters set (S^2)

Output: subset F, subset P, trained ANN MODEL, lowest error in prediction minError

```
1:  Initialize F, P, minError, MODEL
2:  Generate initial population randomly
3:  Do
4:  for each chromosome x in population
5:     Evaluate fitness function:
6:          start (second optimization)
7:          Initialize s-bnp, s-minError, s-model
8:          Generate initial population
9:          Do
10:         for each chromosome y in population
11:            Decode x into input features and y into network parameters
12:            Evaluate fitness function:
13:                 Train the network
14:                 Validate the network
15:                 return RMSE
16:            Compare RMSE to global best in second level optimization:
17:                 if (RMSE< s-minError)
18:                        s-bnp  ← y
19:                        s-minError ← RMSE
20:                        s-model ← net
21:                 end if
22:         end for
23:         if (termination condition not met)
24:            Selection
25:            Crossover
26:            Mutation
27:            Repeat (9)
28:         else
29:            return s-bnp, s-minError, s-model
30:         end if
31:         end (second optimization)
32:    Compare s-minError to global best in first level optimization minError:
33:         if (s-minError < minError)
34:            F ← x
35:            P ← s-bnp
36:            minError ← s-minError
37:            MODEL ← s-model
38:         end
39: end for
40: if (termination condition not met)
41:    Selection
42:    Crossover
43:    Mutation
44:    Repeat (3)
45: Else
46:    return F, P, minError and MODEL.
47: End
48: end
```

Figure 5.11 Steps for the selection of the best neural network model.

5.4.2 Data (Case Study 3)

The same dataset used in Section 5.3 was used in this section. Details about this dataset are shown in Table 3.4 and Table 3.5.

5.4.3 Experiments and Results

5.4.3.1 Experimental Setup

MATLAB was utilized to run experiments. Two sets of characteristics were selected for optimization using the proposed approach: climate input features and neural network parameters. The topology chosen for characteristics optimization was a three-layered FFNN. Some elements of the ANN were constrained to specific bounds to match the elements used in a recent approach [153]. For example, numbers of neurons were limited to 15. The available elements for the first level are shown in Table 3.5. The available elements for the second level are shown in Table 5.1.

As mentioned earlier, the proposed approach required two optimization algorithms since we have two sets of characteristics. The first hybrid GA was deployed to select input features. The second hybrid GA was used to select the best network parameters for each chromosome in each iteration of the first hybrid GA. The same set of parameters were selected for the first and second hybrid GA. The GA details are shown in Table 5.10.

Table 5.10 Hybrid GA characteristics.

Parameter	Hybrid GA 1	Hybrid GA 2
Characteristic	Climate input features	Network parameters
Chromosome type	Bit string	Bit string
Chromosome length	11	9
Population size	10	10
Iterations	30	30
Stall number of iterations	10	10
Scalar	RMSE	RMSE

The chromosome length of the first GA was 11, each bit representing an input feature. The number 1 means that the feature is included, while 0 means that the feature is not included. For the second hybrid GA, 9 bits were used to represent possible ANN parameters. As in Table 5.1, 3 bits were used to represent the training algorithm, 4 bits to represent neurons in the hidden layer, 1 bit to determine activation function in the hidden layer, and 1 bit to determine activation function in the output layer. Chromosome encoding for each level are shown in Figure 5.12 and Figure 5.13 respectively.

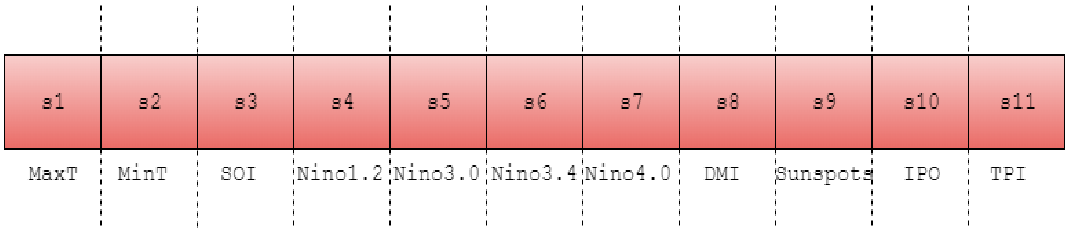

Figure 5.12 Chromosome encoding in the first level (input features selection).

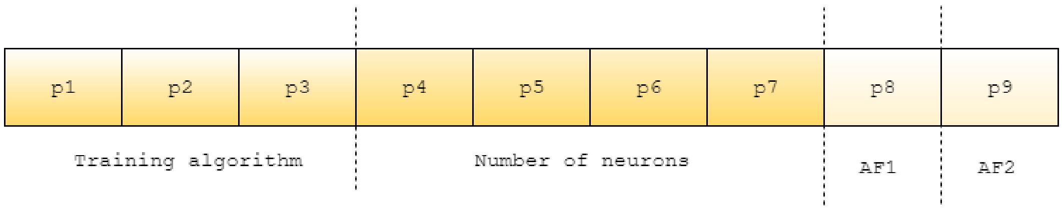

Figure 5.13 Chromosome encoding in the second level (network parameters selection).

5.4.3.2 Climate Features and Network Parameters Selection – Two Levels Optimization (CFPS-TLO) Results

The proposed approach was run 12 times, one for each month's dataset. Some 75% of selected data were used for training, 15% for validation, and 10% for testing. MAE, RMSE, r, and R^2 values were used to measure the accuracy for each month (Table 5.11). MAE values ranged between 14.228 and 182.016. RMSE varied between 17.997 and 217.990 (since March has the highest amount of rainfall in the year). Pearson correlation values varied between 0.833 and 0.980. Determination of coefficient values ranged between 0.693 as the lowest and 0.960 as the highest. The variations between statistical measurements values are due to the variation between monthly rainfall in each month. It is expected that months with higher rainfall averages will have higher prediction errors. February optimization held the highest MAE and RMSE compared to other months.

Table 5.11 MAE, RMSE, and r and R^2 values for the proposed approach's climate features and parameters selection (CFPS-TLO).

Month	MAE	RMSE	r	R^2
January	80.001	99.599	0.938	0.879
February	182.016	217.990	0.838	0.702
March	125.908	162.150	0.911	0.829
April	69.800	83.491	0.930	0.865
May	71.410	78.122	0.887	0.786
June	50.470	58.256	0.918	0.842
July	46.981	60.404	0.833	0.693
August	14.228	17.987	0.980	0.960
September	33.558	38.121	0.916	0.839
October	44.296	60.135	0.899	0.809
November	34.900	45.370	0.980	0.960
December	30.068	37.853	0.973	0.947

5.4.4 Comparative Analysis

Three training algorithms were selected in the 12 optimizations: scaled conjugate gradient BP, Levenberg-Marquardt, and Fletcher-Powell conjugate gradient BP. Scaled conjugate gradient BP was selected in six months, Levenberg-Marquardt in five, and Fletcher-Powell in one (April). The two training algorithms scaled conjugate gradient BP and Levenberg-Marquardt were selected 11 times out of 12 in this study, and nine times out of 12 in the previous study (where climate features and network parameters were selected on the same level). This proves the efficiency of the two algorithms in optimizing the network weights against the others. The same range of 6-15 neurons was obtained with the previous approach. Different selections of features were obtained compared to previous study. At least one feature was selected in the two optimization techniques for the same month. The hyperbolic tangent sigmoid activation function was selected seven times in hidden layer neurons, while the log-sigmoid activation function was used five times. This is the same as in previous study. The same number of selections were obtained in the output layer for the two activation functions.

A rainfall-forecasting model can be formed by combining the output of the optimized network models. To assess the accuracy of the whole mechanism, the testing dataset values in each model were combined to formulate a dataset of 11 years between January 2005 and December 2015. Figure 5.14 shows the combined time series predicted by each of the optimized networks for each month as compared to actual rainfall values between January 2005 and December 2015.

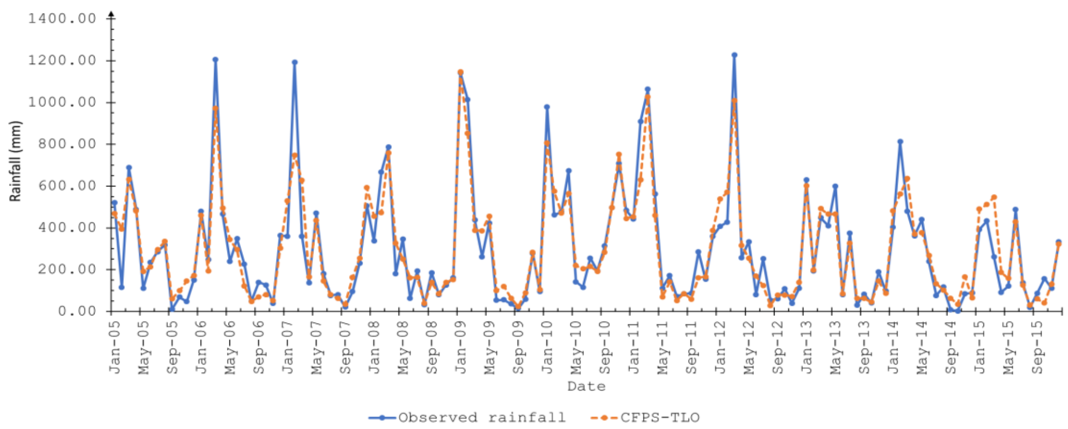

Figure 5.14 Observed rainfall against forecasts generated in each month between January 2005 and December 2015.

The dotted line represents predictions, in which actual values of rainfall are shown in a continuous line. Reasonable accuracy was noticed in most of the instances. The trend of rainfall was estimated perfectly by the models. Different ranges of rainfall were tested. The highest rainfall amount that was recorded in the last three years was not higher than 800 mm. The networks optimized using the proposed approach showed values relative to type of rain. This demonstrated the ability of such forecasting in forecasting severe weather conditions such as floods and drought.

The generated dataset was compared with various forecasting and optimization methods, including climatology, MLR, climate features selection based on a standard GA (CFS-GA), climate and network parameters selection based on a standard GA (CFPS-GA), and climate and network parameters selection based on a hybrid GA (CFPS-HGA). Climatology is a forecasting model which averages the previous amounts of rainfall over a selected month and releases predictions based on that average. It is often used as the base line comparison model in forecasting problems. MLR is a basic statistical model in which the targeted output is modelled as relationships between its predicators (independent variables). It has the aim of finding the best relationship between a set of independent variables (predictors) and a dependent variable (target), which is rainfall in this study. The basic formula of MLR models is as follows.

$$Rain = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_n x_n + \varepsilon$$

n is the number of independent variables, θ_0 is a constant, $x_k$represents a predictor variable (MaxT, MinT, SOI etc.), θ_k represents a coefficient of a predictor variable ($1 \leq k \leq n$), and ε is the error (noise). For each month, a linear regression model was created. The outputs of each model were then aggregated to form the data series between January 2005 and December 2015.

CFS-GA is an optimization method for selecting input features [151]. CFPS-GA is an optimization method for selecting climate and network parameters using a standard GA, in which all model characteristics are shown on the same chromosome. CFPS-HGA is the optimization method for selecting climate and network parameters using a hybrid GA [153]. MAE, RMSE, r, R^2, and IPE values for each method are shown in Table 5.12. The proposed approach produced lower MAE and RMSE compared to the other approaches. In addition, better correlation and determination of coefficient values were obtained when compared to all other models. Finally, the calculated IPE exposed was closer to zero as compared with climatology, MLR, CFS-GA, CFPS-GA, and CFPS-HGA.

Table 5.12 MAE, RMSE, r, R^2, and IPE values for climatology, the alternative approach, and the proposed approach respectively.

Model	**Years**	**MAE**	**RMSE**	**r**	**R^2**	**IPE**
Climatology	2005-2015	147.215	200.323	0.689	0.475	0.984
MLR	2005-2015	149.028	204.592	0.684	0.468	0.999
CFS-GA (Section 4.2)	2005-2015	130.217	178.815	0.763	0.583	0.845
CFPS-GA	2005-2015	132.810	171.339	0.784	0.614	0.820
CFPS-HGA (Section 5.3)	2005-2015	103.813	141.670	0.859	0.738	0.634
CFPS-TLO (proposed approach)	2005-2015	65.305	96.917	0.938	0.880	0.395

Calculation of the SS of the proposed method against the alternatives was based on the RMSE statistical measurement. The results are shown in Figure 5.15. The highest difference was recorded with MLR followed by climatology, CFS-GA, CFPS-GA, and CFPS-HGA. This demonstrates the applicability of the proposed approach against alternative methods.

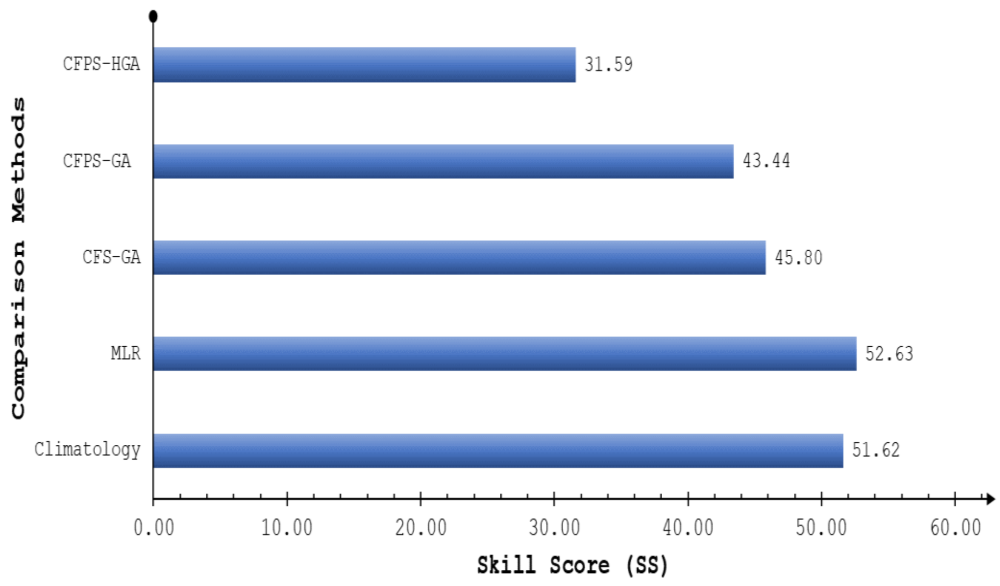

Figure 5.15 Skill score of the proposed approach against alternative forecasting and optimization methods.

To facilitate further analysis, boxplots of absolute error values obtained in each method were compared to actual rainfall amounts between January 2005 and December 2015 (see Figure 5.16). The highest error ranges were obtained with MLR and climatology, then boxplot ranges started decreasing with each of the other methods. The highest range of outliers was obtained with MLR models. The two approaches that are based on the hybrid GA (CFPS-HGA and CFPS-TLO) had lower errors than the others. The ranges of the boxes in both methods (CFPS-HGA and CFPS-TLO) were relatively short. This shows that there is a consistency between predicted and actual values. The proposed approach had the lowest error ranges, as 75% of the values were lower than 88.59 mm monthly. Some 50% were less than 39.73 mm. One outlying value for February 2007 was high (around 455 mm error), where the actual value reported at the station was 1193.04 mm and the predicted value was 747.93 mm. This value was the closest to actual among the alternatives (climatology, 625.60; MLR, 734.53; CFS-GA, 720.71; CFPS-GA, 687.0; and CFPS-HGA, 666.85). This also describes the highest RMSE obtained for February (see Table 5.11).

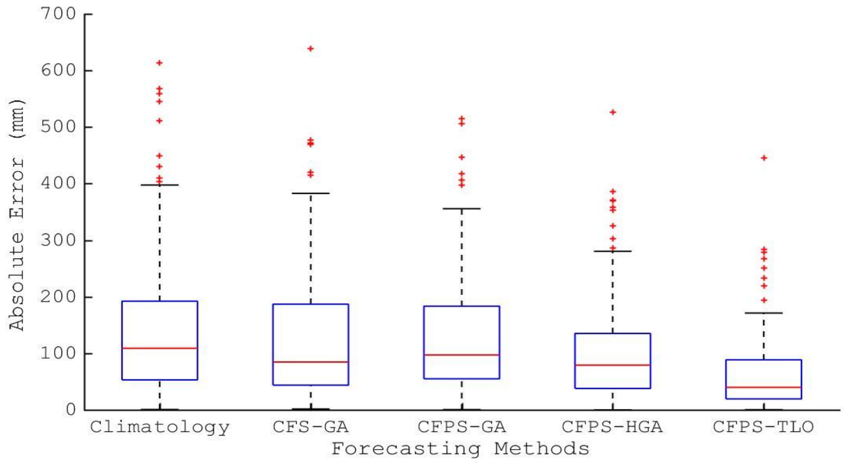

Figure 5.16 Boxplots for errors in rainfall between January 2005 and December 2015 for each method.

