Sudan University of Science and Technology

Faculty of Computer Science and Information Technology

Rainfall Forecasting in Sudan Using Computational Intelligence

(توقعات الأمطار في السودان باستخدام الحوسبة الذكية)

A Dissertation submitting to PhD degree

Presented to fulfillment of the requirements of degree of Doctor of Philosophy in Computer Science

Candidate: Nazim Osman Bushara

Supervisor: Prof. Ajith Abraham

February 2016

Approval Page

Name of Candidate: Nazim Osman Bushara Elhassan

Thesis title: Rainfall Forecasting in Sudan using Computational Intelligence

توقعات الأمطار في السودان باستخدام الحوسبة الذكية

Approved by:

1. External Examiner

Name: Abubakr ElSiddig Mirghani ElHussein

Signature: .... Date: Feb 3rd, '16

2. Internal Examiner

Name: Mohamed E. M. Mustafa Musa

Signature: .... Date: 3/2/2016

3. Supervisor

Name: Ajith Abraham

Signature: .... Date: 3/2/2016

Sudan University of Science and Technology
College of Graduate Studies

Declaration

I, the signing here-under, declare that I'm the sole author of the Ph.D. thesis entitled ……Rain Forecasting in Sudan Using Computational Intelligence……

which is an original intellectual work. Willingly, I assign the copy-right of this work to the College of Graduate Studies (CGS), Sudan University of Science & Technology (SUST). Accordingly, SUST has all the rights to publish this work for scientific purposes.

Candidate's name: Nazim Osman Bushara Elhussan

Candidate's signature: Nazim **Date:** 15-3-2016

إقرار

أنا الموقع أدناه أقر بأننى المؤلف الوحيد لرسالة الدكتوراه المعنونة ……

توقعات الامطار في السودان باستخدام الحوسبة الذكية ……

وهى منتج فكري أصيل . وباختياري أعطى حقوق طبع ونشر هذا العمل لكلية الدراسات العليا جامعه السودان للعلوم والتكنولوجيا ،عليه يحق للجامعه نشر هذا العمل للأغراض العلمية .

اسم الدارس : ناظم عثمان بشارة الحسن

توقيع الدارس : Nazim …… التاريخ : ١٥/٣/٢٠١٦م

Dedication

This thesis is dedicated to spirit of my father, who had always loved me unconditionally and whose good examples have taught me to work hard for the things that I aspire to achieve.

This thesis is also dedicated to my family, for their endless love, support and encouragement.

Acknowledgements

I am grateful to the God for the good health and wellbeing that were necessary to complete this research.

I would like to thank all those in charge of doctoral program by courses and research in computer science at Sudan University of science and technology (professors and coordinators).

I would like to express my sincere gratitude to my supervisor, Prof. Ajith Abraham, whose expertise, understanding, and patience, added considerably to my graduate experience. I appreciate his vast knowledge and skill in many areas and his assistance in writing the proposal, scientific papers, and this thesis.

I wish to thank everyone who helped me complete this thesis. Without their continued efforts and support, I would have not been able to bring my work to a successful completion.

Dr. Mohammed El hafiz and CBS: for obtaining research data.

My wife, Ala El deen and Shams El deen: for their help to produce soft copy of the data set.

Ahmmed El aaqib: for his assistance in dealing with my computer breakdowns

Dr. Tarig Hassan: for advice and the use of his GIS software.

Dr. Mohammed Bakri and Dr. Abd El Hamid: For formatting and editing.

I would also like to thank my family for the support they provided me through my entire life and in particular, I must acknowledge my mother, sisters, wife and children, without whose love, encouragement and endless assistance, I would not have finished this thesis.

Abstract

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time at a given location. Human kind has attempted to predict the weather since ancient times. Generating predictions of meteorological events is very complex process, because the atmosphere is unstable and the systems responsible for the events are the culmination of the instabilities and involve nonlinear interaction between different spatial scales from kilometers to hundreds of kilometers. The chaotic nature of the atmosphere limits the validity of deterministic forecasts, but the increasing economic cost of adverse weather events provides a strong reason to generate more accurate and updated weather forecasts.

Weather forecasting (particularly rainfall prediction) is one of the most imperatives, important and demanding operational tasks and challenge made by meteorological services around the world. It is a complicated procedure that includes numerous specialized fields of knowledge. The task is complicated because in the field of meteorology all decisions are to be taken with a degree of uncertainty, because the chaotic nature of the atmosphere limits the validity of deterministic forecasts. Long term Rainfall prediction is very important for countries whose economy depends mainly on agriculture, like many of the third World countries. It is widely used in the energy industry and for efficient resource planning and management including famine and disease control, rainwater catchment and ground water management. This thesis studies long term rainfall prediction in Sudan using computational intelligence.

Monthly meteorological data obtained from Central Bureau of Statistics, Sudan from 2000 to 2012, for 24 meteorological stations distributed among the country has been used.

The relationship of rainfall in Sudan with some important parameters is investigated and determined the most influencing variables on rainfall among the available ones.

The performance of base and Meta algorithms to deal with rainfall prediction problem is explored and, compared.

A novel method to develop long-term rainfall prediction model by using ensemble technique is proposed. The new novel ensemble model is constructed based of Meta classifier Vote combined with three base classifiers. Several neuro-fuzzy Models using different types of membership functions, different optimization methods and different dataset ratios for training and testing are built.

The proposed models are evaluated and compared by using correlation coefficient, mean absolute error and root mean-squared error as performance metrics. The empirical results illustrate that the ANFIS neuro-fuzzy system and the ensemble Vote+3 models are able to capture the dynamic behavior of the rainfall data and they produced satisfactory results, so they may be very useful in long-term rainfall prediction.

Spatial analysis of rainfall in Sudan is conducted for the interval 2000-2012 on three levels (towns, states and regions) and rainfall maps are obtained.

ملخص

التنبؤ بالطقس هو تطبيق للعلم والتكنولوجيا لتوقع حالة الجو المستقبلية في مكان معين، وقد حاولت البشرية التنبؤ به منذ القدم. انتاج تنبؤات لأحداث الأرصاد الجوية عملية معقدة جدا، لأن الجو غير مستقر والأنظمة المسؤولة عن الأحداث هي تتويج لعدم الاستقرار وتنطوي على تفاعلات غير خطية بين نطاقات مكانية مختلفة تتراوح بين كيلومترات الى مئات الكيلومترات. فالطبيعة الفوضوية للغلاف الجوي تحد من صحة التنبؤات القطعية، لكن التكلفة الاقتصادية المتزايدة للأحداث الجوية السيئة تقدم سببا قويا لتوليد تنبؤات جوية أكثر دقة وتحديثا.

تدرس هذه الأطروحة توقعات هطول الأمطار على المدى الطويل في السودان باستخدام الحوسبة الذكية. حيث يعتبر التنبؤ بالطقس وخصوصاً الأمطار أحد المهام التنفيذية الأكثر إلحاحاً، والتي تمثل تحدي يواجه خدمات الأرصاد الجوية في جميع أنحاء العالم وهو إجراء معقد يتضمن العديد من المجالات المتخصصة للمعرفة، ويعتبر مهمة معقدة لأنه في مجال الأرصاد الجوية جميع القرارات يتعين أتخاذها مع وجود درجة من عدم اليقين، لأن الطبيعة الفوضوية للغلاف الجوي تحد من صحة التنبؤات القطعية. التنبؤ طويل الاجل لهطول الامطار مهم جدا بالنسبة للدول التي يعتمد أقتصادها أساسا على الزراعة، على غرار العديد من بلدان العالم الثالث. كما أنه يستخدم على نطاق واسع في صناعة الطاقة والتخطيط الفعال للموارد وإدارتها بما في ذلك مكافحة المجاعات و الأمراض، تجميع مياه الأمطار وإدارة المياه الجوفية.

أُستخدمت بيانات الأرصاد الجوية التي تتمثل في المتوسطات الشهرية التي تم الحصول عليها من الجهاز المركزي للإحصاء السودان للفترة ٢٠٠٠– ٢٠١٢، لعدد ٢٤ محطة أرصاد جوية موزعة على طول البلاد.

تم التحقق من العلاقة بين الأمطار في السودان و بعض العوامل الهامة وتوصلنا للمتغيرات الأكثر تأثيرا على هطول الأمطار.

أقترحت الدراسة طريقة جديدة لتطوير نموذج توقع طويل الأجل للأمطار بإستخدام تقنية الفرقة. أستند هذا النموذج بشكل أساسي على مصنف التصويت ميتا بجانب ثلاثة مصنفات أساسية، كما تم بناء عدة نماذج عصبية مبهمة بإستخدام أنواع مختلفة من الدوال الأعضاء، طرق التحسين ونسب بيانات للتدريب والاختبار.

تم تقييم النماذج المقترحة ومقارنتها بإستخدام معامل الإرتباط، متوسط الخطأ المطلق وجذر تربيع متوسط الخطأ، كمقاييس للأداء. أظهرت النتائج أن كل من ANFIS neuro-fuzzy system و ensemble

Vote+3 قادران على التقاط السلوك الحركي لهطول الأمطار، لأنهما كشفا عن نتائج مرضية، وعليه قد يكونا مفيدين في التنبؤ طويل المدى بهطول الأمطار.