In addition, the models generated through the proposed method were compared to the Australian Community Climate and Earth-System Simulator (ACCESS-S1) forecasts. ACCESS-S1 is the new seasonal forecasting system for rainfall prediction across Australia. It is based on the UK Met Office's GloSea5-GC2 seasonal prediction system, and consists of 11 ensemble members (e00, e01, …, e10). ACCESS-S1 forecasts are released with 60 km grid resolution and up to six months in advance. Predictions are released at the beginning of each month. ACCESS-S1 was reported to be an operational forecasting system in early 2018 [15]. A set of hindcasts for ACCESS-S1 can be found on the National Computational Infrastructure (NCI) website (http://nci.org.au/) [154]. These hindcasts represent 23 years of testing between 1990-2012. Based on our proposed models, eight years of common testing data can be compared to ACCESS-S1 hindcasts (2005-2012). ACCESS-S1 values generated for grid points near selected stations were collected from NCI website to be compared with the ANNs based models used in this research. MAE, RMSE, r, and R^2 for each method are shown in Table 5.13. Better results were produced, in terms of the statistical measurements, with the proposed method when compared with ACCESS. Reasonable correlation was obtained with the ACCESS forecasting system. Finally, the skill of CFPS-TLO against ACCESS forecasting system over the eight years was found to be 45.75%. As shown in Figure 5.17, lower differences between observed and predicted values were encountered with the proposed method as against ACCESS. A total of 75% of the errors were between 0 and 206 mm with ACCESS. On the other hand, 91 out of the 96 observation errors with models generated through the proposed method ranged between 0 and 200 mm (except for the five outlying values).

Table 5.13 MAE, RMSE, r, and R^2 and IPE values for ACCESS and the proposed method respectively.

Model	Years	MAE	RMSE	r	R^2
ACCESS [8]	2005-2012	135.793	185.451	0.785	0.616
CFPS-TLO (proposed approach)	2005-2012	67.626	100.605	0.944	0.891

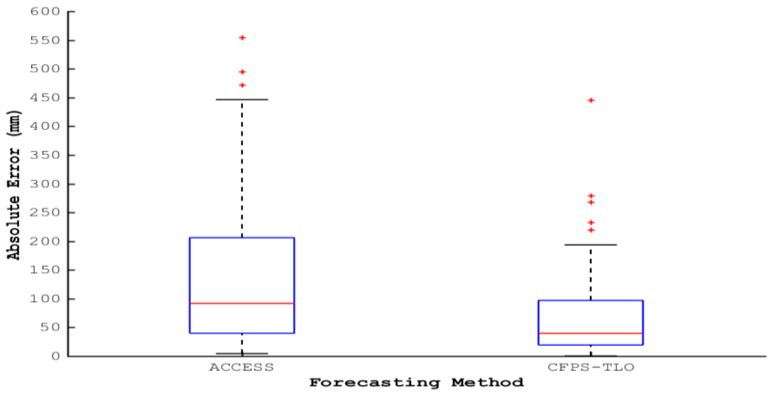

Figure 5.17 Boxplots for errors in rainfall between January 2005 and December 2012 for ACCESS and CFPS-TLO.

MAE represents the average difference between a predicted and an actual value in the sample dataset. The trend of rainfall varies between months in the selected location as shown in Figure 5.3. For example, if the actual rain was 1000 mm and the predicted value was 900 mm for a month with an average of 600 mm, this 100 mm difference is not considered high and the results are a sufficient indicator that unexpected conditions will occur. However, if the actual rain was 170 mm and the predicted values was 70 mm for a month with an average of 60 mm, this 100 mm difference is considered high. This implies that users were not already notified about unexpected conditions. As regards the RMSE, outliers increase the measured value. The RMSE of February is different to values in other months. This is because no value was predicted in 2007. With RMSE, the penalty is higher than with MAE.

As shown in Figure 5.14, the trend of rainfall was captured in most of the cases. This is helpful for the users in need of rainfall forecasts. If the model output displayed values with very different to the expected type of rain, users can go back to earlier decisions. On the other hand, users can keep to their original plans if no deviations in rainfall are shown in predictions. Confidence and trust in these forecasts can be interpreted based on the statistical measurements generated for each model and the expected rainfall trend.

The proposed method aims to select the best optimal solution. The types of forecasts targeted in the study were local forecasts. They are spread over a small grid area close to the weather stations from which data were taken. The forecasts based on ANNs are data dependent. There are thousands of weather stations in Australia. In addition, many farms and producers have their own historical recordings. Rather than applying all of the available options for each location, data can be fed into the proposed algorithm so that it can select the best characteristics automatically for each month in each location. After collecting local variables, this approach can be applied to develop alternative forecasting systems for local areas. In addition, it can be deployed in various areas and machine learning based approaches. Furthermore, it can be applied in ensemble learning, where ensemble generation is dependent on multiple key characteristics including classifiers and the fusion method.

The input features and network parameters selected in this study were constrained to match previous studies for the purposes of comparison. It was clearly noticed that there is a need to increase the range of multiple elements, especially the number of neurons and activation functions. Although the method holds the ability to perform better against alternative approaches, it is considered time-consuming as it takes a long time to optimize the model's characteristics (seven hours). For each possible solution in the search space, another search space is created and optimized. This time is for optimisation, but only once the model is optimised, and a very short time is required to release forecasts (ms). Various methods could be considered to lower the amount of time consumed in optimization, including multi-threading, and feature or parameters reduction (for example, only three training algorithms were selected with different types of monthly rainfall: Levenberg-Marquart, scaled conjugate, and Fletcher-Powell conjugate gradient BP).

5.4.5 Summary

In this study, a new approach for selection and optimization was proposed. It selects and optimizes model characteristics through a multi-level optimization strategy incorporating multiple EAs. It was used to select climate features and ANN parameters in a rainfall prediction application. The proposed method was compared to multiple optimization and forecasting methods, including climatology, climate feature selection using a standard GA, climate and network parameters selection using a standard GA, and climate and network parameters selection using a hybrid GA. MAE, RMSE, r, R^2, and IPE statistical measurements were used to assess the performance of each method. In addition, SS measurement was applied to compare the models' performance. The aggregated time series of models optimized using the proposed approach produced figures of 65.305, 96.917, 0.938, 0.880, 0.401 for MAE, RMSE, r, R^2, and IPE respectively. SSs of 51.62%, 45.80%, 43.44% and 31.59% were recorded when compared to climatology, climate feature selection using a standard GA, climate and network parameters selection using a standard GA, and climate and network parameters selection using a hybrid GA respectively. Finally, eight years of the testing dataset were compared with the Australian Community Climate and Earth-System Simulator (ACCESS-S1), and better accuracy was also recorded.

This approach can be used in ensemble generation in which the ensemble performance is dependent on the combination of classifiers and fusion method. It can be directly applied on stacking fusion, where network parameters could be optimized based on selected ensemble classifiers. Furthermore, the parameters of the proposed hybrid GA can be widely investigated and analysed.

5.5 Chapter Summary

In this chapter, three novel approaches have been proposed and examined. Two of those approaches can be used to generate rainfall forecasting models for locally specified regions. In addition, a new hybrid GA that combines GA with PSO was proposed. The hybrid GA was incorporated in order to select input features and neural network parameters. It was demonstrated that the proposed approaches performed better in forecasting monthly rainfall when compared to alternative approaches and the ACCESS. The superior performance was obtained by using climate input features and parameters selection-two level optimization. It was also found that the ANN models have better skill than existing approaches. These approaches can be applied and extended to multiple weather stations and locations. In addition, it can be applied in various applications that use ANNs and other machine learning algorithms. In the next chapter, ensemble modelling techniques in rainfall prediction are proposed and evaluated.

Chapter 6 Rainfall Forecasting using Ensembles of Neural Networks

In this chapter, a novel ensemble to forecast monthly rainfall for a selected location in Queensland, Australia, is proposed. Multiple ensembles of ANNs were developed to estimate the amount of monthly rainfall for Innisfail, Queensland. In addition, four fusion methods: average fusion, ANN learning fusion, lowest error based ANN fusion, and ANN based PSO fusion were proposed and evaluated. These models were compared with alternative models and climatology, and the results produced by ensemble generated outlooks were more accurate. Among the ensembles with four fusion methods, an ensemble of FFNNs using resilient BP algorithm and PSO produced the highest accuracy (166.71 mm RMSE).

The chapter is organized as follows. Section 6.1 provides a brief introduction to the ensembles of ANNs used in rainfall forecasting. Section 6.2 discusses multiple fusion methods. Section 6.3 describes the dataset used. In Section 6.4 and Section 6.5, experimental results and comparative analysis are detailed. Section 6.6 draws the conclusions of the chapter.

6.1 Introduction

Each ANN model has different generalization capabilities. An advantage with ANN is that they have the ability to approximate non-linear relationships. To improve forecasts, ensemble techniques have been recently introduced in hydrology [94].

Considerable attention has been paid to combining multiple ANNs for weather forecasting, where different numbers of individual classifiers were aggregated in order to forecast [88, 100]. Ensembles perform better with diverse classifiers [87]. A successful ensemble can be easily identified if errors are quite low in its individual classifiers [92].

Parts of this chapter appeared in A. Haidar and B. Verma, "Learning based fusion in ensembles for weather forecasting," In *13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery* (ICNC-FSKD), 2017, pp. 72-78.

Gadgay and Kulkarni combined a number of ANNs using a weighted average to forecast rainfall for Bangalore, India [89]. Results were compared with single ANNs in which better accuracy was reported. Nagahamulla, Ratnayake, and Ratnaweera deployed an ensemble composed of different ANNs (BPNN, RBFNN, and general regression NN) to forecast rain in Colombo, Sri Lanka [99]. The implemented ensemble had better accuracy when compared to single networks with a lack in forecasting extreme rainfall values.

Data fusion methods have been used to combine the output of different models to form the ensemble. In this study, a new approach is proposed for generating ensembles of ANNs to forecast monthly rainfall for local areas in Queensland, Australia. The aim is to investigate the effect of different ensemble models and fusion methods on the generated forecasts.

6.2 Proposed Approach

Two strategies were utilized to find the best model for rainfall forecasting. The first depends on coding the ANN, while the second aims to benefit from existing ANNs and to improve prediction performance.

ANNs attempt to decrease the error between observed and predicted values when applied to deterministic forecasting problems. The network is usually developed to make predictions after learning from previous data. ANNs have different types of models and algorithms, each with different specifications, and they follow certain criteria through the training, validation, and testing phases. Each network varies in terms of the model, number of neurons, number of hidden layers, training algorithms, and activation functions. Hence, each model has different generalization capabilities.

To improve the accuracy of forecasts, we propose a novel approach that is based on combining multiple ANN models using different fusion methods. All the proposed individual topologies were based on FFNN architecture. The architecture of the designed networks consisted of only three layers (one input layer, one hidden layer, and one output layer). Figure 6.1 shows an overview of ensemble architecture.

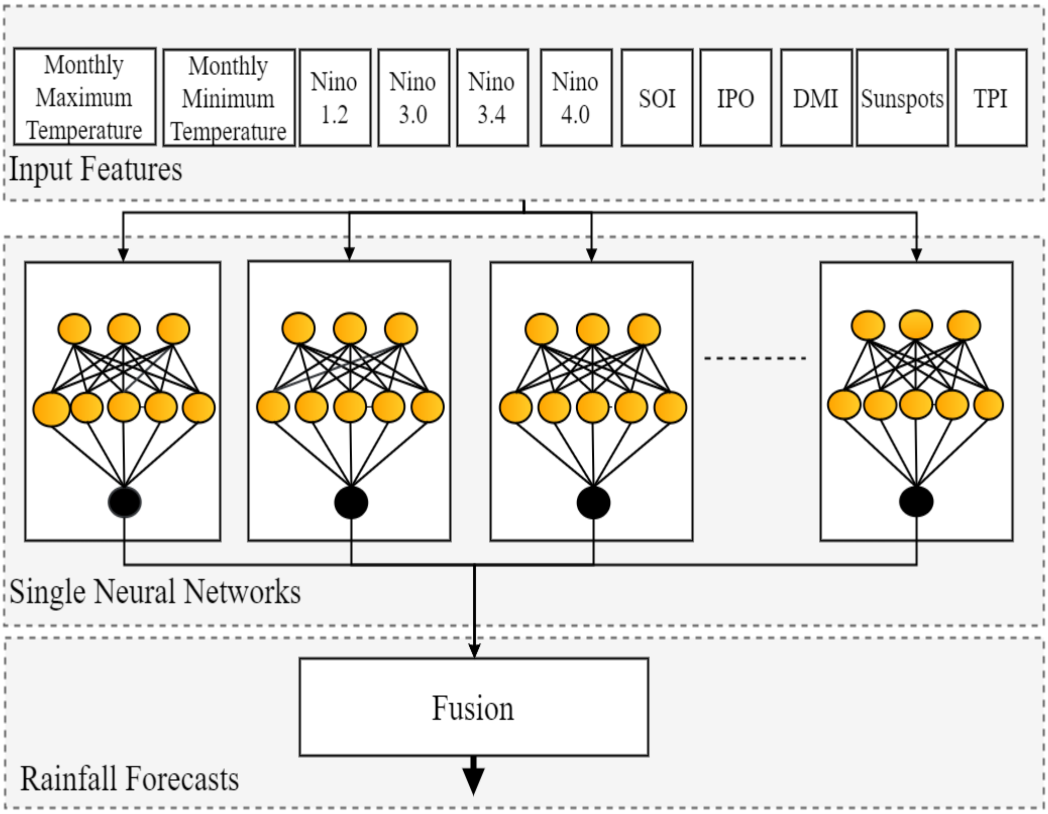

Figure 6.1 Proposed approach.

Multiple parameters can be given as input to the network. They are added to weights and sent to hidden neurons for processing. Connection weights, activation functions, activation functions, and training algorithms determine the output of each neuron.

Four ensemble methods were proposed in an attempt to study the effect of various models on the results of rainfall prediction. The ensemble architecture is shown in Figure 6.1. Each ensemble consisted of a number of single ANNs combined using various data fusion methods.

- **Average Fusion Model (A-FM):** The A-FM is a data fusion method that can be applied with various machine learning algorithms. It is based on averaging the output of individual models. Following this technique, the same priority is given for each topology in the ensemble.

- **Neural Network based Fusion Model (NN-FM):** In this model, a single ANN is used to combine the outputs of previously trained networks. The characteristics of the network are well defined, including number of neurons, training algorithm, and activation functions.

- **Lowest Error-Neural Network based Fusion Model (LENN-FM):** For this model, a group of the developed single ANNs were selected and combined, based on their previous performance. The selection criterion for the models was RMSE. The selected fusion method was an ANN with the same parameters as the NN-FM. Rather than selecting all of the members to form the input of the network, members who had reliable performance are designated as input for the fusion network. The aim of this method is to study the effect of ensemble size and ensemble members on the overall performance of the model. Hence, rather than selecting all of the networks as in the NN-FM, the networks with highest performance were aggregated.

- **Neural Network based PSO Fusion Model (NNPSO-FM):** In this model, an ANN was developed in order to combine the output of all developed single FFNNs. The input dataset consisted of training and validation of dataset output from each model, while the test dataset consisted of each network output on the test dataset. The size of input features is determined based on the number of single ANNs developed. This method is similar to the method in part NN-FM, but instead of selecting all the ANN characteristics PSO was used to select some of the characteristics. The performance of an ANN varies based on the selection of its parameters. A varying number of neurons, initial weights, and training algorithms would affect the overall results. Several training algorithms are usually used with forecasting applications. Hence, those functions were selected statically. PSO was assigned to select the best number of neurons for the ensemble method with each of the selected training algorithms. Each particle in the population represents a possible best solution. Following this optimization, the chromosomes in each iteration have the ability to start with different initial weights. In other words, initial weights were searched to find the best combination to start with. Each particle trained model in each population was stored to guarantee selection of the best model.

6.3 Data (Case Study 4)

Climate attributes were collected with different starting and ending points. These values were arranged so as to have the same duration of time. Datasets came from between January 1908 and December 2015. Data were pre-processed as mentioned in Chapter 3 (Section 3.1.4). It should be mentioned that all the data were used as input in this study (on data manipulation for each month, see Section 3.1.4).