تم عمل التحليل المكاني لهطول الأمطار في السودان للفترة ٢٠٠٠- ٢٠١٢ على ثلاثة مستويات (المدن ، الولايات والأقاليم) ومن ثم تم إظهاره في خرائط للأمطار.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
ARIMA	Auto Regressive Moving Average
AS	Attribute Selection
BP	Back Propagation
BPN	Back Propagation neural network
C4.5	A decision tree induction algorithm
CART	Classification and Regression Tree
CBS	Central Bureau of Statistics
CC	Correlation Coefficient
CI	Computational Intelligence
DE	Differential Evolution algorithm
DM	Data Mining
EFuNN	Evolving Fuzzy Neural Network
ENN	Ensemble Neural Network
ERNN	Elman Recurrent Neural Network
FF-BP	Feed Forward-Back Propagation
FIS	Fuzzy Inference System
FRNN	Fuzzy Recurrent Neural Network
GA	Genetic Algorithm
GDP	Gross Domestic Product

GIS	Geographic Information System
GP	Gaussian Processes
GPS	Global Position System device
GRNN	Generalized Regression Neural Network
HFM	Hopfield Model
K*	K-Star algorithm
KDD	Knowledge Discovery in Databases
K-NN	K-Nearest Neighbour
K–PLSR	Kernel Partial Least Squares Regression
LM	Levenberg Marquardt algorithm
LPM	Linear Perturbation Model
LR	Linear Regression
M5R	M 5 Rules algorithm
MAE	Mean Absolute Error
MANFIS	Modified Adaptive Neuro-Fuzzy Inference System
MCS	Multi-Classifier System
MF	Membership Function
MFGM	Markov–Fourier Gray Model
ML	Machine learning
MLFN	Multilayer Feed Forward Neural network
MLP	Multi-Layered Perceptron
MLR	Multiple Linear Regression
MOS	Model Output Statistics technique
NMSE	Normalized Mean Square Error
NNLPM	Nearest Neighbor Linear Perturbation Model

NS	Nash–Sutcliffe efficiency coefficient
PAC	Probably Approximately Correct
PCA	Principal Component Analysis
PSO	Particle Swarm Optimization
PV	Photovoltaic system modeling
RBF–NN	Radial Basis Function Neural Network
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Network
RT	REP-Tree algorithm
SLM	Simple Linear Model
SLP	Sea Level Pressure
SNN	Single Neural Network
STTF	Short-Term Temperature Forecasting
SVM	Support Vector Machine
SVR	Support Vector Regression
SWE	Snow Water Equivalent
TDNN	Time Delay Neural Network
TLRN	Time Lagged Recurrent Network
UC	User Classifier algorithm
UETS	Unified Evolutionary Training Scheme

1. Introduction

This Chapter is organized to describe the research background, which includes key concepts such as weather forecasting, rainfall prediction, computational intelligence, machine learning, data mining, Meta-learning, ensemble methods, Neuro-fuzzy system and heat maps. Study area and statement of the research problem have been described, followed by the objectives and questions of the research. The last Section concludes with the thesis organization and summary of each Chapter at the end.

1.1 Background

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time at a given location [1] . Human kind has attempted to predict the weather since ancient times.

Today, weather forecasts are made by collecting quantitative data about the current state of the atmosphere and using scientific understanding of atmospheric processes to project how the atmosphere will evolve. The traditional weather forecasting approaches are based on [1]:

(a) Empirical approach

(b) Dynamical approach

The first approach is based upon the occurrence of analogues and is often referred to by meteorologists as analogue forecasting. This approach is useful for predicting local-scale weather if recorded cases are plentiful. The second approach is based on the equations and forward simulations of the atmosphere, and is often referred to as computer modeling. The dynamical approach is only useful for modeling large-scale weather phenomena Because of the grid Coarseness and may not predict short-term weather efficiently.

There are several types of weather forecasts made in relation to time [2]:

- A short-range forecast is a weather forecast made for a time period up to 48 hours.

- Medium range forecasts are for a period extending from about three days to seven days in advance.
- Long-range forecasts are for a period greater than seven days in advance but there are no absolute limits to the period.

The success of the seasonal forecasts depends on a detailed knowledge of how the atmosphere and ocean interact. Short-range forecast predictions, where the forecast is made for a time period for today or tomorrow (up to 48 hours), are generally more accurate than the other types of forecasts. Weather forecasts still have their limitations despite the use of modern technology and improved techniques to predict the weather. For example, weather forecasts for today or tomorrow are likely to be more dependable than predictions about the weather about two weeks from now. Some sources state that weather forecast accuracy falls significantly beyond 10 days [2]. Weather forecasting is complex and not always accurate, especially for days further in the future, because the weather can be chaotic and unpredictable. For example, rain or snow cannot always be predicted with a simple yes or no. Moreover, the Earth's atmosphere is a complicated system that is affected by many factors and can react in different ways.

Long-range weather forecasts are widely used in the energy industry, despite their limited skill; long-range forecasts can still be a valuable tool for managing weather risk. Long-term Prediction of rainfall has several benefits for efficient resource planning and management including agriculture, famine and disease control, rainwater catchment and ground water management [2].

Weather forecasting (particularly rainfall prediction) is one of the most imperatives, important and demanding operational tasks and challenge made by meteorological services around the world. It is a complicated procedure that includes numerous specialized fields of knowledge. The task is complicated because in the field of meteorology all decisions are to be taken with a degree of uncertainty, because the chaotic nature of the atmosphere limits the validity of deterministic forecasts [3].

Rainfall prediction is very important for countries whose economy depends mainly on agriculture, like many of the third World countries [3]. In general, considered climatic phenomena and the precipitation of non-linear phenomena in nature, leading to what is known as the "butterfly effect". Required parameters to predict rainfall, extremely complicated and unclear so that the uncertainty in the prediction using all these criteria enormous even for a short period.

The most prevalent techniques [4] used to predict rainfall is numerical and statistical methods. Although research in these areas takes place for a long time, the successes of these models is rarely concrete because these models have been found to be very accurate in calculation, but not in prediction as they cannot adapt to the irregularly varying patterns of data which can neither be written in form of a function, or deduced from a formula. Numerical weather prediction uses current weather conditions, as input into mathematical models of the atmosphere there is a limited success in forecasting weather parameters using numerical models [5]. The accuracy of the models depends on the initial conditions, which are inherently incomplete. These systems are not able to achieve satisfactory results in domestic cases, short-term, as well as the weak performance in order to predict the long-term seasonal rain even for a large spatial scale. However, the atmospheric circulation is quite sensitive to initial conditions [6].

Predictability by the numerical weather forecast is limited within, say, 1 week at present Because of this characteristic. We have no idea how to handle the chaotic behaviors of the atmosphere and make a long-range forecast by the numerical weather prediction yet.

Statistical models analyze historical data and identify relationships between precursors and consequences. They are distinct from dynamical models in their lack of use of any physical equations. In a statistical approach, certain variables are typically designated as predictors, while others need to be predicted. Two main drawbacks of the statistical models are:

1. Statistical models are not useful to study the highly nonlinear relationships between rainfall and its predictors, even if one considers models like power regression.

2. There is no ultimate end in finding the best predictors. It will never be possible to get different sets of regional and global predictors to explain the variability of the two neighboring regions having distinguished rainfall features.

Computational intelligence (CI) is the study of adaptive mechanisms which enable or facilitate intelligent behavior in complex and changing environments. In other words, CI is mainly about the design of algorithmic models to solve complex problems [7]. As such, computational intelligence combines artificial neural networks, evolutionary computing, swarm intelligence and fuzzy systems. In addition, CI also embraces biologically inspired algorithms such as swarm intelligence and artificial immune systems, which can be seen as a part of evolutionary computation, and includes broader fields such as image processing, data mining, and natural language processing. Furthermore other formalisms: Dempster–Shafer theory, chaos theory and many-valued logic are used in the construction of computational models [8].

In general, Machine learning (ML) involves adaptive mechanisms that enable computers to learn from experience, learn by example and learn by analogy. Learning capabilities can improve the performance of an intelligent system over time. Machine learning mechanisms form the basis for adaptive systems. The most popular approaches to machine learning are artificial neural networks and genetic algorithms [9].

Scientists have tried to forecast the meteorological characteristics using a large set of methods, some of them more accurate than others. Lately, data mining methods developed and can be successfully applied in this domain. Data mining (DM) is about solving problems by analyzing data already present in databases. DM is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semi-automatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities [10].

Data Mining is the search for the relationships and global patterns that exist in large databases but are hidden among vast amount of data, such as the relationship between patient data and their medical diagnosis [11]. This relationship represents valuable knowledge about the database, and the objects in the database, if the hidden database is a faithful mirror of the real world registered by the database. Data Mining refers to using a variety of techniques to identify nuggets of information or decision making knowledge in the database and extracting these in such a way that they can be put to use in areas such as decision support, prediction, forecasting and estimation. The data is often voluminous, but it has low value and no direct use can be made of it. It is the hidden information in the data that is useful [11].

Meteorological data mining [12] is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Useful knowledge can play important role in understanding the climate variability and climate prediction. In turn, this understanding can be used to support many important sectors that are affected by climate like agriculture, vegetation, water resources and tourism.

Meta-learning is a technique that seeks to compute higher-level classifiers (or classification models), called meta-classifiers, that integrate in some principled fashion multiple classifiers computed separately over different databases. Meta-learning means learning from the classifiers produced by the inducers and from the classifications of these classifiers on training data.

Meta-learning improves efficiency by executing in parallel the base-learning processes (each implemented as a distinct serial program) on (possibly disjoint) subsets of the training data set (a data reduction technique). This approach has the advantage, first, of using the same serial code without the time-consuming process of parallelizing it, and second, of learning from small subsets of data that fit in main memory.

Meta-learning improves predictive performance by combining different learning systems each having different inductive bias (e.g. representation, search heuristics, search space) [13]. By combining separately learned concepts, meta-learning is expected to derive a higher-level learned model that explains a large database more accurately than any of the individual learners. Furthermore, meta-learning constitutes a scalable machine learning method since it can be generalized to hierarchical multi-level meta-learning.

The idea of ensemble methodology is to build a predictive model by integrating multiple models. It is well-known that ensemble methods can be used for improving prediction performance [14].

Building an ensemble consists of two steps: (1) constructing varied models and (2) combining their estimates. One may generate component models by, for instance, varying case weights, data values, guidance parameters, variable subsets, or partitions of the input space. Combination can be accomplished by voting, but is primarily done through model estimate weights [15].

Diversity is a crucial condition for obtaining accurate ensemble [16-19]. According to Hu [20], diversified classifiers lead to uncorrelated classifications, which in turn improve classification accuracy. However, in the classification context, there is no complete and agreed upon theory to explain why and how diversity between individual models contributes toward overall ensemble accuracy [21].

An important aspect of ensemble methods is to determine how many base classifiers and which classifiers should be included in the final ensemble. Several algorithms, such as bagging, predetermine the ensemble size, by using a controlling parameter such as number of iterations that can be set by the user. Other ensemble algorithms try to determine the best ensemble size while training. When new members are added to the ensemble, we check if the performance of the ensemble has improved. If it is not, the procedure stops and no new base classifier are trained. Usually these algorithms also have a controlling parameter, which bounds the number of base classifiers in the ensemble. An algorithm that decides when a sufficient number of classification trees have been created was proposed by Robert et al. [22].

Ensemble methodology imitates our second nature to seek several opinions before making a crucial decision. The core principle is to weigh several individual pattern classifiers, and combine them in order to reach a classification that is better than the one obtained by each of them separately. Researchers from various disciplines such as pattern recognition, statistics, and machine learning have explored the use of ensemble methods since the late seventies. Given the growing interest in the field, it is not surprising that researchers and practitioners have a wide variety of methods at their disposal.

An ensemble is largely characterized by the diversity generation mechanism and the choice of its combination procedures. While ensemble approaches to classification usually make use of non-linear combination methods like majority voting; linearly weighted ensembles naturally tackle regression problems. These types of ensembles have a much clearer framework for explaining the role of diversity than voting methods [17].

A hybrid intelligent system is one that combines at least two intelligent technologies [9], For example, combining a neural network with a fuzzy system results in a hybrid neuro-fuzzy system.

Neuro-fuzzy or fuzzy-neural structures [23], has largely extended the capabilities of both technologies in hybrid intelligent systems. The advantages of neural networks in learning and adaptation and those of fuzzy logic systems in dealing with the issues of human-like reasoning on a linguistic level, transparency and interpretability of the generated model, and handling of uncertain or imprecise data, enable building of higher-level intelligent systems. The synergism of integrating neural networks with fuzzy logic technology into a hybrid functional system with low-level learning and high-level reasoning transforms the burden of the tedious design problems of the fuzzy logic decision systems to the learning of connectionist neural networks. In this way the approximation capability and the overall performance of the resulting system are enhanced.

A neuro-fuzzy system is, in fact, a neural network that is functionally equivalent to a fuzzy inference model. It can be trained to develop *If-Then* fuzzy rules and determine membership functions for input and output variables of the system. Expert knowledge can be easily incorporated into the structure of the neuro-fuzzy system. At the same time, the connectionist structure avoids fuzzy inference, which entails a substantial computational burden.

The adaptive neuro-fuzzy inference system is a common approach in which the two techniques such as a neural network and a fuzzy logic get combined [24] to create a complete shell. Basically the system of ANFIS applies the technique of the artificial neural network learning rules to determine and tune the fizzy inference systems' structure and parameters. A number of important features of ANFIS can help the system accomplish a task brilliantly; these features are characterized as easy to implement, fast and accurate learning, strong generalization abilities, excellent explanation facilities through fuzzy rules, and easy to incorporate both linguistic and numeric knowledge for problem solving [25-29].

In Neuro-Fuzzy technique, a neural network learning mechanism is introduced to design the fuzzy system so that the structure and parameters which identify the fuzzy rules are accomplished by adopting and optimizing the topology and the parameters of corresponding neuro fuzzy network based on data sets. The system is considered to be an adaptive fuzzy inference system with the capability of learning fuzzy rules from data and as a connectionist architecture provided with linguistic meaning. Jang developed Neuro-fuzzy inference expert system that works in Takagi-Sugeno type fuzzy inference system [30-32]. This is called Adaptive Neuro Fuzzy Inference System (ANFIS) and is one of the most successful schemes, which combine the benefits of these two powerful paradigms into a single capsule.

There are several features of ANFIS, which enable it to achieve great success in a wide range of scientific applications. The attractive features of an ANFIS include: easy to implement, fast and accurate learning, strong generalization abilities, excellent explanation facilities through fuzzy rules, and easy to incorporate both linguistic and numeric knowledge for problem solving [27].

There is a range of techniques for visualizing the data recorded by an eye tracker such as simple plots, fixation maps and heat maps [33].

A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. Fractal maps and tree maps both often use a similar system of color-coding to represent the values taken by a variable in a hierarchy [34].

Heat maps uses differences in shading or color saturation to encode quantities, which means they are not suited for giving accurate readings, but they allow a lot of data points to be displayed side by side and can be good for spotting patterns and for giving an overview of the data.

Heat maps can simply be colored rectangles or circles in a row or on a grid, but heat mapping can also be combined with other chart types, giving them an extra quantitative axis. A common use case is heat mapping on geographical maps [35].

In the heat maps colors should be used consistently throughout a design; the same colors should not be used to represent different things. When using color to represent categories it's good to think about what people associate with different colors. Using a color the viewer naturally links with the category can reduce their cognitive load [36]. Also keep in mind that some colors are difficult to distinguish for colorblind persons, red and green being the most common.

The number of colors should be kept limited, and the colors used should mainly be soft and muted. Stronger more saturated colors can be used sparingly to highlighting important information. Too many and bright colors makes it hard to focus on the information and bright colors can also affect how the size of areas are perceived [37].When using colors to represent ordinal or quantitative data such as in a heat map a perceptually ordered color scheme should be used. The strongest cue for perceived order is a color's lightness, so a good choice of scale is using a single color hue and gradually altering the lightness, going from light to dark. It's also possible to use a scale with changing hue, as long as the lightness also changes. If the data has a natural midpoint, such as a mean or zero value, a diverging color scheme, with two colors meeting in the middle, can be used.

Using light colors for smaller values and dark colors for larger values is a common cartographic convention [38]. This convention is perhaps confusing when using the common yellow-red color scale for people familiar with black-body radiation and color temperature, since a color gets darker and redder as the temperature decreases.

A rainfall map like heat map, it is a geographical representation of data in which values are represented by colors. Rainfall maps allow users to understand and analyze complex data sets. Rainfall maps make prediction rainfall data more realistic because, they connected rainfall predictions with their geographical locations. They leverage the human visual system to help users gain deeper and faster insights than other visualizations. Users can visually aggregate, determine relevance and detect micro-patterns in their data in ways other visualizations can't match.

1.2 Study Area

Sudan is located in northeastern Africa Between latitudes 8° and 23° north of the equator and longitudes 22° and 38° east (See figure 1.1). Sudan is bordered by Egypt to the north, the Red Sea to the northeast, Eritrea and Ethiopia to the east, South Sudan to the south, the Central African Republic to the southwest, Chad to the west and Libya to the northwest. Sudan is the third largest country in Africa. It had been the largest until the 2011 separation of South Sudan. Sudan's total land area amounts to some 1,886,068 km², with 18,630 km² of irrigated land.