6.4 Experiments and Results

MATLAB was used to perform all of the experiments. Levenberg-Marquardt was used as the learning algorithm for each of the base networks (ensemble components). Hyperbolic tangent sigmoid activation function (tansig) was used between input-hidden and hidden-output layers for each of the single and fusion ANNs. Different sets of initial weights were used while training the networks. This led to different prediction abilities because of the connection weights. Ten single ANNs were trained and used as possible ensemble components.

Single models varied based on the number of neurons and set of initial weights. The same input data were given for each network input layer in each single network model. The available input data were partitioned into sets: 75% for training, 15% for validation, and 10% for testing. The testing portion represented 10 years and 10 months. The main task was to combine abilities of the single networks so as to release better forecasts. All the single networks were run for 1000 epochs, and validation check was used to avoid overfitting.

The single ANNs were combined using the four proposed fusion models: average (A-FM), neural network (NN-FM), networks with lowest errors which were aggregated based on another network from two to ten (LENN-FM), and an optimized ANN based on six training algorithms (NNPSI-FM). In the A-FM model, the outputs of all ANNs were averaged. The ANN used for combining the single networks in NN-FM and LENN-FM used resilient BP as the learning algorithm. The hidden layer consisted of eight neurons, and the activation functions were tansig.

Finally, in the fourth model (NNPSO-FM), ensembles were created using an optimized ANN based on PSO. PSO parameters are shown in Table 6.1. The ANN used for fusion was optimized based on number of neurons. At the same time, random initial weights

Table 6.1 PSO details.

Parameter	**Value**
Initial population	10
Population type	Double
Iterations	100
Range	[4, …, 40]
Fitness function	Three-Layered Feed Forward Network
Scalar value	RMSE

were given. The fourth model (NNPSO-FM) was used six times with each of the following training algorithms.

1. Levenberg-Marquardt BP
2. Resilient BP
3. Bayesian regularization
4. Gradient descent BP
5. BFGS quasi-Newton BP
6. Fletcher-Powell conjugate gradient propagation

MAE, RMSE, and r were used to measure the performance of each model. The closer MAE and RMSE were to zero, the better the forecasts. The range of Pearson correlation (r) varies between -1 and 1. Zero means there is no association between the outlooks and observed values. Further details about the statistical measurements are given in Section 3.2. The calculated measurements of single ANNs are shown in Table 6.2, while results for each of the proposed methods are shown in Table 6.3, Table 6.4, Table 6.5, and Table 6.6 respectively.

Table 6.2 MAE, RMSE and r values of single neural networks.

Number	Neurons	MAE	RMSE	r
1	33	141.26	184.92	0.77
2	23	148.99	192.79	0.749
3	7	149.08	193.44	0.747
4	38	146.87	194.46	0.742
5	46	157.53	195.43	0.762
6	25	149.13	197.35	0.711
7	23	148.25	198.04	0.711
8	32	157.53	198.66	0.74
9	16	154.89	199.46	0.716
10	26	154.21	201.42	0.709

Table 6.3 MAE, RMSE and r for ensemble generated using average fusion model (A-FM).

Ensemble size	MAE	RMSE	r
10	146.74	189.91	0.755

Table 6.4 MAE, RMSE and r for ensemble generated using a specified neural network as fusion model (NN-FM).

Ensemble size	MAE	RMSE	r
10	131.84	181.92	0.753

Table 6.5 MAE, RMSE, and r for ensemble generated using lowest error-neural network fusion model (LENN-FM).

Ensemble size	MAE	RMSE	r
2	123.66	171.4	0.786
3	136.41	188.77	0.763
4	134.1	178.74	0.772
5	129.43	182.74	0.78
6	132.69	188.4	0.755
7	132.66	181.42	0.756
8	137.4	176.99	0.774
9	122.26	173.29	0.783

Table 6.6 MAE, RMSE, and r for ensembles generated using neural network based PSO fusion model (NNPSO-FM).

Training function	Neurons	MAE	RMSE	r
Levenberg-Marquardt BP	5	128.59	178.79	0.769
Resilient BP	**17**	**119.73**	**166.71**	**0.7995**
Bayesian regularization	4	125.10	176.65	0.7721
Gradient descent BP	19	140.14	183.90	0.762
BFGS quasi-Newton BP	8	119.20	173.96	0.7893
Fletche-Powel conjugate gradient BP	31	137.18	178.53	0.7786

The average yearly amount of precipitation encountered over the selected area was approximately 3553.0 mm. Multiple extreme events appeared in the testing dataset. The RMSE of single networks ranged between 201.42 mm, the highest, and 184.92 mm, the lowest. Reasonable correlation was obtained with all of the models. The single ANN that resulted in the highest performance consisted of 33 neurons in the hidden layer. Nine of the ten models had similar a RMSE ranging between 192.79 mm and 201.42 mm. This did not mean that similar prediction performance was shown over the testing series. Because of this, ensemble methods were applied to fuse prediction diversities.

As shown in Table 6.3, averaging all networks decreased the prediction accuracy. The reason beyond this increase in RMSE is the same priority was given to single models. The single models that failed in predicting the amount of rain affected the performance of the overall combination. This reveals that ensemble based on average is not effective in this type of forecasting applications.

With 181.92 mm as RMSE, the ANN based fusion model with a selected number of neurons was more accurate than the average fusion model A-FM (see Table 6.4).

Combining some of the members produced different accuracies (Table 6.5). The errors ranged between 188.4 mm, the highest, and 171.4 mm, the lowest. Two obtained values produced lower accuracy than single models. Considering time and memory, aggregating a small number of individual networks produced the best performance when using lowest error- neural network fusion method.

The ANN and PSO based fusion model (NNPSO-FM) using a resilient BP training algorithm had the best performance in terms of MAE, RMSE, and r as compared to other single and ensemble models (Table 6.6). An ensemble consisting of 10 ANNs as ensemble components and a three-layered FFNN with 17 neurons in the hidden layer produced the highest accuracy (MAE, 119.73 mm; RMSE, 166.71 mm; r, 0.799). A reduction of 18.21 mm was obtained when compared to the best single model. This decrease in RMSE may be caused by the ability of a single model to capture the trend, but not to predict the amount accurately. An ensemble could be effective when low and high predictions are merged.

All of the training algorithms tested with NNPSO-FM increased the prediction performance with a high processing time for optimization using the gradient descent training algorithm. Based on Table 6.3, Table 6.4, Table 6.5, and Table 6.6, ensembles of ANNs with the right characteristics can improve the accuracy of rainfall forecasts.

6.5 Comparative Analysis

The best ensemble was compared with other approaches that have been used to forecast precipitation. Climatology is a basic forecasting system based on averaging the time series values, then forecasting new monthly values based on these averages. It has been used as a comparison reference for GCM forecasting applications [105].

The official rainfall estimations released by the BOM in Australia are based on GCMs. Bagging is an ensemble method in which a set of classifiers are trained using randomly sampled data from the original dataset, and then combined using average [87, 155]. It has been used as a comparison approach in other weather applications [100]. By means of bagging, single networks are trained based on randomly sampled datasets; then outputs are averaged. Ten single ANNs were trained on the basis of randomly sampled data; then outputs from all networks were averaged. Table 6.7 shows MAE, RMSE, and r values for the proposed model against climatology and bagging. The best model generated by the proposed methods (Ensemble-PSO-rp) was more accurate than climatology and bagging in terms of MAE, RMSE, and r. Most of the developed single and ensemble models performed better than climatology. This shows that machine learning based approaches can generate accurate monthly rainfall forecasts. Bagging had a similar performance to climatology.

Table 6.7 MAE, RMSE, and correlation coefficients values for climatology, bagging, and the proposed best ensemble (Ensemble-PSO-rp).

Model	MAE (mms)	RMSE (mms)	r
Climatology	145.65	196.71	0.703
Bagging	149.90	196.14	0.750
Ensemble-PSO-rp	119.73	166.71	0.7995

Figure 6.2, Figure 6.3, and Figure 6.4 represent the forecasted values against actual values obtained by each approach (climatology, bagging, and Ensemble-PSO-rp) for the testing dataset (130 months). Climatology shows the same values for each month through the dataset, as shown in Figure 6.2. Bagged ensemble generated outlooks are shown in Figure 6.3. It is clear that the best ensemble (Ensemble-PSO-rp) was able to capture rainfall pattern with a lack of forecasting accurately rainfall amounts at specific times (see Figure 6.4). The proposed ensemble was able to forecast the pattern of rainfall, and an indication of unusual events was obtained in different parts of the time series (2006, 2010, 2014).

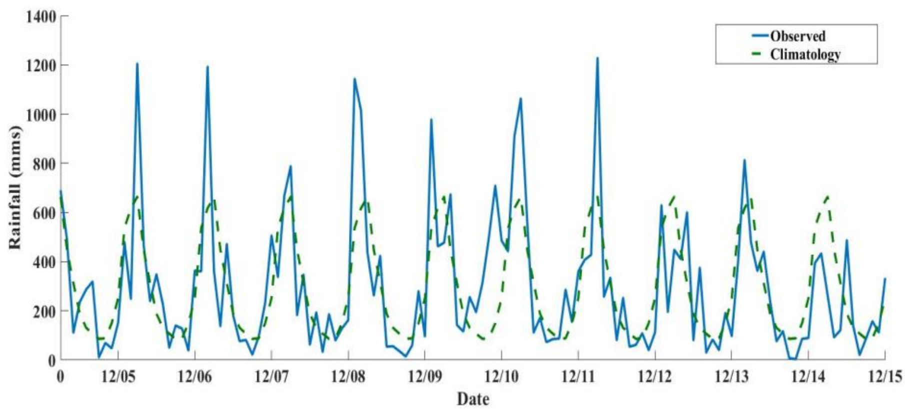

Figure 6.2 Actual rainfall values compared to Climatology between March 2005 and December 2015.

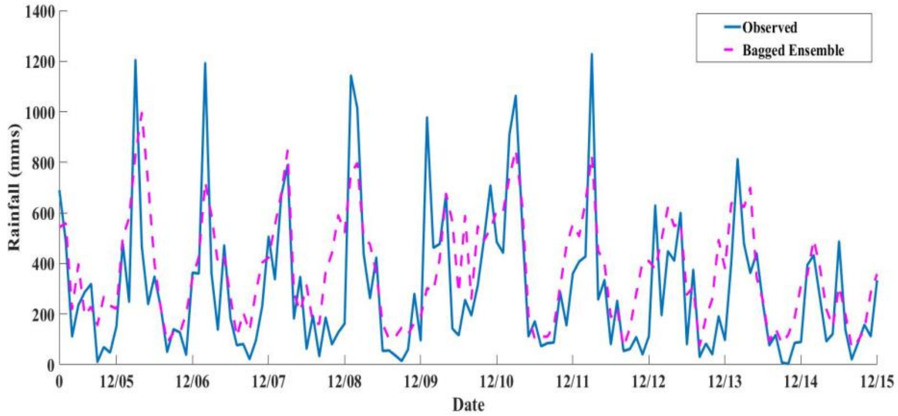

Figure 6.3 Actual rainfall values compared to Bagging between March 2005 and December 2015.

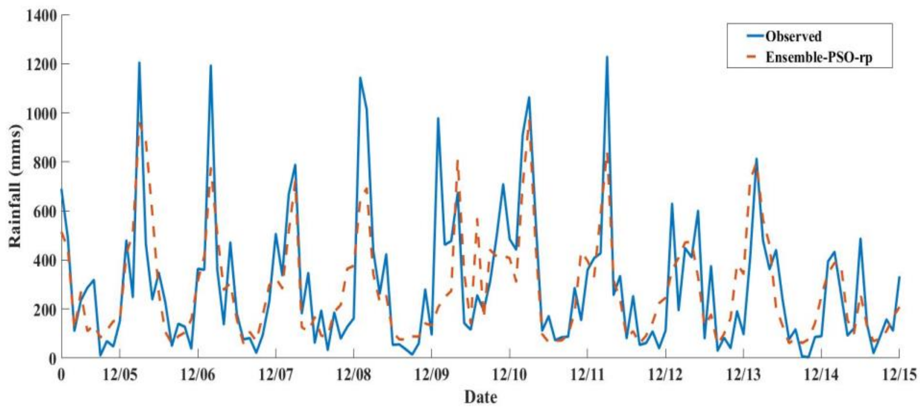

Figure 6.4 Actual rainfall values compared to the best proposed ensemble method (Ensemble-PSO-rp) between March 2005 and December 2015.

6.6 Chapter Summary

In this chapter, novel ensemble models of ANNs developed to forecast the amount of rainfall in Innisfail, Queensland, Australia, were examined. The models consisted of the same ANN type with a different number of neurons in hidden layers and learning algorithms. In addition, four fusion methods were proposed and evaluated. The results have shown that an ensemble with PSO fusion is able to produce better monthly rainfall forecasts. An ensemble consisting of ten single ANNs and an ANN based PSO optimization fusion model was the most accurate (MAE, 119.73 mm; RMSE, 166.71 mm; r, 0.799) among the other single and combined models.

The use of ensemble models is promising and can be efficient if the fusion method is designed properly. Although better accuracy was recorded in comparison to other approaches, the ensembles were not able to forecast rainfall accurately at some stages. The ensemble components were trained using values from every month.

It was observed that this way of adding input features to the network did not produce the best accuracy because of seasonality in data. Although training all of the values in one network would decrease the computation cost and complexity, performance would be affected.

This study was a first step towards enhancing our understanding of using ensembles of ANNs in rainfall prediction. The limitations of this approach will be investigated in the next chapter. The network models will be trained on each month dataset, and the ensemble components will be extended and further optimized in order to enhance performance.

Chapter 7 Optimized Neural Network Ensembles in Rainfall Forecasting

In this study, an ensemble of ANNs was created and optimized in order to estimate monthly rainfall. The ensemble utilizes single ANNs as components and combines them using an ANN fusion method. A novel ensemble components selection approach was used. Ensemble components were selected based on a hybrid GA that combines a standard GA with a PSO optimization algorithm (as was proposed in Chapter 5). Various statistical measurements were calculated to assess the accuracy of the proposed ensembles against single ANNs, climatology, and ensembles generated through alternative selection approaches. A better performance was obtained with the proposed ensembles as compared to alternative models.

7.1 Introduction

The ensemble systems were combined with the diverse classifiers in an attempt to enhance the classification performance. Maqsood, Kahn, and Abraham used multiple ANNs to predict daily weather attributes such as temperature, humidity, and wind [88]. Nagahamulla, Ratnayake, and Ratnaweera developed an ensemble of multiple ANN topologies to forecast rainfall for a location in Sri Lanka [99]. These models were compared to single networks and were more accurate in prediction. Saba, Rehman, and AlGhamdi combined a MLP and a RBFNN to overcome the limitations of each model [101]. Results were compared to single MLP and RBF models. The error recorded was less than both single networks.

Several EAs such as GA and PSO have been applied in forecasting applications. These algorithms were applied in order to select the best features that would produce the highest accuracy among the proposed possible solutions (as shown in previous chapters). In addition, EAs were applied to find optimal ANNs characteristics, such as weights,

Parts of this chapter appeared in: A. Haidar, B. Verma, and T. Sinha, "A novel approach for optimizing ensemble components in rainfall prediction," in *IEEE Congress on Evolutionary Computation* (CEC), 2018, pp. 1817-1824.

connections and architecture [139]. PSO has been used to find the best ANN architecture and weights [139]. Jin, Huang, and Zhao utilized PSO to develop different BPNNs, and an ensemble of the developed models was generated [98]. Jiang and Jiansheng used BP, GA, and simulated annealing to form a model that gives the optimal weights and connections for forecasting rainfall [77]. EAs have also been used with various machine learning algorithms to predict rainfall for different time-frames. Soe and Kim applied a GA that uses SVM and K-Nearest Neighbour (KNN) in attempt to predict whether or not there would be a heavy rainfall in the next three hours [156]. Three years with daily values were used to perform the study. The fitness function of the GA was either SVM or KNN. The results showed that selected subsets of features produced lower results than using all of the features. In addition, Nagahamulla, Ratnayake, and Ratnaweera introduced a GA and KNN algorithm to select the best ensemble members for forecasting rainfall for two locations in Sri Lanka [100]. The best results were obtained by the GA. This supports the idea of using portions of the available features to maintain better accuracy.