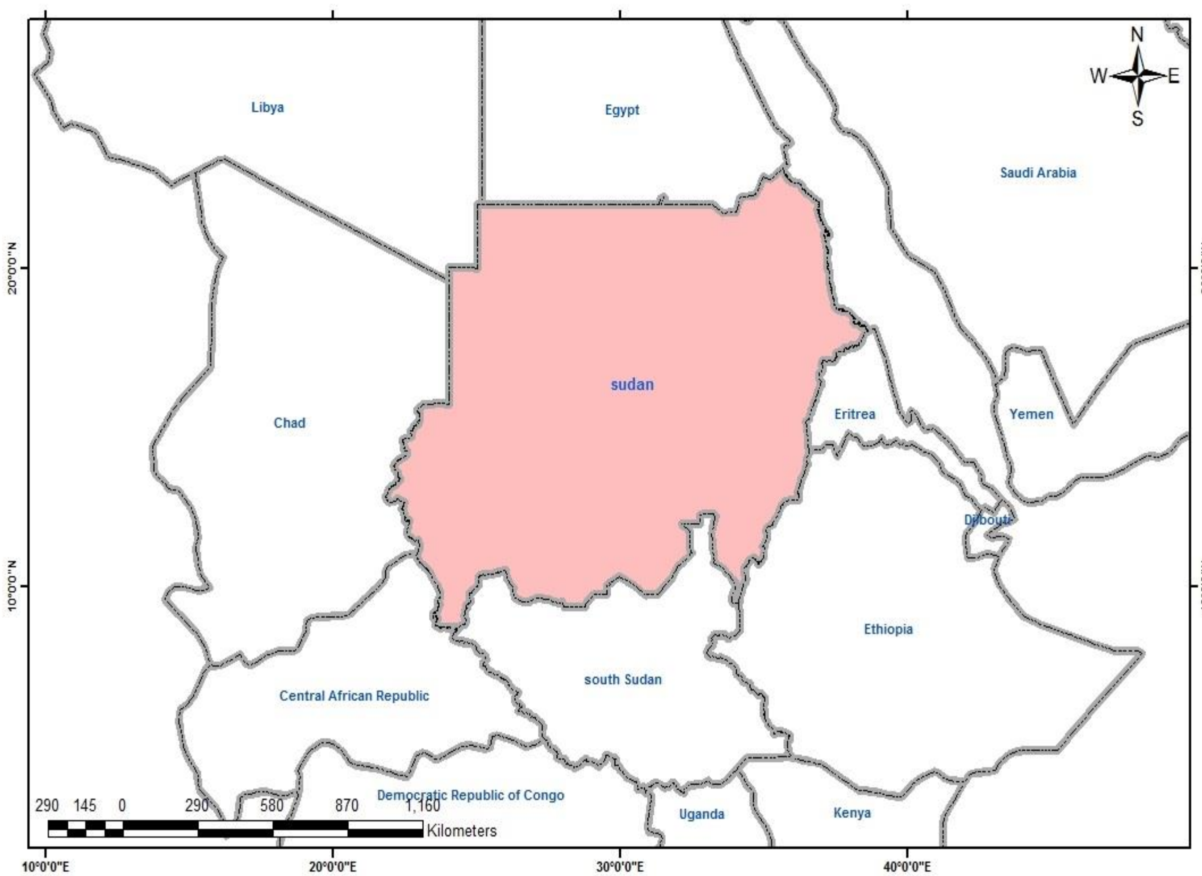


Figure 1.1. The geographical location of Sudan and neighboring countries.

1.2.1 Climate

Although Sudan lies within the tropics, the climate ranges from hyper-arid in the north to tropical wet-and-dry in the far southwest. Temperatures do not vary greatly with the season at any location; the most significant climatic variables are rainfall and the length of the dry season. Variations in the length of the dry season depend on which of two airflows predominates dry northeasterly winds from the Arabian Peninsula or moist southwesterly winds from the Congo River basin [39].

From January to March, the country is under the influence of the dry northeaster lies, continental trade winds [40]. There is practically no rainfall countrywide except for a small area in northwestern Sudan in where the winds have passed over the Mediterranean bringing occasional light rains. The sky is fully clear of clouds and the sunshine duration is near the theoretical maximum. By early April, the moist south westerlies have reached southern Sudan, bringing heavy rains and thunderstorms. By July the moist air has reached Khartoum, and in August it extends to its usual northern limits around Abu Hamad, although in some years the moist air mass may not even reach Khartoum. The flow becomes weaker as it spreads north. In September the dry north easterlies begin to strengthen and to push south and by the end of December they cover the entire country. Yambio, close to the border with Zaire, has a nine-month rainy season (April–December) and receives an average of 1,142 millimeters (45.0 in) of rain each year; Khartoum has a three-month rainy season (July–September) with an annual average rainfall of 161 millimeters (6.3 in); Barbara receives showers in August that produce an annual average of only 74 millimeters (2.9 in) [40].

In some years, the arrival of the south westerlies and their rain in central Sudan can be delayed, or they may not come at all. If that happens, drought and famine follow. The decades of the 1970s and 1980s saw the south westerlies frequently fail, with disastrous results for the Sudanese people and economy [40].

Temperatures are highest at the end of the dry season when cloudless skies and dry air allow them to soar. The far south, however, with only a short dry season, has uniformly high temperatures throughout the year. In Khartoum, the warmest months are May and June, when average highs are 41.9 °C (107.4 °F) and temperatures can even reach 48 °C (118.4 °F) [39]. Northern Sudan, with its short rainy season, has very high daytime temperatures year round, except for winter months in the northwest where there is precipitation from the Mediterranean in

January and February. Conditions in highland areas are generally cooler, and the hot daytime temperatures during the dry season throughout central and northern Sudan fall rapidly after sunset. Lows in Khartoum average 15 °C (59 °F) in January and have dropped as low as 6 °C (42.8 °F) after the passing of a cool front in winter [39].

The Haboob, a violent dust storm, can occur in central Sudan when the moist southwesterly flow first arrives (May through July). The moist, unstable air forms thunderstorms in the heat of the afternoon. The initial down flow of air from an approaching storm produces a huge yellow wall of sand and clay that can temporarily reduce visibility to zero [39].

Desert regions in central and northern Sudan are among the driest and the sunniest places on Earth: the sunshine duration is always uninterrupted year-round and soar to above 4,000 h in the best cases, or about 91% of the time and the sky is cloudless all the time. Areas around Wadi Halfa and along the Egyptian border can easily pass many years or many decades without seeing any rainfall at all. They are also among the hottest places during their summertime and their "wintertime”: averages high temperatures routinely exceed 40 °C (104 °F) for four to nearly six months a year to reach a maximum peak of about 45 °C (113 °F) in some places and averages high temperature remain above 24 °C (75.2 °F) in the northernmost region and above 30 °C (86 °F) in places such as Atbara or Meroe [40].

1.2.2 Agriculture

In Sudan Agriculture is one of the leading sectors of the country because of its water and human resources. Water resources include Nile River and its tributaries, rain, Lagoons seasonal and groundwater

Human resources: in the last census in 2008 the population of Sudan is 30.9 million, and estimated in 2015 by the Central Bureau of Statistics (CBS) to be 38.44 million [41]. The relative distribution of economic activity according to the 2008 census showed that 43.7% of the population works in agriculture, forestry and fishing. The relative distribution of occupations according to the same source was 27.4% are farming, herding and fishing [41].

The importance of agriculture in Sudan:

It represents 34.8% of GDP in 2012.

It contributes about 15.3% of the country's non-oil exports for the year 2010. 27.4% of the population acts in the agricultural sector as the fifth population census Indicators 2008 [42].

According to the Ministry of Agriculture and Irrigation the agricultural sector production divided into the Plant production and animal production [41]. Plant production is divided into several Agricultural sectors which appears in Table 1.1.

Table 1.1: Agricultural sectors, areas and the most important crops.

Sector	Area	The most important production areas	The most important crops produced
Irrigated sector and spate sector.	It represents 9% of the total cultivated area per year.	El Gezira project, Rahad agricultural project, New Halfa Agricultural Commission, Suki project, Northern, River Nile, Khartoum, Sinnar, Kassala, White Nile, the Red Sea.	Corn, wheat, cotton, pulses, spices, vegetables, fruit, fodder, sunflower, peanut.
Rainy semi-automated sector.	The average tenure between 1000-1500 square meters for individuals, while the average tenure for companies to hundreds of thousands of acres.	El Gedaref, Sinnar, White Nile, Kassala, Blue Nile, South Kordofan.	Corn, sesame, millet, sunflower, guar, Egyptian cotton.
Traditional rain-fed sector.	This sector is characterized by the small size of holdings.	Great Kordofan, Great Darfur, El Gezira, Sinnar, Kassala, White Nile, Blue Nile.	Millet, corn, sesame, peanuts, hibiscus, watermelon seeds, cowpea, maize.

Animal Production (Animal Resources and Fisheries products):

The Animal Resources and Fisheries sector is Occupies second order after the oil sector; it contributes about 49% of the contribution of the agricultural sector as a whole. Also it contributes about 17% of non-oil export earnings. And contribute to food security by providing meat, fish, milk and dairy products and eggs in addition to its contribution to the leather goods, fertilizer and energy load and traction [41].

According to the Ministry of Livestock, Fisheries and Livestock pastures have doubled numbers of livestock from 59.8 million head of estimates in the 1990/1991 year to 104.3 million head in 2012, an increase of 70.6% and an increase of 89.86% for sheep and an increase of 98.7 goats and 40.95% for cows and 67.86% for camels [41].

1.2.3 Forests

After the separation of the south in 2010, forest management faced fundamental change in the area and the composition of forests in Sudan. Sudan has lost two thirds of its forests area, 68% of the total forest area went to the south of Sudan, and the remaining portion covered by the desert and semi-desert is much more than half the area. This change constitutes surprises milestone in the history of forest cover the country that they were the reasons behind the decline supplier through the ages is the traditional reasons for the removal of trees to provide energy and building materials to the conversion of forest land for agriculture and housing to negatively affect by forest fires and over-grazing. Although the forest cover has decreased significantly and sequentially over the years where forests covered 40% of the area of Sudan early last century, and then decreased to 34% in 1958 and to 29% in 2000, but it rolled to about 14% after the separation of the south area. This new circumstance resulted a need to create new forests convergence the current forest area, the forests must be cover an area of up to 20% of the country area according to forest policy of 1986 which is currently being work out.

Following this dry region, province of a few savannah rain and covers 27% of the area of the Sudan, and represents the majority of the current area of the country and contains most of its forest coat including Arabic gum belt and other natural forests and most of its Nile forests [41].