To constitute an ensemble, three constraints are usually targeted: individual components, number of individuals, and the fusion method. Selecting a subset of the available classifiers to form the ensemble may sometimes reveal superior results as against selecting all of the available classifiers [157, 158]. Ensemble selection is the process of selecting classifiers in the ensemble, and deciding on its final size [159]. Choosing the best sub-groups of classifiers is done through different mechanisms, including trial and error, and evolution based algorithms. If using trial and error, a lot of time is needed, and the cost of selecting the best classifiers is expensive.

This chapter makes the following original contributions.

- A new ensemble approach for building weather forecasting models using specific locations and various forecasting lead times is proposed.
- A new algorithm for selecting ensemble components is proposed.
- A new components selection approach that uses neural fusion and a hybrid GA is proposed.

The chapter is organized as follows. Section 7.2 describes the proposed ensemble components selection approach. The collected weather variables used in the study are described in Section 7.3. Section 7.4 discusses the experimental results and comparative analysis. A summary and recommendations for future work are presented in Section 7.5.

7.2 Proposed Approach

An overview of the proposed approach is shown in Figure 7.1.

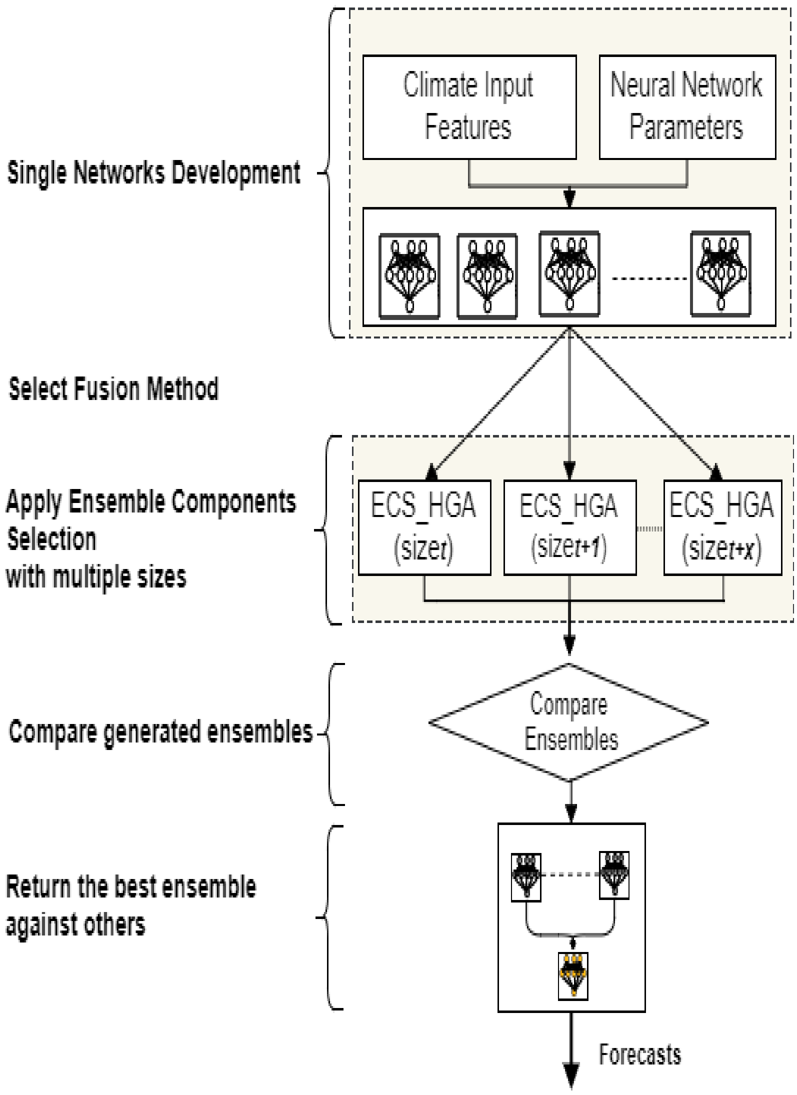

Figure 7.1 Overview of the proposed approach.

To develop an ensemble model, three steps are followed: creating possible ensemble components, selecting the fusion function, and combining the ensemble components. The study focussed on ensemble components selection, where an ANN is used as the fusion method. Four main steps were followed to select the best ensemble:

- First, single ANNs are developed.
- Second, the ANN fusion method for rainfall prediction is developed.

- Third, the ensemble component selection approach is applied multiple times with different maximum ensemble sizes (between t and x as shown in Figure 7.1: $t \geq 2$, $x \leq n$, n is the number of single ANNs).
- Fourth, the generated ensembles in each size are evaluated against each other and the best ensemble is selected.

7.2.1 Single Neural Networks Development

Changing network parameters would lead to a different training process, and the generation of diverse results. These parameters include: initial weights, training algorithm, activation functions between layers, and neurons in the hidden layers. Therefore, in an attempt to increase the diversity of single models, various options were used when developing single ANNs. Thus, a pool of possible single networks can be listed as follows.

$$\rho = [\delta_1, \delta_2, \dots, \delta_n] \tag{7.1}$$

n is the number of developed ANNs, δ_i is the network identifier, $1 \leq i \leq n$.

7.2.2 Neural Network Fusion

Multiple fusion methods were utilized to combine single ANNs. A fusion method based on another ANN was used to combine the selected networks. The outputs of the single networks were added as input to the fusion method. Forecasting rainfall at time t based on ANNs ensemble could be described as follows.

$$rain_t = \left(f \left(\delta_1(D_{t-g}), \delta_2(D_{t-g}), \dots, \delta_m(D_{t-g}) \right) \right) \tag{7.2}$$

f is the fusion function, m is the ensemble size ($m \leq n$), D_t is the input climate features vector at time t, δ_i is a trained ANN in which $1 \leq i \leq n$, g is the lag time.

7.2.3 Ensemble Components Selection using a Hybrid Genetic Algorithm

The aim of this approach is to determine the combination of ensemble components β.

$$\beta \subseteq \rho \tag{7.3}$$

$$\beta = \left[\delta_{e_1}, \dots, \delta_{e_m}\right] \tag{7.4}$$

In this equation:

$1 \leq e_i \leq n$,

m is the number of networks selected in the ensemble,

$2 \leq m \leq x$

x is the maximum size the ensemble can have,

$x \leq n$,

n is the number of the networks in a pool,

so that when β is given to the ANN fusion method f, a lowest error err is obtained.

$$err = f(\beta) \tag{7.5}$$

Selecting subsets of the created single networks may result in better accuracy when combined. For this reason, x was not selected to n in this approach. The maximum size x was also modified when evaluating the approach.

To select β, various methods can be followed. EAs have been utilized to select ensemble members. GA is a metaheuristic algorithm inspired by natural evolution. With GAs, possible combinations interact together to find the best solution. GA has been widely applied in selection and optimization problems. Rather than selecting classifiers randomly or based on performance, GA can be incorporated to select ensemble components. Having an average or majority voting fusion method in the fitness function will ensure that elitism succeeds. But when an ANN is incorporated as a fusion method, the performance of the same chromosome in two consecutive iterations may vary. The main reason for this variation is the random initial weight selection that occurs every time the network (fusion method) is trained. Because GA does not remember the performance of previous iterations, the new elites may be less accurate when trained with different initial weights. To overcome this limitation, a hybrid GA that utilizes GA operations and selection criteria with PSO functionality (saving global best) was incorporated (see Figure 7.2). The steps followed in selecting ensemble components are shown in Figure 7.3.

The GA uses chromosomes (parents) from previous iterations to create new chromosomes (offspring) with higher performance or better accuracy. That is, the GA relies on selecting the best chromosome in the final population. The GA does not remember the accuracy of previous generations, but with elitism the best chromosomes are guaranteed to transfer to the next iteration. If better solutions are found, these chromosomes are selected as the best and transferred to next iteration.

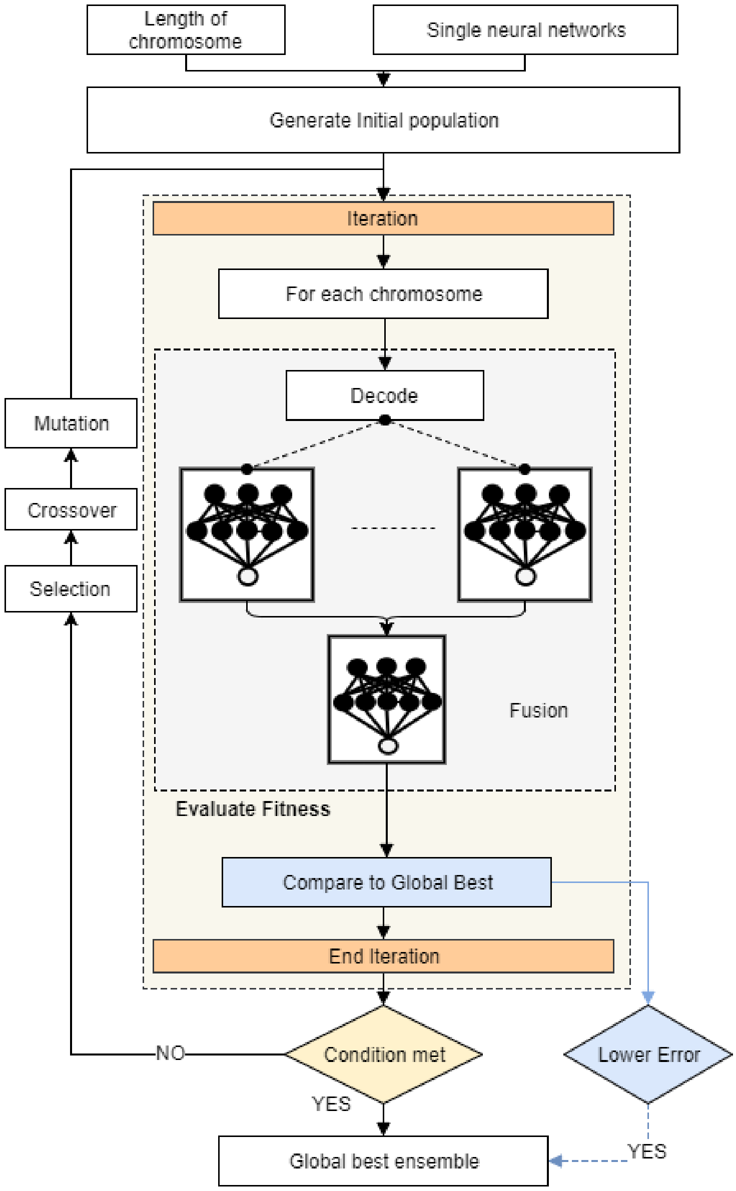

Figure 7.2 Ensemble components selection using a hybrid genetic algorithm.

PSO is metaheuristic algorithm that was introduced by Eberhart and Kennedy [138]. Like the GA, it is population based and elements in a population are called particles. Particles traverse the search space in an attempt to discover the best global solution. In each iteration, the particle's performance is compared to its previous performance, and to the global best solution. If a lower error than the global best is produced, the selected particle becomes the global best solution for the problem being optimized. This functionality was combined with GA operations so that each chromosome in the GA will be compared to a global best. The global best will be returned when condition is met. The main concept of the GA was modified so that the best chromosome through all iterations will be selected.

By these means, the ensemble that reveals the highest accuracy will be the guaranteed end product.

Algorithm: Ensemble Components Selection using Hybrid Genetic Algorithm (ECS_HGA)

Input: pool ρ, ensemble size x
Output: selected best chromosome
Initialize global best chromosome $\beta \leftarrow$ empty
Initialize β error $err\_\beta \leftarrow$ maximum error.
Generate random initial population
for each chromosome ∂ in population
    Decode ∂ to genes $\delta_{e_1}, \ldots, \delta_{e_m}$ $(m \leq x)$
    Train fitness function based on decoded genes.
    $err \leftarrow$ evaluate fitness function.
    if $err < err\_\beta$
        $err\_\beta \leftarrow err$
        $\beta \leftarrow \partial$
    End
End
if condition not met
    Selection
    Crossover
    Mutation
    Repeat
Else
    return β
End

Figure 7.3 Steps followed to select best ensemble components with ensemble size of x.

The created single ANNs were encoded as binary numbers. Hence, the length of the genes is proportional to the size of the network pool.

$$N = 2^m \tag{7.6}$$

Where *m* is the gene length and N is number of possible ensemble components. *mk* binary numbers identify each ensemble network, where *k* is the chromosome length (ensemble size). The developed GA was applied with multiple chromosome lengths. The basic reason for running the GA multiple times with different chromosome sizes is the very low probability of having an empty gene in a chromosome. The probability of having *m* (*m* is the gene length) consecutive zeros is as follows.

$$1/2^m \tag{7.7}$$

In order to have *m* consecutive zeros at a specific location in a chromosome, the probability would be as follows.

$$P\left(\bigcup_{i=1}^{k} Z_i\right) = 1 - P\left(\bigcap_{i=1}^{k} Z_i^c\right) = 1 - (1 - 1/2^m)^k \tag{7.8}$$

Z is the probability of having consecutive zeros at specific locations (0, m, 2m, …, (k-1) m), and Z^c is the complement of Z, $1 \leq i < k$.

Therefore, there is a small probability of having an empty gene while moving through an ensemble based on GA selection because of the type of gene representation. In addition, the larger the chromosome, the lower the probability of having consecutive zeros. Thus, when setting the number of ensembles to *x*, there is a high probability that the size of the selected ensemble will be *x*. To overcome this, a GA was utilized to select ensemble components multiple times with different chromosome sizes (lengths) for each set of components, as shown in the third step of the proposed approach.

7.3 Data (Case Study 5)

Innisfail is a town located on the northeastern side of Queensland, Australia, with an annual average rainfall of 3553.00 mm. It has been selected as the study area because of its proximity to multiple agricultural areas. Rainfall values between January 1908 and December 2015 were collected from the BOM, which is located in Melbourne, Australia. The details of the created dataset are shown in Section 3.1.

7.4 Experiments and Results

7.4.1 Experimental Setup

7.4.1.1 Single Neural Networks (Pools)

MATLAB ANNs and optimization toolboxes were used to conduct the experiments. Some 12 datasets representing the 12 months in Innisfail were used to create 12 pools of ANNs, each containing 512 FFNNs with the following parameters.

- Four training algorithms: Levenberg-Marquardt BP, Resilient BP, Bayesian Regularization, and Scaled Conjugate Gradient BP.
- The number of neurons ranged between two and 32.
- Three activation functions were used: linear function (purelin), hyperbolic tangent (tansig), and log sigmoid (logsig).
- There were random initial weights.

In rainfall forecasting, the testing dataset is usually selected as the last portion of the available data. Therefore, blocks technique were used when partitioning data. The first

75% of the dataset in each month was used for training. The next 15% were used for validation and the last 10% were used as a testing dataset.

7.4.1.2 Chromosome Encoding

Each of the developed single ANNs represented a possible component for each developed ensemble. Binary values were used to represent each single network. A total of 512 networks were created. Hence, nine binary values represented each possible component in each EA. An example of a chromosome is shown in Figure 7.4. SN represents a single network and h is the size of the ensemble.

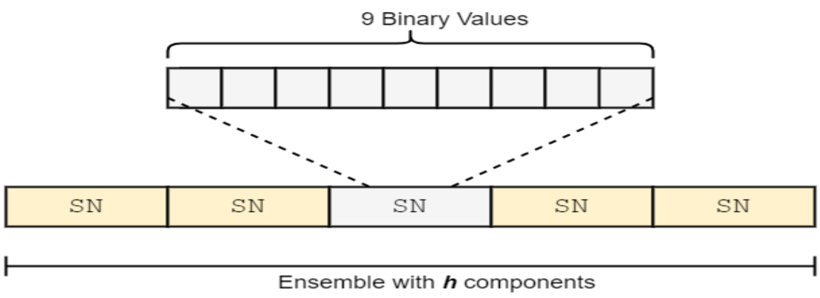

Figure 7.4 Chromosome encoding.

7.4.1.3 Fitness Function

The proposed fusion method was a three-layered FFNN. The input of the ANN in the fitness function was the output of the training data in single networks. The ANN consisted of 13 neurons in the hidden layer. The hyperbolic tangent activation function was used between input-hidden and hidden-output layers. A resilient BP training algorithm was selected to modify weights of connections. RMSE value represented the scalar value of the fitness function. A summary of network details is shown in Table 7.1.

Table 7.1 Fitness function parameters.

Parameter	Value
Type	Feed Forward Neural Network
Number of layers	Three (input-hidden-output)
Neurons	13
Training algorithm	Resilient BP
Activation functions	Hyperbolic tangent

7.4.1.4 Hybrid Genetic Algorithm Parameters

The parameters of the hybrid GA used in selecting ensemble components are shown in Table 7.2.