Types of forest in Sudan:

Forests in Sudan are divided according to the administrative classification to:

Forests belong to the states and federal forests. Also the forests are divided by natural division into:

Natural Nile forests: show along the banks of the Nile River in the (Northern, River Nile, El Gezira, White Nile, Blue Nile) and planted forests. Accordingly to the National Forestry Commission, the reserved forest area which under the booking was 12.1 billion hectares in 2012 [42].

1.2.4 Arabic Gum

Arabic Gum belt in Sudan extends between latitudes 10-14 where 200-400 mm of rain rate per year. Arabic gum has been considered one of the most important national goods. Sudan has a leadership role in terms of production represents 75 % to 85% of the total global production.

Arabic Gum belt represents five area of the Sudan covers nine states of North and South Kordofan, West, North and South Darfur, White Nile, Blue Nile, Sinnar, Gedaref. It support more than five million people in the provision of livelihoods and in their stability. According to the National Forestry Commission, the production of Arabic gum 28.5 thousand tons [42].

1.2.5 Labor Force

The occupational distribution of the labor force for total Sudan is presented as that more than one third of the labor force of both sexes, 33.2 percent of the males and more than half of the females in 2008 worked in the major category of occupation which includes skilled agricultural and animal husbandry workers, farmers, farm managers and supervisors, skilled forestry workers, fishermen, hunters and related workers [43]. In the urban areas, the predominant occupation which attracted 19.8 percent of the urban labor force was the elementary occupation whereas agriculture, animal husbandry and forestry workers, fishermen and hunters was the major occupational group in rural areas with over 50 percent of the labor force [43].

The major occupational group in which more labor force engaged in Northern Sudan in 2008 was elementary occupations followed by skilled agricultural and fishery workers [44].

In 2008, the predominant occupation for urban males was craft and related trades workers and for urban females were services workers, shop and market sales, while the skilled agriculture and fishery were the main occupation for both males and females in rural areas.

As shown in Table 1.2 during the inter-censual period (1993-2008) the proportion of labour force in the major occupational group of skilled agricultural, animal husbandry and fishery workers decreased in Northern Sudan from 52.9 percent to 27.4 percent [44, 45].

Table 1.2: Distribution of skilled agricultural and fishery workers labor force Sudan, 1993 and 2008

Year	Total			Urban			Rural			Nomad		
	Total	Male	Female	Total	Male	Female	Total	Male	Female	Total	Male	Female
1993	52.92	45.08	73.54	9.68	9.35	11.56	71.57	64.71	85.48	-	-	-
2008	27.4	23.7	41.5	4.4	4.6	3.8	44.2	38.7	62.1	13.9	12.2	24.0

1.3 Problem Statement

The problem of generating predictions of meteorological events is more complex than that of generating predictions of planetary orbits [1]. This is because the atmosphere is unstable and the systems responsible for the events are the culmination of the instabilities and involve nonlinear interaction between different spatial scales from kilometers to hundreds of kilometers. The chaotic nature of the atmosphere limits the validity of deterministic forecasts [3]. The increasing economic cost of adverse weather events provides a strong reason to generate more accurate and updated weather forecasts [46].

Weather forecasting (particularly rainfall prediction) is one of the most imperatives, important and demanding operational tasks and challenge made by meteorological services around the world. It is a complicated procedure that includes numerous specialized fields of knowledge. The task is complicated because in the field of meteorology all decisions are to be taken with a degree of uncertainty, because the chaotic nature of the atmosphere limits the validity of deterministic forecasts [3].

Long term Rainfall prediction [4] is very important for countries whose economy depends mainly on agriculture, like many of the third World countries. Also it widely used in the energy industry and for efficient resource planning and management [2] including agriculture, famine and disease control, rainwater catchment and ground water management.

1.4 Research Objectives

The main objective of this research is to develop rain prediction models by using computational intelligence methodologies and machine learning techniques for different stations in Sudan and compares them to determine which one is most accurate and has highest performance. The other objectives are as follow:

- Discover the influencing variables that affect rainfall prediction.
- Discover the patterns of weather in Sudan.
- To show that computational intelligence methodologies and machine learning techniques can be used as aneffective approaches to construct long-term predictive models for rainfall forecasting from meteorological data.
- Create heat maps based on predictions for different regions of Sudan.

1.5 Research Questions

The main question behind this research is how to develop rain prediction models by using computational intelligence and machine learning approaches for different stations in Sudan? And then find the more accurate model. There are also other important questions behind this research such as:

Are there any patterns of weather in some regions of Sudan?

What are the influencing variables? In other word which variables (meteorological such as temperature, wind speed and direction and humidity) can be used to most effectively predict rain fall?

What are the different models that would be work?

1.6 Organization of the Thesis

Chapter 2 presents the literature review of the research. It presents an overview of using the various computational intelligence tools such as (Artificial Neural networks (ANNs), Fuzzy Expert Systems, Evolutionary Algorithms, Machine learning and data mining approaches, Ensemble and Meta approaches and Adaptive Neuro-Fuzzy Inference System (ANFIS) technique) in weather forecasting, describing the main contributions on this field and providing taxonomy of the existing proposals according to the type of tools used. Also the chapter provides review analysis; it focused on the capabilities and limitations of several techniques that used in the prediction of several weather phenomena such as rainfall, temperature, wind, flood, tidal level etc. Finally, geographical information system and producing heat maps are discussed in much detail. And up-to-date literature review for many disciplines is displayed.

Chapter 3 describes the research methodology. Some issues such as attribute selection process, the methodology for base and Meta algorithms, the proposed framework of ensemble classifiers, the proposed Structure of ANFIS and The methodology to produce rainfall maps are explained in this chapter, which will direct the aim of the research to achieve all the objectives. The materials used in this research such as datasets and evaluation measures are also introduced in this chapter.

Chapter 4 presents the results of attribute selection process to determine the most influencing variables that affect on the long term rainfall prediction in Sudan. Then we compare between the experimental results of the 10 base algorithms. Finally we compare between the experimental results of the individual Meta algorithms. Comparison processes designed to determine the best models to predict the rainfall among both basic and individual Meta models.

Chapter 5 begins with the results of the proposed ensemble models of Meta Vote method combining with various base classifiers. Then it displays comparisons between the experimental results of several ANFIS models using different types of membership functions, different optimization methods and different dataset ratios for training and testing. Also it shows comparison between the proposed ensemble Vote+3 algorithms and the best ANFIS model. Finally it appears the results of the performance measure criteria of our proposed ensemble Vote+3 algorithms and ANFIS comparing with other related models form literature.

Chapter 6 presents the spatial analysis of rainfall in Sudan for the interval 2000-2012 in three levels (towns, states and regions). It shows the rainfall maps which produced as results of the spatial analysis in different levels. Also it discusses the rainfall maps in details by discussing each related set of maps separately to extract general patterns that control rain phenomenon and its distribution over different geographical areas in Sudan.

Finally, Chapter 7 concludes the thesis that mentions general conclusions of the achieved results, major contributions and future work of this research.

1.7 Summary

In this Chapter, the introduction of this research is described. Firstly, the background of the research is presented. In this Section, the overview of weather forecasting, rainfall prediction, computational intelligence, machine learning, data mining, Meta-learning, ensemble methods, ANFIS technology and heat maps are explained, followed by study area and problem of long term rainfall prediction are then described. Based on the given issues, the objectives and purposes of this research are stated. In addition, research questions are provided accordingly. Finally, the organization of this thesis is presented.

2. Literature Review

2.1 Introduction

Weather forecasting is the application of science and technology to predict the state of the atmosphere for a future time at a given location [3]. Human kind has attempted to predict the weather since ancient times. Many researchers attempt to predict weather using numerical methods [5], statistical methods [1], machine learning techniques [6], hybrid intelligence systems [47] and ensemble strategies [48].

This chapter reviews the existing literature on weather forecasting particularly rainfall prediction. First the chapter briefly commences with a background review of weather forecast.

Second, the Chapter gives an overview of taxonomy based on computational intelligence methodology used such as artificial neural networks, fuzzy expert systems, Evolutionary Algorithms, machine leaning and data mining approaches, ensemble and Meta approaches and Adaptive Neuro-Fuzzy Inference System technique. This is gives a big picture about the modern research trends.

Third, the Chapter provides an up-to-date literature review of remote sensing, geographical information system and producing heat maps to many disciplines such as risk assessment, weather and climate, agriculture, Health and Disease, and environmental studies.

Fourth, the Chapter represents a review analysis, compares between different techniques that used to forecast weather. The comparison highlights the advantages and limitations of each technique. This is essential to reveal the core challenges in weather forecasting. Moreover the concepts involved in these techniques are essential to the construction of the proposed methodologies in this research.

2.2 Background Review

Soft computing techniques and machine learning approaches have been widely used in several applications of weather forecasting, for example they have been applied in rain fall prediction, temperature forecasting, rainfall runoff modeling, flood forecasting and wind forecasting. The results proved that they are better than conventional approaches.

In the nineties, Lee et al. [49] proved that radial basis function (RBF) networks produced good prediction and better than the linear models which produced poor prediction for Rainfall Prediction.

Tang, et al. [50] applied various artificial neural network (ANN) models for prediction and analysis in meteorology and oceanography data and they have found ANN technique is very useful. In the study of rainfall runoff modeling, Dawson and Wilby [51] illustrated the ability of ANN to cope with missing data and to "learn" from the event currently being forecast in real time makes it better choice than conventional lumped or semi distributed flood forecasting models.

In a similar research, Trigo and Palutikof [52] used ANN for simulation of daily temperature for climate change over Portugal. And they have compared the performances of linear models and ANN model using a set of rigorous validation techniques. Finally, they concluded that the non-linear ANN model is more efficient than the linear models. In a comparative study of short-term rainfall prediction models for real time flood forecasting, Toth et al. [53] founded that the time series analysis technique based on ANN provides significant improvement in the flood forecasting accuracy in comparison to the use of simple rainfall prediction approaches.

Luk et al. [54] implemented and compared three types of ANNs suitable for rainfall prediction i.e. multilayer feed forward neural network, Elman partial recurrent neural network and time delay neural network. During a study of radial basis function neural network (RBFNN) for rainfall runoff model, Chang et al. [55] concluded that RBFNN is a good technique for a rainfall runoff model for three hours ahead floods forecasting. Michaelides et al. [56] proved that ANN is a suitable technique for the study of the medium and long-term climatic variability. The ANN models trained were capable of detecting even minor features and discrimination between various classes. In comparative study Maqsood et al. [57] illustrated that Hopfield Model (HFM) for weather forecasting in southern Saskatchewan, Canada is relatively less accurate and RBFN is relatively more reliable for the weather forecasting problems and in comparison the ensembles of neural networks produced the most accurate forecast. Cannon and Whitfield [58] introduced in their climate change studies the bagging (or bootstrap aggregation) method as an ensemble neural network (ENN) approach and showed the suitability of ENNs for downscaling techniques. Combining outputs of several member models can significantly improve generalization performance, because the generalization error of the final predictive model is controlled.

Jeong and Kim [59] developed new rainfall-runoff models that can be used for ensemble stream flow prediction. The new models used two types of artificial neural networks, i.e. single neural network (SNN) and ensemble neural network (ENN). Both ANN models used the early stopping method to optimize generalization performance during training. The bagging method was used in that study for the ENN to control the generalization error better than the SNN. The ANN models were applied to make 1-month ahead probabilistic forecasts for inflows to the Daecheong multipurpose dam in Korea. The calibrated ANN models were compared with each other first. The results illustrated that the ENN is less sensitive to the input variable selection and the number of hidden nodes than the SNN is, and the ENN, in general, produced smaller RMSEs than the corresponding SNN, which implies that the ENN can reduce the generalization error more efficiently than the SNN can. Comparing the SNN and ANN with a rainfall-runoff model TANK, which has been widely used in Korea, with respect to their simulation accuracy, this study found that the new ANN models performed better than TANK for 9 out of 10 test cases. Finally, the study tested TANK and the ENN using some probabilistic forecasting accuracy measures and showed of the test months from 1996 to 2001, the skills of the ENN were better than those of TANK. During the dry season in particular, the ENN improved its ESP performance considerably better than that of TANK. Therefore, the study concluded that an ENN should be substituted for the existing rainfall-runoff model, TANK, for the ESP probabilistic forecasting system for the Daecheong dam inflows in Korea.

Somvanshi et al. [60] illustrated that ANN model can be used as a suitable forecasting tool to predict the rainfall, which out performs the ARIMA (Autoregressive Integrated Moving Average) model. Reddy and Maity [61] implemented artificial intelligence techniques for forecasting regional rainfall and they found that this technique shows reasonably good performance for monthly and seasonal rainfall forecasting. Bustami et al. [62] used ANN for rainfall and water level prediction and the empirical results illustrate that ANN is an effective method in forecasting both missing precipitation and water level data. Hayati and Mohebi [63] used ANN for short-term temperature forecasting (STTF) and founded that MLP network has the minimum forecasting error and can be considered as a good method to model the STTF systems.

In a comparative study between Artificial Intelligence and Artificial Neural Network for rainfall runoff modeling, Aytek et al. [64] illustrated that genetic programming (GP) formulation performs quite well compared to results obtained by ANNs and is quite practical for use. It is concluded from the results that GP can be proposed as an alternative to ANN models. Hocaoglu et al. [65] developed adaptive neuro-fuzzy inference system for missing wind data forecasting. In a Case Study on Jarahi Watershed, Solaimani, [66] studied Rainfall-runoff Prediction located in a semiarid region of Iran Based on Artificial Neural Network and found that Artificial Neural Network method is more appropriate and efficient to predict the river runoff than the classical regression model. Shamseldin [67] examined the effectiveness of rainfall-runoff modeling with ANNs by comparing their results with the Simple Linear Model (SLM), the seasonally based Linear Perturbation Model (LPM) and the Nearest Neighbor Linear Perturbation Model (NNLPM) and concluded that ANNs could provide more accurate discharge forecasts than some of the traditional models.

Nayak et al. [68] offered an application of an adaptive neuro-fuzzy inference system (ANFIS) to hydrologic time series modeling, and it was observed that the ANFIS model preserves the potential of the ANN approach fully, and eases the model building process.

Exploring the new concept, soft computing models based on ANNs and an Evolving Fuzzy Neural Network (EFuNN) for predicting the rainfall time series, Abraham and Philip [69] analyzed 87 years of rainfall data in Kerala state, the southern part of Indian Peninsula. Authors attempted to train 5 soft computing based prediction models with 40 years of rainfall data. Also in the same context Maqsood et al. [70] applied a soft computing model based on a RBFN for 24-h weather forecasting of southern Saskatchewan, Canada. The model is trained and tested using hourly weather data of temperature, wind speed and relative humidity in 2001. The performance of the RBFN is compared with those of multi-layered perceptron (MLP) network, Elman recurrent neural network (ERNN) and Hopfield model (HFM) to examine their applicability for weather analysis. Reliabilities of the models are then evaluated by a number of statistical measures. The results indicate that the RBFN produces the most accurate forecasts compared to the MLP, ERNN and HFM.

Luenam et al. [71] presented a Neuro-Fuzzy approach for daily rainfall prediction, and their experimental results show that overall classification accuracy of the neuro-fuzzy classifier is satisfactory. Vamsidhar et al. [72] used the back propagation neural network model for predicting the rainfall, basis on humidity, dew point and pressure in India. In the training phase, authors obtained 99.79% of accuracy and 94.28% in the testing phase. From these results they have concluded that rainfall can be predicted in the future using the same method. Patil and Ghatol [73] used various ANN models such as radial basis functions and multilayer perceptron with Levenberg Marquardt and momentum learning rules for predicting rainfall using local parameters and they found the models fit for the same task.

Joshi and Patel [74] studied Rainfall-Runoff modeling using ANN, in the that study they have compared three neural network techniques, Feed Forward Back Propagation (FFBP), Radial Basis Function (RBF) and Generalized Regression Neural Network (GRNN) and they have seen that GRNN flow estimation performances were close to those of the FFBP, RBF and MLR. The theoretical basis of the RBF approach lies in the field of interpolation of multivariate functions. The solution of exact interpolating RBF mapping passes through every data point. Different number of hidden layer neurons and spread constants were tried in study.

As we have observed that many of the researchers have used ANN models and soft computing models for forecasting Rainfall, Temperature, Wind and Flood etc., El-Shafie et al. [75] studied and compared Dynamic versus Static neural network models for rainfall forecasting, they have developed soft computing models using Multi-Layer Perceptron Neural Networks (MLPNN), RBFNN and Adaptive Neuro-Fuzzy Inference Systems (ANFIS), finally they concluded that the dynamic neural network namely IDNN could be suitable for modeling the temporal dimension of the rainfall pattern, thus, provides better forecasting accuracy. Sawaitul et al. [76] presented an approach for classification and prediction of future weather using back propagation algorithm, and discussed different models which were used in the past for weather forecasting, finally the study concludes that the new technology of wireless medium can be used for weather forecasting process.

Nelson et al. [77] discussed the issue of whether ANNs can learn seasonal patterns in a time series. They trained networks with both de-seasonalized and the raw data, and evaluated them using 68 monthly time series from the M-competition. Their results indicate that the ANNs are unable to adequately learn seasonality and that prior de-seasonalization of seasonal time series is beneficial to forecast accuracy. However, Sharda and Patil [78] concluded that the seasonality of the time series does not affect the performance of ANNs and ANNs are able implicitly to incorporate seasonality.

Several empirical studies found that ANNs seem to be better in forecasting monthly and quarterly time series Kang [79] and Hill et al. [80, 81] than in forecasting yearly data. This may be due to the fact that monthly and quarterly data usually contain more irregularities (seasonality, cyclicity, nonlinearity, noise) than the yearly data, and ANNs are good at detecting the underlying pattern masked by noisy factors in a complex system.

2.3 Taxonomy Based on Computational Intelligence Methodology Used

2.3.1 Artificial Neural Networks (ANNs)

Abhishek et al. [82] used Feed Forward Network for predicting the rainfall. In the study, monthly rainfall was used as input data for training and testing model. The authors analyzed 50 years of rainfall data for seasonal monsoon (8 months) in Udupi, Karnataka. The empirical results showed that Multi-layer Algorithm is better than Single-layer algorithm terms of performance and Back Propagation is the best algorithm compared to Layer Recurrent and Cascaded feed Forward back Propagation. They concluded that the results of their study were fairly good and high degree of accuracy was obtained. However they obtained large MSE (0.44) because their input data is not quit enough to train the neural network so Larger the amount of input data, lower is the MSE after training. Authors employed only two of the rainfall predictors (Humidity and the average Wind Speed) and there are other predictors such as temperature, wind direction, sunshine, pressure etc. were not taken in to consideration.