Table 7.2 Hybrid genetic algorithm parameters.

Parameter	Value
Population	20
Iterations	100
Crossover	Crossover scattered
Mutation	uniform
Fitness function	FFNN
Gene length	9
Scalar value	RMSE

7.4.2 Results of Ensemble Components Selection using a Hybrid Genetic Algorithm (ECS_HGA)

Since chromosomes were encoded as binary strings, there is a small probability of having an empty gene. Consequently, if the chromosome size was specified as x, it is likely that the generated ensemble would have the size x. Therefore, ensemble components selection for each month was applied multiple times with multiple ensemble maximum sizes.

The ECS_HGA method was deployed with multiple ensemble sizes. The maximum number that the generated ensembles could have was 11. Maximum ensemble sizes varied between two and 11. Therefore, the chromosome lengths varied between 18, 27, 36, …, 99 (9 is each gene element length). Ten optimizations were applied for each month to find the ensemble with highest accuracy, and this led to 120 optimizations for the 12 months. MAE, RMSE, r, and R^2 values for the best ensemble selected in each month are shown in Table 7.3. The size of the best combination is shown in the second column (ensemble size). Different numbers of classifiers generated the best ensemble in each month. MAE varied between 0.008 and 0.051. RMSE ranged between 0.055 in February and 0.009 in August. Reasonable correlation and determination of coefficients were obtained through all the months.

Table 7.3 MAE, RMSE, r, and R^2 values for the best selected ensemble using GA in each month (ECS_HGA).

Month	Size	MAE	RMSE	r	R^2
January	6	0.026	0.036	0.939	0.881
February	11	0.051	0.055	0.911	0.83
March	4	0.016	0.019	0.990	0.98
April	10	0.030	0.035	0.894	0.799
May	7	0.022	0.030	0.879	0.773
June	5	0.019	0.023	0.896	0.803
July	7	0.008	0.011	0.964	0.929
August	3	0.008	0.009	0.956	0.913
September	7	0.010	0.012	0.951	0.904
October	4	0.013	0.016	0.960	0.922
November	3	0.018	0.020	0.962	0.926
December	6	0.015	0.018	0.955	0.912

7.5 Comparative Analysis

In this section, the proposed approach is compared to single ANNs and an alternative selection method. Nagahamulla, Ratnayake, and Ratnaweera applied a standard GA to select ensemble components (ECS_SGA) for a location in Sri Lanka [100]. The ensemble components were fused using another ANN. In addition, the proposed approach was compared to ensemble components selection using a PSO algorithm (ECS_PSO). Although this mechanism has not been proposed elsewhere, it was implemented for the purpose of analysing the effect of the hybrid GA. The implemented approach is shown in Figure 7.5

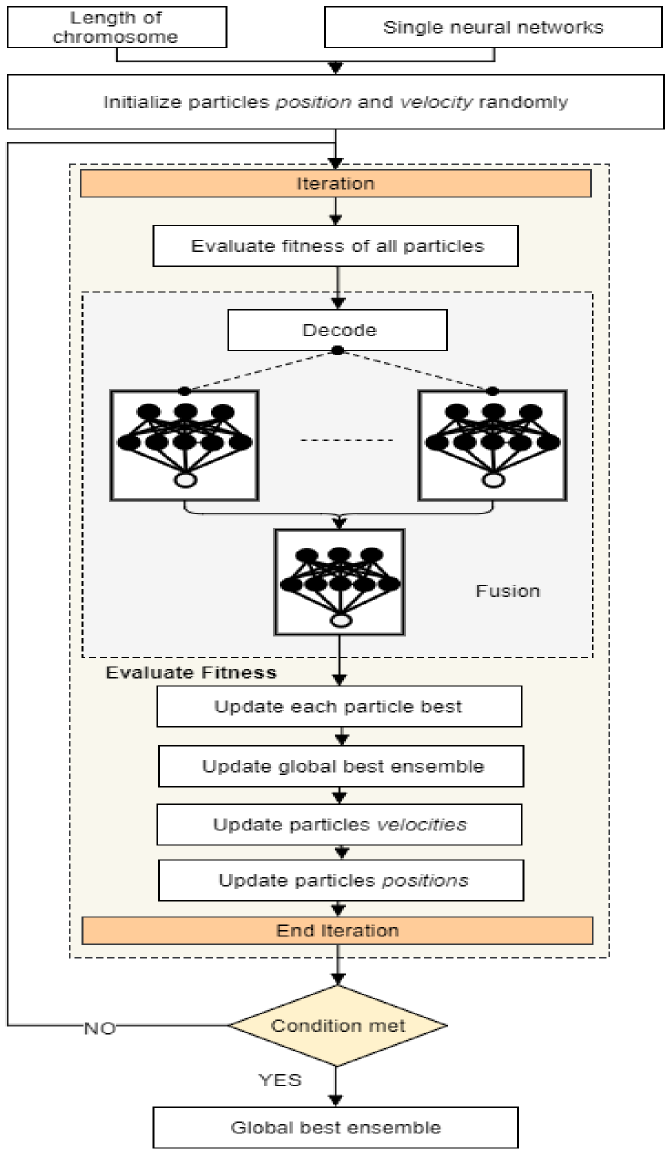

Figure 7.5 Ensemble components selection using particle swarm optimization.

In comparison with single networks, MAE, RMSE, r, and R^2 values were measured for each network in each pool (each month). The 12 ANNs with the highest accuracy for each dataset are shown Table 7.4.

RMSE values ranged between 0.064, the highest in February, and 0.019, the lowest in August. Reasonable correlation values were obtained with all of the single networks, except for June. Months with higher rainfall amounts had higher MAE and RMSE, except for March. The ensembles generated through the proposed method produced lower MAE and RMSE as compared to single ANNs in all the datasets.

Table 7.4 MAE, RMSE, r, and R^2 for the best single network obtained over each month dataset.

Month	MAE	RMSE	r	R^2
January	0.046	0.060	0.837	0.701
February	0.057	0.064	0.864	0.746
March	0.037	0.045	0.947	0.896
April	0.038	0.045	0.828	0.685
May	0.040	0.041	0.744	0.554
June	0.038	0.042	0.583	0.340
July	0.019	0.023	0.796	0.634
August	0.016	0.019	0.801	0.642
September	0.019	0.022	0.750	0.562
October	0.028	0.035	0.730	0.533
November	0.026	0.035	0.884	0.781
December	0.024	0.038	0.859	0.738

To develop the alternative selection approaches, the same fitness function was used (Table 7.1). The details of the standard GA and PSO parameters are shown in Table 7.5. The alternative approaches were applied for multiple chromosome sizes (2, 3, …, 11) for each month. MAE, RMSE, r, and R^2 values, when selecting ensemble components using a standard GA and PSO, are summarized in Table 7.6 and Table 7.7. Better accuracy was obtained with the proposed approach in terms of MAE, RMSE, r, and R^2 in all the months. ECS_SGA was more accurate than the single networks in ten out of 12 optimizations.

Table 7.5 PSO and standard GA parameters.

Algorithm Parameter	GA	PSO
Population	20	20
Iterations	30	30
Crossover	Crossover scattered	--
Mutation	Uniform	---
Fitness function	FFNN	FFNN
Gene length	9	--
Particle length	--	9

Table 7.6 MAE, RMSE, r, and R^2 values for the best selected ensemble using GA in each month (ECS_SGA).

Month	MAE	RMSE	r	R^2
January	0.038	0.055	0.828	0.685
February	0.055	0.071	0.840	0.706
March	0.023	0.027	0.980	0.961
April	0.039	0.046	0.746	0.557
May	0.026	0.034	0.833	0.694
June	0.023	0.030	0.789	0.622
July	0.012	0.014	0.928	0.862
August	0.009	0.010	0.958	0.918
September	0.011	0.013	0.945	0.893
October	0.018	0.020	0.932	0.869
November	0.019	0.023	0.955	0.913
December	0.026	0.037	0.807	0.651

Table 7.7 MAE, RMSE, r, and R^2 values for the best selected ensemble using PSO in each month (ECS_PSO).

Month	MAE	RMSE	r	R^2
January	0.036	0.043	0.892	0.795
February	0.05	0.057	0.896	0.803
March	0.021	0.026	0.981	0.962
April	0.031	0.036	0.897	0.804
May	0.029	0.037	0.797	0.635
June	0.023	0.027	0.839	0.704
July	0.01	0.013	0.946	0.895
August	0.01	0.012	0.94	0.884
September	0.011	0.013	0.93	0.865
October	0.015	0.018	0.942	0.887
November	0.017	0.021	0.958	0.919
December	0.026	0.032	0.839	0.704

Figure 7.6 represents the RMSE value obtained with the best single networks in each month as against the RMSE value obtained with the best ensemble generated in each month using the proposed approach (ECS_HGA). Figure 7.7 represents the RMSE value

obtained with the best ensemble generated in each month using ECS_SGA as against the RMSE value obtained with the best ensemble in each month using the proposed approach (ECS_HGA). Figure 7.8 represents the RMSE value obtained with the best ensemble generated in each month using ECS_PSO as against the RMSE value obtained with the best ensemble in each month using the proposed approach (ECS_HGA). Better performance was obtained with the proposed approach in all months compared to the three forecasting methods. The highest decrease in RMSE was in March, which is the month with the highest annual average compared to others. This demonstrates the effectiveness of this selection method against single ANNs, ECS_SGA, and ECS_PSO.

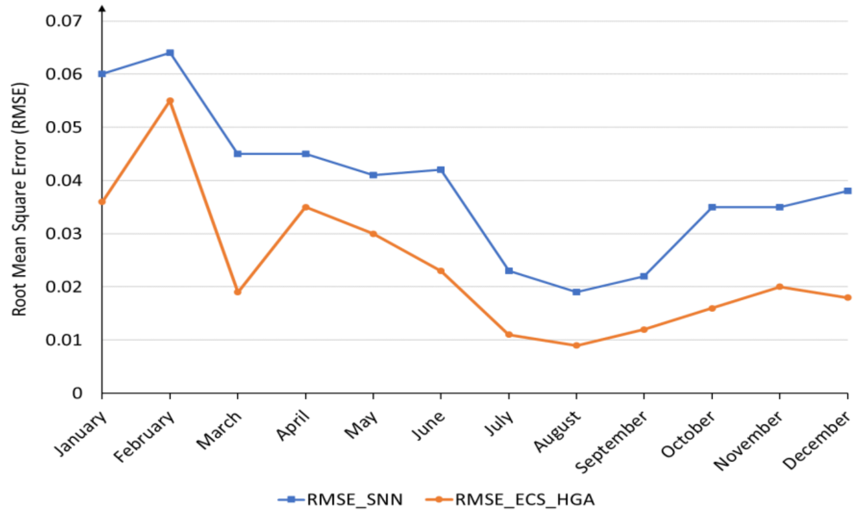

Figure 7.6 RMSE obtained for best single network against best ensemble generated using ECS_HGA for each month.

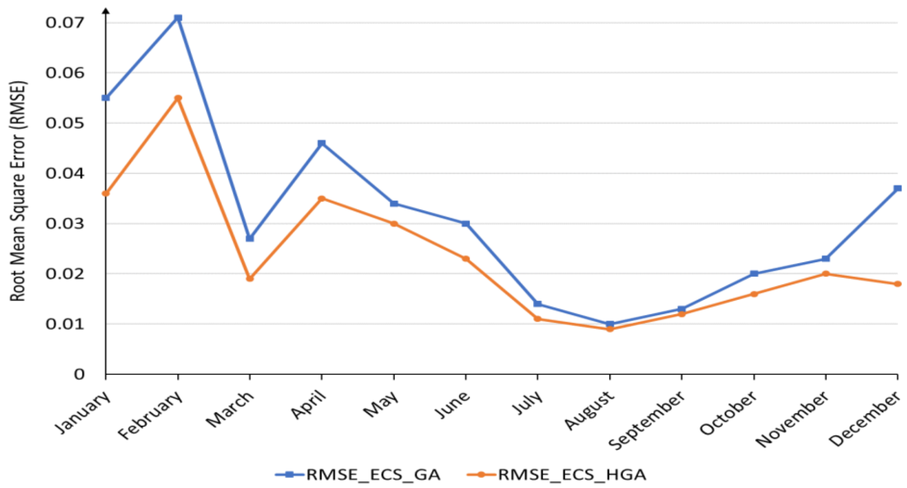

Figure 7.7 RMSE obtained for best ensemble generated using ECS_SGA against ensemble generated using ECS_HGA for each month.

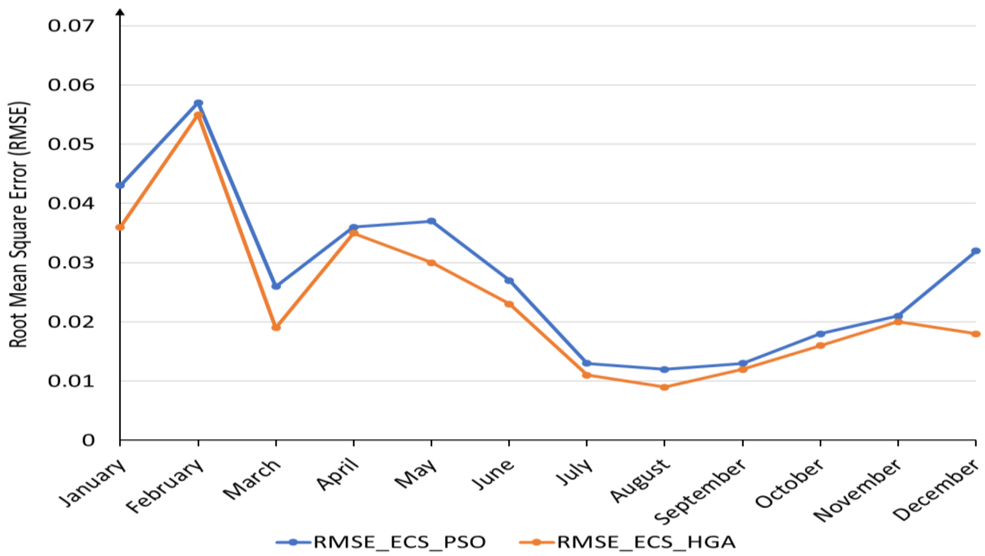

Figure 7.8 RMSE obtained for best ensemble generated using ECS_PSO against ensemble generated using ECS_HGA for each month.

To investigate the whole approaches, the outputs of the best single network in each month were aggregated to form a dataset between January 2005 and December 2015. In addition, the same aggregation was applied for the outputs of each ensemble selected for each month using ECS_SGA, ECS_PSO, and ECS_HGA. MAE, RME, r, and R^2 for climatology, single ANNs, generated ensembles through ECS_SGA, generated ensembles through ECS_PSO, and generated ensembles through ECS_HGA are shown in Table 7.8.

Table 7.8 MAE, RMSE, r and R^2 over the aggregated time series of climatology, single neural networks, ECS_ SGA, ECS_PSO and ECS_HGA.

Method	MAE	RMSE	r	R^2
Climatology	0.055	0.075	0.689	0.474
SNN	0.032	0.041	0.916	0.840
ECS_SGA	0.025	0.036	0.936	0.876
ECS_PSO	0.023	0.031	0.954	0.910
ECS_HGA	0.020	0.027	0.966	0.933

The four methods were more accurate than climatology. Better RMSE was obtained with ECS_SGA, ECS_PSO, and ECS_HGA when compared to single ANNs. Higher correlation values were also generated with the ensemble selection approaches. This shows the effectiveness of ensembles in increasing forecasting accuracy in rainfall applications. Better results were obtained with ECS_HGA.

To show the effectiveness of this method, an Analysis of Variance (ANOVA) test was applied. Three samples were taken: the correlation values obtained using ECS_SGA, the correlation values obtained using single ANNs, and the correlation values obtained using ECS_SGA. The null hypothesis was set to (h0), which means that there is no significant difference between ECS_HGA accuracy and single ANNs, ECS_SGA, and ECS_SGA accuracy.

The null hypothesis was that the three accuracies are similar. The alternative hypothesis (h1) was that the three groups are different. The significance level was assigned to α=0.05. General details of each sample are shown Table 7.9.

Table 7.9 ANOVA samples summaries.

Sample	**ECS_HGA**	**SNN**	**ECS_SGA**	**ECS_PSO**
Count	12	12	12	12
Sum	11.257	9.623	10.541	10.857
Average	0.9381	0.8019	0.8784	0.9048
Standard deviation	0.0346	0.0934	0.0792	0.557

Table 7.10 ANOVA details.

One-Way ANOVA						
Source of Variation	**SS**	**df**	**MS**	**F**	**P-value**	**F-crit**
Between	0.121	3	0.0403	8.3592	0.000165	2.81646583
Within	0.2123	44	0.0048			
Total	0.3333	47				

Sum of Squares (SS), Degrees of freedom (df), Mean Square (MS), and F were determined. If F is greater than the critical value of F (F-crit), the null hypothesis is rejected. As shown in Table 7.10, F was greater than F-crit (F>F-crit). This shows that there is a significant difference between the accuracy of the three samples, and the null hypothesis is rejected. Therefore, there was significant difference between the accuracy of ECS_HGA and other approaches.

7.6 Chapter Summary

This chapter presented a new ensemble approach for building weather forecasting models using specific locations and various forecasting lead times. In addition, a new algorithm for selecting ensemble components was examined. Furthermore, a new components selection approach that uses neural fusion and a hybrid GA was proposed. It was shown that ensembles of ANNs increase the accuracy of predictions. It was also shown that selecting ensemble components using the hybrid GA is more effective than alternative selection techniques.

Chapter 8 Conclusion

This thesis investigated the usefulness of ANNs in building weather prediction models. Several approaches were proposed, implemented, and assessed. This chapter provides a summary of the research described in this thesis, and summarizes its contributions. Finally, possible future directions are discussed.

8.1 Summary and Research Contributions

In this research, new approaches were proposed and evaluated to develop medium-long term forecasting models. ANNs were utilized as the main building blocks of these approaches because of their capability in approximate non-linear relationships between input features and targeted output. EAs were combined with ANNs and ensembles of ANNs to enhance the performance of the forecasting models. In the initial stages, climate indices were searched and collected (Chapter 3). The use of single ANNs was investigated using all of the months, and reasonable accuracy was obtained but there was no ability to release accurate forecasts in months with high annual averages. The dataset was then partitioned so that each ANN was trained using each month's values only.

Climate indices have been linked to Australian rainfall variability in parts of the year. Therefore, a feature selection approach using a GA to select predictors for a specific location at a specific time was proposed. This approach was evaluated against a single ANN that uses all input features as predictors, and better accuracy was recorded. Another technique that uses a PSO algorithm to select input features in order to categorize rainfall values was also proposed. This approach was compared to a single ANN that uses all input features to categorize rainfall, and better classification accuracy was observed. These two approaches were discussed in Chapter 4.

In Chapter 5, the selection technique was extended to include not only climate input features but also network parameters. Throughout Chapter 4, limitations were observed even though better accuracy was obtained. This is because the selected network parameters may not be the optimal parameters, so there was a need to search for the best network parameters. In addition, the ANN initial weights were selected from the same subset. The same set of initial weights was used every time in the training process. Therefore, there could be another set, which might reveal better performance. This, it is necessary to search for the optimal set of initial weights. To overcome these limitations, the selection approach was extended to include both climate input features and ANN parameters. In addition, a new hybrid GA was proposed that saves the global best solution. This was incorporated from functionalities in PSO. This is different to elitism in which the best chromosome is put into the next iteration. The proposed approach was then evaluated against the feature selection approach, and better performance was obtained. Then, the proposed approach was extended to include multiple optimization levels. The main reason behind this extension was to optimize the characteristics of the forecasting model on each level only. In other words, the climate input features and network parameters were not included in the same chromosome. For each input features subset, the best network parameters were optimized. This approach produced the best accuracy compared to alternatives. In addition, it was evaluated against the ACCESS forecasting model, and better accuracy was obtained.

In Chapter 6, a new type of forecasting system was proposed and evaluated. Ensemble classifiers were introduced as possible forecasting models for locally specified regions. To design an ensemble, single classifiers were created first, then combined using a specific fusion method. In this chapter, multiple fusion methods were proposed and evaluated. These fusion methods were divided mainly into two categories: fusion using the average, and fusion using an ANN. Three approaches in which ANNs are used as the fusion methods were proposed. The best fusion method was an ANN based PSO fusion method, in which better accuracy was observed compared to other approaches. It was shown that the average fusion method is not useful for rainfall prediction for locally specified regions. The selected approach was compared to bagging and climatology, and a lower error was observed. Although better accuracy in comparison with alternative approaches was obtained, the ensemble was approximated to a non-linear function that combines both low and high monthly averages. The generated approach was not able to predict peak rainfall values. This was related to seasonality in the data.

In Chapter 7, a novel approach was proposed for selecting ensemble components. One of the main conclusions in Chapter 6 was that the ANN fusion method had reasonable performance. Therefore, the ANN fusion method was kept as the fusion method for the proposed ensemble. In addition, it was concluded that using every month’s dataset may not result in good accuracy when forecasting specific months. For this reason, each month's dataset only was used. Finally, it was noticed that the created single classifiers may not be optimal. For this reason, a number of single ensemble components were created and selected using a hybrid GA. The proposed approach was compared with an alternative selection approach, in addition to a PSO based selection approach. Better accuracy was obtained with the proposed selection approach.

The following table (Table 8.1) summarizes all of the experiments conducted in this research. The approaches listed in the table are categorized as proposed (P) or compared (C). The location, input feature, type, method, examined variable, and size of the dataset are also shown in the table.

Table 8.1 Summary of experiments in this research.

Case Study	ID	Location	Input features	Type	P/C	Method	Variable examined	Size
1	I	Bingera	10	SNN	P	GA	Input features	One month
	I	Bingera	10	SNN	C	-	-	One month
2	III	Innisfail	28	SNN	P	PSO	Input features	One month
	IV	Plane Creek	28	SNN	P	PSO	Input features	One month
	V	Bingera	28	SNN	P	PSO	Input features	One month
	VI	Maryborough	28	SNN	P	PSO	Input features	One month
	VII	Yamba	28	SNN	P	PSO	Input features	One month
	III	Innisfail	28	SNN	C	-	-	One month
	IV	Plane Creek	28	SNN	C	-	-	One month
	V	Bingera	28	SNN	C	-	-	One month
	VI	Maryborough	28	SNN	C	-	-	One month
	VII	Yamba	28	SNN	C	-	-	One month
3	II	Innisfail	11	SNN	P	HGA	Input features and network parameters	One month
	II	Innisfail	11	SNN	C	GA	Input features and network parameters	One month
	II	Innisfail	11	SNN	C	GA	Input features	One month
	II	Innisfail	11	SNN	P	Two Hybrid GAs	Input features and network parameters	One month

	II	Innisfail	11	MLR	C	-	-	One month
	II	Innisfail	-	Climatology	C	-	-	-
4	VIII	Innisfail	11	ENN	P	-	Average	All month’s
	VIII	Innisfail	11	ENN	P	-	ANN fusion	All month’s
	VIII	Innisfail	11	ENN	P	-	Lowest error network fusion	All month’s
	VIII	Innisfail	11	ENN	P	PSO	PSO network fusion	All month’s
	VIII	Innisfail	11	ENN	C	-	Bagging	All month’s
	VIII	Innisfail	11	SNN	C	-	-	All month’s
	VIII	Innisfail	-	Climatology	C	-	-	-
5	III	Innisfail	28	ENN	P	HGA	Ensemble components	One month
	III	Innisfail	28	ENN	C	PSO	Ensemble components	One month
	III	Innisfail	28	ENN	C	SGA	Ensemble components	One month
	III	Innisfail	28	SNN	C	-	-	One month
	III	Innisfail	-	Climatology	C	-	-	-

The answers obtained to the research questions are summarized below.

- *What climate indices can be used to predict rainfall, and what is the effect of these indices on forecasts using an ANN model?*
  It was shown in Chapters 4 and 5 that for each month in each selected location, different subsets of climate indices can be used to forecast rainfall. A climate index may affect rainfall variability in certain parts of the year. In addition, a climate index may be beneficial for rainfall prediction in one location, and have no effect in another location. With that in mind, ENSO indicators were commonly selected for each month using various approaches. This was in harmony with previous research in which ENSO conditions have been linked to Australian climate variability.

- *What is the best method for selecting possible predictors and ANN parameters?*
  It was shown in Chapter 5 that selecting network parameters and input features produces models with better rainfall prediction performance. It was concluded that selecting network parameters for each input features subset produces the best

accuracy (Section 5.3). In addition, it was found that the inclusion of a hybrid GA when applying the selection approach enhances the performance of the forecasting model.

- *What is the effect of using ensembles of ANNs on a model's performance?*

  The use of ensemble techniques in ANNs is very effective, but ensemble components and the fusion method should be extensively optimized in order to deliver accurate forecasts. Based on investigations in Chapter 6, using an ANN to fuse single classifiers enhanced prediction accuracy. It was also found that combining the classifiers with certain fusion methods (average) in this type of application may lead to undesirable outcomes. As a result, the conclusion reached in Chapter 7 was that selecting ensemble components can enhance prediction accuracy.

- *What is the best method to select ensemble components to obtain highest accuracy?*

  In Chapter 7, we demonstrated the effectiveness of an ensemble modelling technique that selects ensemble components using EAs. All of the proposed and evaluated approaches had reasonable accuracy, and superior performance was obtained by a mechanism that uses a hybrid GA. It was concluded that it is not guaranteed that the formation of an ensemble from the classifiers with highest accuracy in the pool would lead to an ensemble with better performance. The best ensemble would contain various combinations of classifiers with various accuracy ranges.

- *Are ANN forecasts better than existing forecasting models? Could ANNs represent a possible alternative for existing complex forecasting models?*

  ANNs have shown accurate performance in approximating rainfall values and categories using various weather input features through various parts of this research. We compared the proposed ANN approach to ACCESS, which was introduced in the middle stages of this research. ACCESS-S1 is the first version of the BOM's monthly forecasting model. A set of hindcasts was found on the NCI website. We compared to our best model, and better accuracy was observed. The accuracy of optimized ANNs recorded in this application area is promising. Their ability in approximating the non-linear relationships between weather attributes has been demonstrated in this research. With the vast spread of data, and new recordings released over time, ANNs could be used as possible forecasting models. Knowing that a sufficient amount of data can be found in many weather stations across Australia, the proposed approaches in this research could be generalized and used.

Some of the conclusions drawn in this research are summarized below.

- It was concluded that feature selection is mandatory in ANN these types of applications, since each feature may affect the location at specific times of the year.
- ANNs can be reliable forecasting models if their elements are optimized properly.
- Because of seasonality and weather fluctuations, it is better to use each month's dataset to predict rainfall values. It was hard to predict peak rain values using a dataset that contained all of the months.
- Ensemble components should be wisely selected when predicting rain. Ensemble fusion using the average had low accuracy, and contradicted related research in other locations.
- The selection and optimization approaches proposed in this research can be used in any other application. A trial and error method can result in a good solution, but optimization is better in finding optimal solutions.
- The implementation cost of ANN models is low compared to other models which require super computation power.
- In some areas, such as large agricultural farms and cattle areas, farmers make their own recordings. Using of these approaches, new location based forecasting models can be produced to predict rainfall.
- The more information given to the network about antecedent weather conditions, the better the accuracy that can be obtained, especially with rainfall antecedent values. But this type of forecasting model can be used for one-month rainfall prediction. In other words, the more information given to the network, the less the information and duration given to users.

8.2 Future Work

This research represents a basic starting point for the use of machine learning algorithms as an alternative to forecasting models in atmospheric research. The optimization and selection approaches proposed in this research can be applied to predict any other weather variable, such as temperature and wind.

It was demonstrated in Chapters 4, 5, 6, and 7 that ANNs and ensembles of ANNs could deliver reasonable performance when optimized properly. The proposed methods can be applied to any weather station in Australia, and the generated network/ensemble can be used for releasing predictions. The whole approach can be extended by applying extrapolation techniques to combine the prediction output in each weather station.

Furthermore, additional ANN topologies and other machine learning algorithms can be investigated and analysed in order to assess their capability in forecasting rainfall. Ensemble techniques can be extended by incorporating new fusion methods.

The GA and PSO were mainly incorporated to this research. Additional optimization algorithms including the memetic algorithms can be investigated and explored.

The ability of new climate indices and weather attributes in forecasting Australian rainfall could be examined. For example, climate indices such as the Southern Annular Mode index could be utilized and used as possible predictors for Australian rainfall variability after being processed (recorded since 1957). In addition, several pre-processing techniques could be applied to extend the amount of available data of specific climate indices that were excluded from this research because of its small size in comparison to collected indices.

REFERENCES

[1] P. McIntosh, P. Carberry, M. Stafford Smith, B. Cullen, and A. Ash, "Constraints and opportunities in applying seasonal climate forecasts in agriculture," *Australian Journal of Agricultural Research,* vol. 58, no. 10, pp. 952-965, 2007.

[2] S. Graham, "Weather Forecasting Through the Ages: Feature Articles," 2002.

[3] L. Bosher and K. Chmutina, *Disaster risk reduction for the built environment*. John Wiley & Sons, 2017.

[4] A. Wiin-Nielsen, "The birth of numerical weather prediction," *Tellus A,* vol. 43, no. 4, pp. 36-52, 1991.

[5] L. F. Richardson, *Weather prediction by numerical process*. Cambridge University Press, 2007.

[6] P. Roudier *et al.,* "The role of climate forecasts in smallholder agriculture: Lessons from participatory research in two communities in Senegal," *Climate Risk Management,* vol. 2, pp. 42-55, 2014/01/01/ 2014.

[7] T. Gunda, J. T. Bazuin, J. Nay, and K. L. Yeung, "Impact of seasonal forecast use on agricultural income in a system with varying crop costs and returns: an empirically-grounded simulation," *Environmental Research Letters,* vol. 12, no. 3, p. 034001, 2017/02/22 2017.

[8] J. A. Silva and C. J. Matyas, "Relating Rainfall Patterns to Agricultural Income: Implications for Rural Development in Mozambique," *Weather, Climate, and Society,* vol. 6, no. 2, pp. 218-237, 2014.

[9] R. C. Van den Honert and J. McAneney, "The 2011 Brisbane floods: Causes, impacts and implications," *Water,* vol. 3, no. 4, pp. 1149-1173, 2011.

[10] J. A. Osborne, N. E. Stoeckl, Y. L. Everingham, N. G. Inman-Bamber, and R. Welters, "The economic value of conditioning harvest start date on long-lead seasonal climate forecasts," in *2011 Conference of the Australian Society of Sugar Cane Technologists*, Mackay, Australia, 2011, pp. 37-Ag 37.

[11] Y. L. Everingham, A. J. Clarke, C. C. M. Chen, S. v. Gorder, and P. McGuire, "Exploring the capabilities of a long lead climate forecasting system for the NSW sugar industry," in *2007 Conference of the Australian Society of Sugar Cane Technologists,* Mackay, Australia, 2007: Australian Society of Sugar Cane Technologists, pp. 9-17.

[12] W. Drosdowsky and M. C. Wheeler, "Predicting the onset of the north australian wet season with the POAMA dynamical prediction system," *Weather and Forecasting,* vol. 29, no. 1, pp. 150-161, 2014.

[13] A. Cottrill *et al.*, "Seasonal climate prediction in the Pacific using the POAMA coupled model forecast system," in "National Library of Australia Cataloguing-in-Publication entry," CACWCR Technical Report No. 048, 2012.

[14] D. Hudson, A. G. Marshall, Y. Yin, O. Alves, and H. H. Hendon, "Improving intraseasonal prediction with a new ensemble generation strategy," *Monthly Weather Review,* vol. 141, no. 12, pp. 4429-4449, 2013.

[15] D. Hudson *et al.*, "ACCESS-S1: The new Bureau of Meteorology multi-week to seasonal prediction system," *Journal of Southern Hemisphere Earth Systems Science,* vol. 67, no. 3, pp. 132-159, 2017.

[16] D. Hudson, L. Shi, O. Alves, M. Zhao, H. Hendon, and G. Young, "Performance of ACCESS-S1 for key horticultural regions," in "Bureau of Meteorology, Melbourne " Bureau of Meteorology, , Melbourne, Australia 2017.

[17] R. P. Shukla, K. C. Tripathi, A. C. Pandey, and I. M. L. Das, "Prediction of Indian summer monsoon rainfall using NINO indices: A neural network approach," *Atmospheric Research,* vol. 102, no. 1, pp. 99-109, 2011.

[18] C.-F. Tsai, Y.-C. Lin, D. C. Yen, and Y.-M. Chen, "Predicting stock returns by classifier ensembles," *Applied Soft Computing,* vol. 11, no. 2, pp. 2452-2459, 3// 2011.