Mohammad [83] designed artificial neural network to predict two weather temperatures (average high and average low) for one month ahead in Baghdad, Iraq. Author used feed-forward neural network with back propagation learning algorithm. For network’s training and testing, the author used meteorological daily data for three years (2007-2010). Empirical results suggested that the ANN model has good performance, and low cost of implementation.

De and Debnath [84] proposed ANN model to forecast the mean monthly surface temperature in the monsoon months (June, July and August) over India. Authors developed multilayer feed forward neural network used back propagation algorithm. Three models are generated for both maximum and minimum temperature each model for one of the monsoon months. Data of these three months for the period 1901-2003 were used. The results obtained from the study showed that the Artificial Neural Network has been found to produce a forecast with small prediction error, also he has established that the third model (for the month of August) is the best predictive model over the other two models with the percentage of prediction error = 0.00995147 below than 5% in both maximum and minimum temperature.

Hayati and Mohebi [63] designed short-term temperature forecasting (STTF) Systems for one day ahead for Kermanshah city, west of Iran. Authors used back propagation as the learning algorithm. As back propagation training algorithms are often too slow for practical problems they tried to accelerate the convergence by used several high performance algorithms and they found scaled conjugate gradient was suitable for that purpose. Authors used MLP to train and test using ten years (1996-2006) meteorological data. For accuracy of prediction, they split data into four seasons and then for each seasons one network is presented. The global set of patterns is divided into two randomly selected groups, the training group, corresponding to 67% of the patterns, and the test group, corresponding to 33% of patterns. Two random days in each season are selected as unseen data, which have not been used in training. MSE is used to measure the performance. Tan-sig is used as activation function at each hidden layers and pure-linear is used at each layer. The empirical results showed that MLP network is best suited for this research, it has a good performance and reasonable prediction accuracy was achieved for this model.

Baboo and Shereef [85] used back propagation neural network for predicting the temperature based on the training set provided to the neural network. The ANN is trained and tested using real training data set. Authors used complete daily one year weather data. Through the implementation of this system, it is illustrated, how an intelligent system can be efficiently integrated with a neural network prediction model to predict the temperature. The results shows that when iteration count goes below 1000 the RMSE is more and when it reaches 5000 the error value is up to 0. There are various parameters like the no. of layers, epochs, no. of neurons at each layer etc. and ANN is trained with 200 data and tested for unseen data the result varies with 2.16% errors. The research illustrates a Min RMSE of 0.0079 and Max error of 1.2916.

Devi et al. [86] studied how neural networks are useful in forecasting the weather. Three-layered neural network was designed and trained with the existing dataset and obtained a relationship between the existing non-linear parameters of weather. So many parameters are taken and their relationships are taken into consideration those factors for the temperature forecasting. Like temperature, humidity, dew point, visibility, atmospheric pressure, sea level, wind speed, wind direction. The data is normalized using min-max normalization to scale the dataset into the range of 0 to 1. Authors compared the performance between feed forward network and Radial basis function network to check which is better for the temperature forecasting. The Results indicated that Propagation feed forward network had the best performance and taken for further development for prediction of temperature.

Mayilvaganan and Naidu [87] tried to forecast groundwater level of a watershed using ANN and Fuzzy Logic. They have developed three-layer feed-forward ANN using the sigmoid function and the back propagation algorithm. Authors concluded that ANNs perform significantly better than Fuzzy Logic.

El-Shafie et al. [88] attempted to use neural network and regression technique for rainfall-runoff prediction and the empirical results illustrated that the feed forward ANN can describe the behavior of rainfall-runoff relation more accurately than the classical regression model.

Lekkas et al. [89] developed a multilayer back propagation network and found that BPN does not always find the correct weight and biases fort the optimum solution, whereas their results supported the hypothesis that ANNs can produce qualitative forecast.

Luk et.al [54] developed three alternative types of ANNs, namely multilayer feed forward neural network (MLFN), Elman partial recurrent neural networks, and time delay neural network (TDNN) and the models provided reasonable predictions of the rainfall depth one time-step in advance.

Geetha and Samuel [90] predicted Rainfall in Chennai, India using back propagation neural network model, by their research the mean monthly rainfall is predicted and the results illustrated that the model can perform well both in training and independent periods.

Naik and Pathan [91] proposed new method of weather forecasting using Feed forward ANN with Levenberg Back Propagation Algorithm for training. The results showed that FFNN is appropriate for weather forecasting.

Tiron and Gosav [92] proposed a feed forward neural network approach for short-term prediction of the rainfall field from radar data from the province of Moldova, Romania. The reflectivity data sets extend over July 2008. The ANN system with reflectivity values as input variables was trained to predict the rain rate on the ground. The ANN network was trained with the learning algorithm based on the back-propagation of errors. The results of the study indicated that the use of artificial neural network as a rainfall forecasting system is feasible and efficient.

Dombayc and Golcu [93] tried to predict daily mean ambient temperatures by use of an ANN model in Denizli, South-Western Turkey. They used the meteorological data of the years 2003- 2005 and 2006 as the training and testing data respectively. They analyzed different ANN networks and selected a feed-forward back propagation algorithms consists of 3 inputs, 6 hidden neurons and 1 output.

Afzali et al. [94] developed an artificial neural network for ambient air temperature prediction in Kerman city located in the south east of Iran. The mean, minimum and maximum ambient air temperature during the year 1961-2004 was used as the input parameter in Feed Forward Network (FNN) and Elman Network. The output of the models is composed of one day and one-month ahead air temperature prediction. The experiments illustrated that ANN approach is a desirable model in ambient air temperature prediction, while the results from Elman network are more precise than FNN network.

Perea et al. [95] analyzed an energy consumption predictor for greenhouses using a multi-layer perceptron (MLP) artificial neural network (ANN) trained by means of the Levenbergh-Marquardt back propagation algorithm. The predictor uses cascade architecture, where the outputs of a temperature and relative humidity model are used as inputs for the predictor, in addition to time and energy consumption. The performance of the predictor was evaluated using real data obtained from a greenhouse located at the Queretaro State University, Mexico. This study shows the advantages of the ANN with a focus through analysis of variance (ANOVA). Energy consumption values estimated with an ANN were compared with regression-estimated and actual values using ANOVA and mean comparison procedures. Results show that the selected ANN model gave a better estimation of energy consumption with a 95% significant level. The study resents an algorithm based in ANOVA procedures and ANN models to predict energy consumption in greenhouses.

Paras et al. [96] used feed forward artificial neural networks with back propagation for supervised learning using the data recorded at Pantnagar station situated in Tarai region of Uttarakhand state, India since April 1996 to March 1999 and is available as weekly average. The trained ANN was used to predict the future weather conditions. The results are very encouraging and it is found that the feature based forecasting model can make predictions with high degree of accuracy. The model can be suitably adapted for making forecasts over larger geographical areas.

Chang et al. [55] tried RBFN to develop a rainfall–runoff model for three-hour-ahead flood forecasting. They have used dataset of the Lanyoung River collected during typhoons for training, testing and validating the network. After the study they found that that the RBF NN can be considered as an appropriate technique for predicting flood flow.

Maqsood et al. [57] projected ensemble model performance is contrasted with multi-layered perceptron network (MLPN), Elman recurrent neural network (ERNN), RBFN, Hopfield model (HFM) predictive models and regression techniques. They have used dataset of temperature, wind speed and relative humidity to train and test the different models. With each model, 24-h-ahead forecasts are prepared for the winter, spring, summer and fall seasons. Furthermore, the performance and reliability of the seven models are then evaluated by a number of statistical measures. Among the direct approaches employed, empirical results indicate that HFM is relatively less accurate and RBFN is relatively more reliable for the weather forecasting problem. In comparison, the ensemble of neural networks and RBFN produced the most accurate forecasts.

In a comparative study, Maqsood et al. [70] applied a soft computing model based on a RBFN for 24-h weather forecasting of southern Saskatchewan, Canada. The model is trained and tested using hourly weather data of temperature, wind speed and relative humidity. The performance of the RBFN is compared with those of multi-layered perceptron (MLP) network, Elman recurrent neural network (ERNN) and Hopfield model (HFM) to examine their applicability for weather analysis. Reliabilities of the models are then evaluated by a number of statistical measures. The results indicate that the RBFN produces the most accurate forecasts compared to the MLP, ERNN and HFM.

Santhanam and Subhajini [97] developed two neural network models for weather forecasting, based on various factors obtained from meteorological experts such as temperature, air pressure, humidity, cloudiness, precipitation, wind direction and wind speed. Authors evaluated the performance of Radial Basis Function (RBF) with Back Propagation (BPN) neural network. The back propagation neural network and radial basis function neural network were used to test the performance in order to investigate effective forecasting technique. The Results showed that the prediction accuracy of RBF was 88.49% while the prediction accuracy of BPN was 81.99. The results indicated that proposed radial basis function neural network is better than back propagation neural network.

Abdul–Kader [98] evaluated the use of two different artificial neural network models namely, RBF and back propagation neural networks to forecast temperature in some Egyptian towns. The gained simulated results showed that the popular feed-forward neural network, which trained by differential evolution algorithm (DE) is the most accurate model to use as a temperature predicator. Especially in the uniform temperature distribution (minimum or maximum temperature) which can be considered the most suitable technique for temperature forecasting.