[19] W. T. Lim, L. Wang, Y. Wang, and Q. Chang, "Housing price prediction using neural networks," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2016, pp. 518-522.

[20] L. Zhang, B. Verma, and D. Stockwell, "Class-semantic color-texture textons for vegetation classification," in *International Conference on Neural Information Processing (ICONIP 2015)*, 2015: Springer, pp. 354-362.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[22] S. S. Talathi, "Hyper-parameter optimization of deep convolutional networks for object recognition," in *2015 IEEE International Conference on Image Processing (ICIP)*, 2015, pp. 3982-3986.

[23] B. Han and X. Bian, "A hybrid PSO-SVM-based model for determination of oil recovery factor in the low-permeability reservoir," *Petroleum,* vol. 4, no. 1, pp. 43-49, 2017/06/19/ 2017.

[24] M. A. Ahmadi and A. Ahmadi, "Applying a sophisticated approach to predict CO2 solubility in brines: application to CO2 sequestration," *International Journal of Low-Carbon Technologies,* vol. 11, no. 3, pp. 325-332, 2016.

[25] A. Shafiei, M. A. Ahmadi, S. H. Zaheri, A. Baghban, A. Amirfakhrian, and R. Soleimani, "Estimating hydrogen sulfide solubility in ionic liquids using a machine learning approach," *The Journal of Supercritical Fluids,* vol. 95, pp. 525-534, 11// 2014.

[26] M. A. Ahmadi and S. R. Shadizadeh, "New approach for prediction of asphaltene precipitation due to natural depletion by using evolutionary algorithm concept," *Fuel,* vol. 102, pp. 716-723, 12// 2012.

[27] K. Muralitharan, R. Sakthivel, and R. Vishnuvarthan, "Neural network based optimization approach for energy demand prediction in smart grid," *Neurocomputing,* vol. 273, no. Supplement C, pp. 199-208, 2018/01/17/ 2018.

[28] Y. Jungwon and K. and Sungshin, "Locally-weighted polynomial neural network for daily short-term peak load forecasting," *International Journal of Fuzzy Logic and Intelligent Systems,* vol. 16, no. 3, pp. 163-172, 09 2016.

[29] P. Wang, G. Zhao, and X. Yao, "Applying back-propagation neural network to predict bus traffic," in *2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, 2016, pp. 752-756.

[30] W. W. Guo, L. D. Li, and G. Whymark, "Simulating wheat yield in New South Wales of Australia using interpolation and neural networks," in *17th International Conference on Neural Information Processing*, Sydney,Australia, 2010: Springer Berlin Heidelberg, pp. 708-715.

[31] G. Rub, R. Kruse, M. Schneider, and P. Wagner, "Data mining with neural networks for wheat yield prediction," in *8th Industrial Conference in Data Mining (ICDM 2008)*, Leipzig, Germany, 2008: Springer Berlin Heidelberg, pp. 47-56.

[32] S. Thuankaewsing, S. Pathumnakul, and K. Piewthongngam, "Using an artificial neural network and a mathematical model for sugarcane harvesting scheduling," in *IEEE International Conference on Industrial Engineering and Engineering Management*, Singapore, 2011: IEEE, pp. 308-312.

[33] R. Haidar, I. Koprinska, and B. Jeffries, "Sleep apnea event detection from nasal airflow using convolutional neural networks," in *24th International Conference on Neural Information Processing (ICONIP 2017)*, Guangzhou, China, 2017: Springer International Publishing, pp. 819-827.

[34] W. Sun, T.-L. B. Tseng, J. Zhang, and W. Qian, "Enhancing deep convolutional neural network scheme for breast cancer diagnosis with unlabeled data," *Computerized Medical Imaging and Graphics*, vol. 57, pp. 4-9, 2017.

[35] P. M. Leod, B. Verma, and M. Zhang, "Optimizing configuration of neural ensemble network for breast cancer diagnosis," in *2014 International Joint Conference on Neural Networks (IJCNN)*, 2014, pp. 1087-1092.

[36] N. D. Truong *et al.*, "Convolutional neural networks for seizure prediction using intracranial and scalp electroencephalogram," *Neural Networks*, vol. 105, pp. 104-111, 2018/09/01/ 2018.

[37] A. M. Bagirov and A. Mahmood, "A comparative assessment of models to predict monthly rainfall in australia," *Water Resources Management*, journal article vol. 32, no. 5, pp. 1777-1794, March 01 2018.

[38] H. Vathsala and S. G. Koolagudi, "Prediction model for peninsular Indian summer monsoon rainfall using data mining and statistical approaches," *Computers & Geosciences*, vol. 98, no. Supplement C, pp. 55-63, 2017/01/01/ 2017.

[39] H. Sanikhani *et al.*, "Survey of different data-intelligent modeling strategies for forecasting air temperature using geographic information as model predictors," *Computers and Electronics in Agriculture*, vol. 152, pp. 242-260, 2018/09/01/ 2018.

[40] F. Mekanik and M. A. Imteaz, "Capability of artificial neural networks for predicting long-term seasonal rainfalls in east Australia," in *20th International Congress on Modelling and Simulation (Modsim2013)*, 2013, pp. 2674-2680.

[41] R. C. Deo and M. Şahin, "Forecasting long-term global solar radiation with an ANN algorithm coupled with satellite-derived (MODIS) land surface temperature

(LST) for regional locations in Queensland," *Renewable and Sustainable Energy Reviews,* vol. 72, no. Supplement C, pp. 828-848, 2017/05/01/ 2017.

[42] K. Gnana Sheela and S. N. Deepa, "Performance analysis of modeling framework for prediction in wind farms employing artificial neural networks," *Soft Computing,* journal article vol. 18, no. 3, pp. 607-615, 2014.

[43] T. Kashiwao, K. Nakayama, S. Ando, K. Ikeda, M. Lee, and A. Bahadori, "A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency," *Applied Soft Computing,* vol. 56, pp. 317-330, 7// 2017.

[44] P. T. Nastos, A. G. Paliatsos, K. V. Koukouletsos, I. K. Larissi, and K. P. Moustris, "Artificial neural networks modeling for forecasting the maximum daily total precipitation at Athens, Greece," *Atmospheric Research,* vol. 144, pp. 141-150, 7/1/ 2014.

[45] K. Nadig, W. Potter, G. Hoogenboom, and R. McClendon, "Comparison of individual and combined ANN models for prediction of air and dew point temperature," *Applied Intelligence,* vol. 39, no. 2, pp. 354-366, 2013/09/01 2013.

[46] A. Pasini and R. Langone, "Attribution of precipitation changes on a regional scale by neural network modeling: A case study," *Water,* vol. 2, no. 3, pp. 321-332, 2010.

[47] S. A. A. Fazel, H. Mirfenderesk, R. Tomlinson, and M. Blumenstein, "Towards robust flood forecasts using neural networks," in *2015 International Joint Conference on Neural Networks (IJCNN),* Killarney, Ireland, 2015, pp. 1-6.

[48] M. Toukourou, A. Johannet, G. Dreyfus, and P.-A. Ayral, "Rainfall-runoff modeling of flash floods in the absence of rainfall forecasts: the case of “Cévenol flash floods”," *Applied Intelligence,* vol. 35, no. 2, pp. 178-189, 2011/10/01 2011.

[49] J. Zou, Y. Han, and S.-S. So, "Overview of artificial neural networks," *Methods in molecular biology (Clifton, N.J.),* vol. 458, p. 15, 2008.

[50] M. S. Baawain, M. H. Nour, A. G. El-Din, and M. G. El-Din, "El Nino southern-oscillation prediction using Southern Oscillation Index and Nino3 as onset indicators: Application of artificial neural networks," *Journal of Environmental Engineering and Science,* vol. 4, no. 2, pp. 113-113, 2005.

[51] A. Krenker, A. Kos, and J. Bešter, *Introduction to the artificial neural networks*. INTECH Open Access Publisher, 2011.

[52] J. Krenek, K. Kuca, O. Krejcar, P. Maresova, V. Sobeslav, and P. Blazek, "Artificial neural network tools for computerised data modeling and processing," in *15th International Symposium on Computational Intelligence and Informatics (CINTI),* 2014, pp. 255-260.

[53] D. Svozil, V. Kvasnicka, and J. í. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems,* vol. 39, no. 1, pp. 43-62, 1997/11/01/ 1997.

[54] M. N. French, W. F. Krajewski, and R. R. Cuykendall, "Rainfall Forecasting in Space and Time Using a Neural Network," *Journal of Hydrology,* vol. 137, no. 1, pp. 1-31, 1992/08/15 1992.

[55] X. Gan, L. Chen, D. Yang, and G. Liu, "The research of rainfall prediction models based on matlab neural network," in *2011 IEEE International Conference on Cloud Computing and Intelligence Systems (CCIS)* 2011, pp. 45-48.

[56] P. Singh and B. Borah, "Indian summer monsoon rainfall prediction using artificial neural network," *Stochastic Environmental Research and Risk Assessment*, vol. 27, no. 7, pp. 1585-1599, Oct 2013.

[57] M. P. Darji, V. K. Dabhi, and H. B. Prajapati, "Rainfall forecasting using neural network: A survey," in *2015 International Conference on Advances in Computer Engineering and Applications (ICACEA)*, 2015, pp. 706-713.

[58] N. Kurian, T. Venugopal, J. Singh, and M. M. Ali, "A soft-computing ensemble approach (SEA) to forecast Indian summer monsoon rainfall," *Meteorological Applications*, vol. 24, no. 2, pp. 308-314, 2017.

[59] K. C. Luk, J. E. Ball, and A. Sharma, "An application of artificial neural networks for rainfall forecasting," (in English), *Math. Comput. Model.*, Article; Proceedings Paper vol. 33, no. 6-7, pp. 683-693, Mar-Apr 2001.

[60] S. Chaudhuri and S. Chattopadhyay, "Neuro-computing based short range prediction of some meteorological parameters during the pre-monsoon season," *Soft Computing*, journal article vol. 9, no. 5, pp. 349-354, 2005.

[61] S. Chattopadhyay, "Feed forward artificial neural network model to predict the average summer-monsoon rainfall in India," *Acta Geophysica*, journal article vol. 55, no. 3, pp. 369-382, 2007.

[62] L. Ayalew, D. P. F. Moller, and G. Reik, "Using artificial neural networks (ANN) for real time flood forecasting, the Omo river case in southern Ethiopia," in *2007 Summer Computer Simulation Conference*, San Diego, California, 2007, 1358125: Society for Computer Simulation International, pp. 1-7.

[63] M. Nasseri, K. Asghari, and M. J. Abedini, "Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network," *Expert Systems with Applications*, vol. 35, no. 3, pp. 1415-1421, 10// 2008.

[64] S. Karmakar, M. Kowar, and P. Guhathakurta, "Long-range monsoon rainfall pattern recognition and prediction for the subdivision 'epmb' Chhattisgarh using deterministic and probabilistic neural network," in *Seventh International Conference on Advances in Pattern Recognition*, 2009, pp. 367-370.

[65] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi, "An artificial neural network model for rainfall forecasting in Bangkok, Thailand," *Hydrology and Earth System Sciences*, vol. 13, no. 8, pp. 1413-1425, 2009.

[66] A. Sedki, D. Ouazar, and E. El Mazoudi, "Evolving neural network using real coded genetic algorithm for daily rainfall–runoff forecasting," *Expert Systems with Applications*, vol. 36, no. 3, pp. 4523-4527, 2009/04/01/ 2009.

[67] Z.-l. Wang and H.-h. Sheng, "Rainfall prediction using generalized regression neural network: Case study Zhengzhou," in *2010 International Conference on Computational and Information Sciences (ICCIS)*, 2010, 2010, pp. 1265-1268.

[68] T. N. d. Castro, F. Souza, J. M. B. Alves, R. S. T. Pontes, M. B. M. Firmino, and T. M. d. Pereira, "Seasonal rainfall forecast using a neo-fuzzy neuron model," in *9th IEEE International Conference on Industrial Informatics*, 2011, pp. 694-698.

[69] K. P. Moustris, I. K. Larissi, P. T. Nastos, and A. G. Paliatsos, "Precipitation forecast using artificial neural networks in specific regions of Greece," *Water Resources Management,* journal article vol. 25, no. 8, pp. 1979-1993, 2011.

[70] H. Nagahamulla, U. Ratnayake, and A. Ratnaweera, "Monsoon rainfall forecasting in Sri Lanka using artificial neural networks," in *6th International Conference on Industrial and Information Systems*, 2011, pp. 305-309.

[71] S. Campisi-Pinto, J. Adamowski, and G. Oron, "Forecasting urban water demand via wavelet-denoising and neural network models. Case study: City of Syracuse, Italy," *Water Resources Management,* journal article vol. 26, no. 12, pp. 3539-3558, September 01 2012.

[72] K. Abhishek, M. P. Singh, S. Ghosh, and A. Anand, "Weather forecasting model using artificial neural network," *Procedia Technology,* vol. 4, pp. 311-318, // 2012.

[73] N. A. Charaniya and S. V. Dudul, "Focused time delay neural network model for rainfall prediction using Indian Ocean Dipole index," in *2012 Fourth International Conference on Computational Intelligence and Communication Networks (CICN)*, 2012: IEEE, pp. 851-855.

[74] P. T. Nastos, K. P. Moustris, I. K. Larissi, and A. G. Paliatsos, "Rain intensity forecast using artificial neural networks in Athens, Greece," *Atmospheric Research,* vol. 119, no. Supplement C, pp. 153-160, 2013/01/01/ 2013.

[75] X. Meng, "Weather forecast based on improved genetic algorithm and neural network," in *International Conference on Information Engineering and Applications (IEA)*, London, 2013: Springer London, pp. 833-838.

[76] F. Mekanik, M. A. Imteaz, S. Gato-Trinidad, and A. Elmahdi, "Multiple regression and artificial neural network for long-term rainfall forecasting using large scale climate modes," *Journal of Hydrology,* vol. 503, pp. 11-21, 10/30/ 2013.

[77] L. Jiang and J. Wu, "Hybird evolutionary algorithms for artificial neural network training in rainfall forecasting," in *10th International Symposium on Neural Networks (ISSN 2013),* Dalian, China, 2013: Springer Berlin Heidelberg, pp. 359-366.

[78] L. Jiang and J. Wu, "Hybrid PSO and GA for neural network evolutionary in monthly rainfall forecasting," in *5th Asian Conference on Intelligent Information and Database Systems (ACIIDS 2013),* Kuala Lumpur, Malaysia, 2013: Springer Berlin Heidelberg, pp. 79-88.

[79] B. K. Rani, K. Srinivas, and A. Govardhan, "Rainfall prediction with TLBO optimized ANN," *Journal of Scientific & Industrial Research,* vol. 73, no. 10, pp. 643-647, Oct 2014.

[80] S. Khedhiri, "Artificial neural network for forecasting rainfall pattern in Prince Edward island, Canada," *International Journal of Environmental Studies*, vol. 72, no. 2, pp. 331-340, 2015.

[81] Mislan, Haviluddin, S. Hardwinarto, Sumaryono, and M. Aipassa, "Rainfall monthly prediction based on artificial neural network: a case study in Tenggarong

station, East Kalimantan - Indonesia," *Procedia Computer Science,* vol. 59, pp. 142-151, // 2015.

[82] R. C. Deo and M. Şahin, "Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia," *Atmospheric Research,* vol. 161-162, no. Supplement C, pp. 65-81, 2015/07/01/ 2015.

[83] J. Wu, J. Long, and M. Liu, "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm," *Neurocomputing,* vol. 148, pp. 136-142, 1/19/ 2015.

[84] F. Mekanik, M. A. Imteaz, and A. Talei, "Seasonal rainfall forecasting by adaptive network-based fuzzy inference system (ANFIS) using large scale climate signals," *Climate Dynamics,* vol. 46, no. 9, pp. 3097-3111, 2016.

[85] L. Jaedong and L. and Jee-Hyong, "Constructing efficient regional hazardous weather prediction models through big data analysis," *International Journal of Fuzzy Logic and Intelligent Systems,* vol. 16, no. 1, pp. 1-12, 03 2016.

[86] J. A. Le, H. M. El-Askary, M. Allali, and D. C. Struppa, "Application of recurrent neural networks for drought projections in California," *Atmospheric Research,* vol. 188, no. Supplement C, pp. 100-106, 2017/05/15/ 2017.

[87] T. G. Dietterich, "Ensemble methods in machine learning," in *Multiple Classifier Systems: First International Workshop, MCS 2000 Cagliari, Italy, June 21–23, 2000 Proceedings*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000, pp. 1-15.

[88] I. Maqsood, M. R. Khan, and A. Abraham, "An ensemble of neural networks for weather forecasting," *Neural Computing & Applications,* vol. 13, no. 2, pp. 112-122, 2004.

[89] B. Gadgay and S. Kulkarni, "Novel ensemble neural network models for better prediction using variable input approach," *International Journal of Computer Applications,* vol. 39, no. 18, pp. 37-45, 2012.

[90] G. P. Zhang and V. L. Berardi, "Time series forecasting with neural network ensembles: An application for exchange rate prediction," *Journal of the Operational Research Society,* vol. 52, no. 6, pp. 652-664, 2001.

[91] M. Fernández-Redondo, C. Hernández-Espinosa, and J. Torres-Sospedra, "Multilayer feedforward ensembles for classification problems," in *11th International Conference on Neural Information Processing: (ICONIP 2004),* Calcutta, India, 2004: Springer Berlin Heidelberg, pp. 744-749.

[92] R. Chen and J. Yu, "An improved bagging neural network ensemble algorithm and its application," in *Third International Conference on Natural Computation (ICNC 2007),* 2007, vol. 5, pp. 730-734.

[93] G. P. Zhang, "A neural network ensemble method with jittered training data for time series forecasting," *Inf. Sci.,* vol. 177, no. 23, pp. 5329-5346, 2007.

[94] S. Araghinejad, M. Azmi, and M. Kholghi, "Application of artificial neural network ensembles in probabilistic hydrological forecasting," *Journal of Hydrology,* vol. 407, no. 1–4, pp. 94-104, 9/15/ 2011.

[95] I. Maqsood, M. Khan, and A. Abraham, "Weather forecasting models using ensembles of neural networks," in *Intelligent Systems Design and Applications*, vol. 23, A. Abraham, K. Franke, and M. Köppen, Eds. (Advances in Soft Computing: Springer Berlin Heidelberg, 2003, pp. 33-42.

[96] I. Maqsood, M. R. Khan, G. H. Huang, and R. Abdalla, "Application of soft computing models to hourly weather analysis in southern Saskatchewan, Canada," *Engineering Applications of Artificial Intelligence,* vol. 18, no. 1, pp. 115-125, 2005/02/01/ 2005.

[97] S. S. Monira, Z. M. Faisal, and H. Hirose, "A neural network ensemble incorporated with dynamic variable selection for rainfall forecast," in *12th ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, 2011: IEEE, pp. 7-12.

[98] L. Jin, Y. Huang, and H. S. Zhao, "Ensemble prediction of monthly mean rainfall with a particle swarm optimization-neural network model," in *13th International Conference on Information Reuse and Integration (IRI)* 2012, pp. 287-294.

[99] H. Nagahamulla, U. Ratnayake, and A. Ratnaweera, "An ensemble of artificial neural networks in rainfall forecasting," in *2012 International Conference on Advances in ICT for Emerging Regions (ICTer)*, 2012, pp. 176-181.

[100] H. Nagahamulla, U. Ratnayake, and A. Ratnaweera, "Selecting most suitable members for neural network ensemble rainfall forecasting model," in *Recent Advances on Soft Computing and Data Mining: Proceedings of The First International Conference on Soft Computing and Data Mining (SCDM-2014) Universiti Tun Hussein Onn Malaysia, Johor, MalaysiaJune 16th-18th, 2014*, T. Herawan, R. Ghazali, and M. M. Deris, Eds. Cham: Springer International Publishing, 2014, pp. 591-601.

[101] T. Saba, A. Rehman, and J. S. AlGhamdi, "Weather forecasting based on hybrid neural model," *Applied Water Science,* journal article vol. 7, no. 7, pp. 3869-3874, November 01 2017.

[102] C. M. Kishtawal, S. Basu, F. Patadia, and P. K. Thapliyal, "Forecasting summer rainfall over India using genetic algorithm," *Geophysical Research Letters,* vol. 30, no. 23, pp. n/a-n/a, 2003.

[103] Y. L. Everingham, A. J. Clarke, and S. Van Gorder, "Long lead rainfall forecasts for the Australian sugar industry," *International Journal of Climatology,* vol. 28, no. 1, pp. 111-117, 2008.

[104] A. Clarke and S. Van Gorder, "Improving El Niño prediction using a space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content," *Geophysical Research Letters,* vol. 30, no. 7, 2003.

[105] S. Hawthorne, Q. J. Wang, A. Schepen, and D. Robertson, "Effective use of general circulation model outputs for forecasting monthly rainfalls to long lead times," *Water Resources Research,* vol. 49, no. 9, pp. 5427-5436, Sep 2013.

[106] X. He, H. Guan, X. Zhang, and C. T. Simmons, "A wavelet-based multiple linear regression model for forecasting monthly rainfall," *International Journal of Climatology,* vol. 34, no. 6, pp. 1898-1912, 2013.

[107] A. Schepen, Q. J. Wang, and D. Robertson, "Evidence for using lagged climate indices to forecast australian seasonal rainfall," *Journal of Climate*, vol. 25, no. 4, p. 1230, 2012.

[108] D. H. Cobon and N. R. Toombs, "Forecasting rainfall based on the Southern Oscillation Index phases at longer lead-times in Australia," *Rangeland Journal,* vol. 35, no. 4, pp. 373-383, 2013.

[109] A. Pui, A. Sharma, A. Santoso, and S. Westra, "Impact of the El Niño-Southern Oscillation, Indian Ocean Dipole, and Southern Annular Mode on daily to subdaily rainfall characteristics in east Australia," (in English), *Monthly Weather Review,* vol. 140, no. 5, pp. 1665-1682, May 2012 2012.

[110] G. A. Tularam, "Relationship between El Niño Southern Oscillation Index and rainfall (Queensland, Australia)," *WIT Press*, vol. 5, no. 4, pp. 378 - 391, 2010-12-02 2010.

[111] J. S. Risbey, M. J. Pook, P. C. McIntosh, M. C. Wheeler, and H. H. Hendon, "On the remote drivers of rainfall variability in Australia," *Monthly Weather Review,* vol. 137, no. 10, pp. 3233-3253, 2009.

[112] R. J. B. Fawcett and R. C. Stone, "A comparison of two seasonal rainfall forecasting systems for australia," *Australian Meteorological and Oceanographic Journal*, vol. 60, no. 1, pp. 15-23, Mar 2010.

[113] A. B. o. Meteorology, *Record-breaking La Nina events : an analysis of the La Nina life cycle and the impacts and significance of 2010-11 and 2011-12 La Nina events in Australia*. Bureau of Meteorology, 2012.

[114] Y. Iseri, G. C. Dandy, H. R. Maier, A. Kawamura, and K. Jinno, "Medium term forecasting of rainfall using artificial neural networks," in *International Congress on Modelling and Simulation: Advances and Applications for Management and Decision Making*, 2005, pp. 1834-1840.

[115] K. Trenberth and N. C. f. A. R. Staff. (2016). *The climate data guide: Nino SST indices (Nino 1+2, 3, 3.4, 4; ONI and TNI).* [Online]. Available: https://climatedataguide.ucar.edu/climate-data/nino-sst-indices-nino-12-3-34-4-oni-and-tni.

[116] B. Dong and A. Dai, "The influence of the Interdecadal Pacific Oscillation on temperature and precipitation over the globe," *Climate Dynamics,* journal article vol. 45, no. 9, pp. 2667-2681, 2015.

[117] M. J. Salinger, J. A. Renwick, and A. B. Mullan, "Interdecadal Pacific Oscillation and south Pacific climate," *International Journal of Climatology*, vol. 21, no. 14, pp. 1705-1721, 2001.

[118] B. J. Henley, J. Gergis, D. J. Karoly, S. Power, J. Kennedy, and C. K. Folland, "A tripole index for the Interdecadal Pacific Oscillation," *Climate Dynamics*, journal article vol. 45, no. 11, pp. 3077-3090, 2015.

[119] J. Kenyon and G. C. Hegerl, "Influence of modes of climate variability on global precipitation extremes," *Journal of Climate*, vol. 23, no. 23, pp. 6248-6262, 2010.

[120] N. Schneider and B. D. Cornuelle, "The forcing of the Pacific Decadal Oscillation," (in English), *Journal of Climate*, vol. 18, no. 21, pp. 4355-4373, 2005 Nov 01

2017-02-04 2005.
[121] M. Newman *et al.*, "The Pacific Decadal Oscillation, Revisited," (in English), *Journal of Climate*, vol. 29, no. 12, pp. 4399-4427, 2016 Jun 15

2017-02-04 2016.
[122] D. C. Verdon and S. W. Franks, "Long-term behaviour of ENSO: Interactions with the PDO over the past 400 years inferred from paleoclimate records," *Geophysical Research Letters*, vol. 33, no. 6, pp. 1-5, 2006.
[123] P. D. Jones, T. Jonsson, and D. Wheeler, "Extension to the North Atlantic oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland," *International Journal of Climatology*, vol. 17, no. 13, pp. 1433-1450, 1997.
[124] Bureau of Meteorology. *Climate data online* [Online]. Available: http://www.bom.gov.au/climate/data/.
[125] Royal Netherlands Meteorological Institute Climate Explorer. [Online]. Available: https://climexp.knmi.nl/start.cgi.
[126] Solar Influences Data Analysis Center. [Online]. Available: http://sidc.oma.be/.
[127] Earth System Research Laboratory. [Online]. Available: https://www.esrl.noaa.gov/.
[128] C. K. Folland, J. Shukla, J. Kinter, and M. J. Rodwell. (2002). *C20C: The climate of the twentieth century project* [Online]. Available: http://cola.gmu.edu/c20c/.
[129] D. S. Wilks, "Forecast Verification," in *International Geophysics*, vol. Volume 100, S. W. Daniel, Ed.: Academic Press, 2011, pp. 301-394.
[130] A. Elshorbagy, G. Corzo, S. Srinivasulu, and D. P. Solomatine, "Experimental investigation of the predictive capabilities of data driven modeling techniques in hydrology - part 1: Concepts and methodology," *Hydrol. Earth Syst. Sci.*, vol. 14, no. 10, pp. 1931-1941, 2010.
[131] R. C. Deo and M. Şahin, "Application of the extreme learning machine algorithm for the prediction of monthly effective drought index in eastern Australia," *Atmospheric Research*, vol. 153, pp. 512-525, 2// 2015.
[132] Y. L. Everingham, R. C. Muchow, R. C. Stone, and D. H. Coomans, "Using Southern Oscillation Index phases to forecast sugarcane yields: A case study for northeastern Australia," *International Journal of Climatology*, vol. 23, no. 10, pp. 1211-1218, Aug 2003.
[133] C. R. Rivero, H. D. Patiño, and J. A. Pucheta, "Short-term rainfall time series prediction with incomplete data," in *2015 International Joint Conference on Neural Networks (IJCNN)*, 2015, pp. 1-6.
[134] N. S. Philip and K. B. Joseph, "A neural network tool for analyzing trends in rainfall," *Computers & Geosciences*, vol. 29, no. 2, pp. 215-223, 3// 2003.
[135] S. Luke, *Essentials of metaheuristics*. Lulu Com, 2013.
[136] S. Sumathi and P. Surekha, *Computational intelligence paradigms: theory and applications using MATLAB* (no. Book, Whole). Boca Raton: CRC Press, 2010.

[137] O. Babatunde, L. Armstrong, J. Leng, and D. Diepeveen, "A genetic algorithm-based feature selection," *British Journal of Mathematics & Computer Science,* vol. 4, no. 21, pp. 889-905, 2014.

[138] R. Eberhart and J. Kennedy, "A new optimizer using particle swarm theory," in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science, 1995. MHS'95.*, , 1995: IEEE, pp. 39-43.

[139] J. Wu and E. Chen, "A novel nonparametric regression ensemble for rainfall forecasting using particle swarm optimization technique coupled with artificial neural network," in *6th International Symposium on Neural Networks (ISNN 2009)*, Wuhan, China, 2009: Springer Berlin Heidelberg, pp. 49-58.

[140] V. Kothari, J. Anuradha, S. Shah, and P. Mittal, "A survey on particle swarm optimization in feature selection," in *4th International Conference on Global Trends in Information Systems and Software Applications* India, 2012: Springer Berlin Heidelberg, pp. 192-201.

[141] C.-F. Tsai, W. Eberle, and C.-Y. Chu, "Genetic algorithms in feature and instance selection," *Knowledge-Based Systems,* vol. 39, pp. 240-247, 2// 2013.

[142] S. R. Devi, P. Arulmozhivarman, C. Venkatesh, and P. Agarwal, "Performance comparison of artificial neural network models for daily rainfall prediction," *International Journal of Automation and Computing,* journal article vol. 13, no. 5, pp. 417-427, October 01 2016.

[143] M. A. Ahmadi, "Prediction of asphaltene precipitation using artificial neural network optimized by imperialist competitive algorithm," *Journal of Petroleum Exploration and Production Technology,* journal article vol. 1, no. 2, pp. 99-106, 2011.

[144] C.-Y. Chiu and B. Verma, "An evolutionary algorithm based optimization of neural ensemble classifiers," in *18th International Conference on Neural Information Processing (ICONIP 2011)*, Shanghai, China, 2011: Springer Berlin Heidelberg, pp. 292-298.

[145] M. A. Ahmadi, R. Soleimani, M. Lee, T. Kashiwao, and A. Bahadori, "Determination of oil well production performance using artificial neural network (ANN) linked to the particle swarm optimization (PSO) tool," *Petroleum,* vol. 1, no. 2, pp. 118-132, 6// 2015.

[146] H. Hu, X. Hu, and X. Guan, "A method for crude oil output decline rate forecasting based on optimization of neural network by GA," in *2017 3rd International Conference on Information Management (ICIM)*, 2017, pp. 164-168.

[147] S. Asadi, J. Shahrabi, P. Abbaszadeh, and S. Tabanmehr, "A new hybrid artificial neural networks for rainfall–runoff process modeling," *Neurocomputing,* vol. 121, no. Supplement C, pp. 470-480, 2013/12/09/ 2013.

[148] I. A. Hameed, R. T. Bye, and O. L. Osen, "A comparison between optimization algorithms applied to offshore crane design using an online crane prototyping tool," in *International Conference on Advanced Intelligent Systems and Informatics* Cairo, Egypt, 2017: Springer International Publishing, pp. 266-276.

[149] T. Yamasaki, T. Honma, and K. Aizawa, "Efficient optimization of convolutional neural networks using particle swarm optimization," in *2017 IEEE Third International Conference on Multimedia Big Data (BigMM)*, 2017, pp. 70-73.

[150] H. Ding and W. Dong, "Chaotic feature analysis and forecasting of Liujiang river runoff," *Soft Computing*, journal article vol. 20, no. 7, pp. 2595-2609, 2016.

[151] A. Haidar and B. Verma, "A genetic algorithm based feature selection approach for rainfall forecasting in sugarcane areas," in *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*, 2016, pp. 1-8.

[152] A. Haidar and B. Verma, "Monthly rainfall categorization based on optimized features and neural network," in *30th Australasian Joint Conference on Artificial Intelligence* Melbourne, Australia, 2017: Springer International Publishing, pp. 208-220.

[153] A. Haidar and B. Verma, "A novel approach for optimizing climate features and network parameters in rainfall forecasting," *Soft Computing*, journal article pp. 1-12, August 03 2017.

[154] National Computational Infrastructure. (Mar. 24, 2018). [Online]. Available: https://geonetwork.nci.org.au/geonetwork/srv/eng/catalog.search#/metadata/f5023_1291_6579_1882.

[155] Z.-H. Zhou, *Ensemble methods: Foundations and algorithms*. CRC press, 2012.

[156] J.-H. Seo and Y.-H. Kim, "Genetic feature selection for very short-term heavy rainfall prediction," in *6th International Conference on Convergence and Hybrid Information Technology (ICHIT 2012)*, G. Lee, D. Howard, J. J. Kang, and D. Ślęzak, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 312-322.

[157] Z.-H. Zhou, J.-X. Wu, Y. Jiang, and S.-F. Chen, "Genetic Algorithm based Selective Neural Network Ensemble," in *17th International Joint Conference on Artificial Intelligence* Seattle, WA, USA, 2001, 1642200: Morgan Kaufmann Publishers Inc., pp. 797-802.

[158] Z.-H. Zhou, J. Wu, and W. Tang, "Ensembling neural networks: Many could be better than all," *Artificial Intelligence*, vol. 137, no. 1–2, pp. 239-263, 5// 2002.

[159] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, journal article vol. 33, no. 1, pp. 1-39, 2010.