Maqsood et al. [99] presented a comparative study of different neural network models for forecasting the weather of Vancouver, British Columbia, Canada. For developing the models, they used one year's data comprising of daily maximum and minimum temperature, and wind-speed. They used Multi-Layered Perceptron (MLP) and an Elman Recurrent Neural Network (ERNN), which were trained using the one-step-secant and Levenberg- Marquardt algorithms. To ensure the effectiveness of neuro-computing techniques, they also tested the different connectionist models using a different training and test data set. Their goal is to develop an accurate and reliable predictive model for weather analysis. Radial Basis Function Network (RBFN) exhibits a good universal approximation capability and high learning convergence rate of weights in the hidden and output layers. Experimental results obtained have shown RBFN produced the most accurate forecast model as compared to ERNN and MLP networks.

Awchi [100] investigated the potential of Radial Basis Function (RBF) neural networks for the prediction of reference Evapotranspiration (ETo). The study utilized daily climatic data of temperature, relative humidity, sunshine hours, wind speed, and rainfall for five years collected from Mosul meteorological station, north of Iraq. Thirteen RBF networks each using varied input combination of climatic variables have been trained and tested. The network output is compared with estimated daily Penman-Monteith ETo values. To evaluate the performance of RBF networks, the same networks in the studied cases were re-trained using the well-known feed forward-back propagation (FF-BP) networks. In addition, the effect of including a time index within the inputs of considered networks is investigated. The study showed that the RBF network is seen to emulate the FF-BP in its performance and can be effectively used for ETo prediction. Besides, it is much easier to built and much faster to train. It is noticed that the networks' output are very highly correlated to estimated ETo, especially when concerning all the climatic parameters. The study results reveal that adding a time index to the inputs highly improves the ETo prediction of the studied cases.

Lin and Chen [101] used radial basis function network (RBFN) to construct a rainfall-runoff model for the Fei–Tsui Reservoir Watershed in northern Taiwan for predicting real time stream flows. The fully supervised learning algorithm has been presented for the parametric estimation of the network. The results showed that the RBFN could be successfully applied to build the relationship between rainfall and runoff. Moreover, the proposed network trained using the fully supervised learning algorithm provides better training and testing accuracy than the network trained using the hybrid-learning algorithm does. The proposed network also gives better forecasts.

Chow and Cho [102] have developed recurrent Sigma-Pi neural network for rainfall forecasting system in Hong Kong. The results were very promising, and the neural-based rainfall forecasting system is capable of providing a rainstorm-warning signal one hour ahead. They have concluded that the neural network based now casting system is capable of providing a reliable rainfall now casting.

In comparative study of Jordan and Elman networks for rainfall-runoff modeling for the upper area of Wardha River in India, Deshmukh and Ghatol [103] developed the models by processing online data over time using recurrent connections. The prediction results of the Jordan network indicated a satisfactory performance in the three hours ahead of time prediction. The conclusions also indicated that the Jordan network is more versatile than Elman model and can be considered as an alternate and practical tool for predicting short term flood flow.

Gong et al. [104] have tried Elman neural network models for wind power forecasting. The relevant data sequences provided by numerical weather prediction are decomposed into different frequency bands by using the wavelet decomposition for wind power forecasting. The Elman neural networks models are established at different frequency bands respectively, then the output of different networks are combined to get the eventual prediction result. For comparison, Elman neutral network and BP neutral network are used to predict wind power directly. Several error indicators are given to evaluate prediction results of the three methods. The simulation results showed that the Elman neural network can achieve good results and that prediction accuracy can be further improved by using the wavelet decomposition simultaneously.

Meng and Wu [105] proposed a novel hybrid Radial Basis Function Neural Network (RBF–NN) ensemble model is proposed for rainfall forecasting based on Kernel Partial Least Squares Regression (K–PLSR). In the process of ensemble modeling, the first stage the initial data set is divided into different training sets by used Bagging and Boosting technology. In the second stage, these training sets are input to the RBF–NN models of different kernel function, and then various single RBF–NN predictors are produced. Finally, K–PLSR is used for ensemble of the prediction purpose. Authors findings reveal that the K–PLSR ensemble model can be used as an alternative forecasting tool for a Meteorological application in achieving greater forecasting accuracy.

2.3.2 Fuzzy Expert Systems

Özelkan et al. [106] compared the performance of regression analysis and fuzzy logic in studying the relationship between monthly atmospheric circulation patterns and precipitation. Liu and Chandrasekar [107] developed a Fuzzy Logic and Neuro-Fuzzy system for classification of a hydrometeor type based on polarimetric radar measurements where fuzzy logic was used to infer a hydrometeor type, and the neural network-learning algorithm was used for automatic adjustment of the parameters of the fuzzy sets in the fuzzy logic system according to the prior knowledge. Luenam et al. [71] presented a Neuro-Fuzzy approach for daily rainfall prediction, and their experimental results show that overall classification accuracy of the neuro-fuzzy classifier is satisfactory.

Mahabir et al. [108] investigated the applicability of fuzzy logic modeling techniques for forecasting water supply for the Lodge Creek and Middle Creek basins, located in southeastern Alberta, Canada. By applying fuzzy logic, a water supply forecast was created that classified potential runoff into three forecast zones: ‘low’, ‘average’ and ‘high’. Spring runoff forecasts from the fuzzy expert systems were found to be considerably more reliable than the regression models in forecasting the appropriate runoff zone, especially in terms of identifying low or average runoff years. Based on the modeling results in these two basins, it is concluded that fuzzy logic has a promising potential for providing reliable water supply forecasts.

Bae et al. [109] developed Adaptive Neuro-Fuzzy Inference System (ANFIS) to predict the optimal dam inflow. The model used dataset of rainfall, inflow, temperature, relative humidity observation data and monthly weather forecasts. The subtractive clustering algorithm was adopted to enhance the performance of the ANFIS model and the hybrid-learning algorithm was adopted to enhance model performance. The ANFIS model for monthly dam inflow forecasts was tested in cases with and without weather forecasting information. The results demonstrated that a neuro-fuzzy system is appropriate for dam inflow forecasts. The model gave better performances where the various field data were available.

Abraham and Philip [69] attempted to train 5 soft computing based prediction models with 40 years of rainfall data. For performance evaluation, network predicted outputs were compared with the actual rainfall data. Simulation results reveal that soft computing techniques are promising and efficient. They used an artificial neural network using back propagation (variable learning rate), adaptive basis function neural network, neural network using scaled conjugate gradient algorithm and an Evolving Fuzzy Neural Network (EFuNN) for predicting the rainfall time series. The test results given by EFuNN algorithm were the best. Lowest RMSE was obtained using EFuNN (0.090) and it was 0.095, 0.094, 0.092 and 0.093 for BP, BP-VLR and SCG and ABF neural networks respectively. Also they found EFuNN adopts a one-pass (one epoch) training technique, which is highly suitable for online learning. Hence online training can incorporate further knowledge very easily. Compared to pure BP and BP-VLR, ABFNN and SCGA converged much faster. Finally they concluded that EFuNN outperformed neurocomputing techniques with the lowest RMSE test error and performance time.

Aliev et al. [110] proposed, fuzzy recurrent neural network (FRNN) based time series forecasting method for solving forecasting problems, and they found that the performance of the proposed method for forecasting fuzzy time series shows its high efficiency and effectiveness for a wide domain of application areas ranging from weather forecasting to planning in economics and business.

2.3.3 Evolutionary Algorithms

Aytek et al. [64] illustrated that genetic programming (GP) formulation performs quite well compared to results obtained by ANNs and is quite practical for use. It is concluded from the results that GP can be proposed as an alternative to ANN models.

Jiang and Wu [111] investigated the effectiveness of the hybrid Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) evolved neural network for rainfall forecasting and its application to predict the monthly rainfall in a catchment located in a subtropical monsoon climate in Guilin of China. They adopted a hybrid Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) for the automatic design of NN by evolving to the optimal network configuration(s) within an architecture space, namely PSOGA–NN. The PSO is carried out as a main frame of this hybrid algorithm while GA is used as a local search strategy to help PSO jump out of local optima and avoid sinking into the local optimal solution early. The proposed technique is applied over rainfall forecasting to test its generalization capability as well as to make comparative evaluations with the several competing techniques, such as GA–NN, PSO–NN and NN. The experimental results showed that the GAPSO–NN evolves to optimum or near–optimum networks in general and has a superior generalization capability with the lowest prediction error values in rainfall forecasting. Finally they concluded that the predictions using the GAPSO–NN approach can significantly improve the rainfall forecasting accuracy.

Riley and Venayagamoorthy [112] used a recurrent neural network (RNN) for Photovoltaic (PV) system modeling. They used particle swarm optimization (PSO) for modifying the network weights to train the network and minimize the sum of the mean absolute error (MAE). Also they have compared a traditional modeling approach using the Sandia Photovoltaic Array Performance Model to a new method of characterization using a recurrent neural network (RNN). The results showed that modeling and characterizing an existing PV system with a recurrent neural network may provide adequate results for existing PV systems, although in this case, the RNN model did not perform as well as the component-based model. Thus, it seems that in the case where component parameters are known, a traditional PV modeling approach may yield more accurate model results. Also The RNN model correctly learned the relationships between the weather data and performance data.

2.3.4 Machine learning and Data Mining Approaches

Knowledge Discovery in Databases (KDD) is an automatic, exploratory analysis and modeling of large data repositories. KDD is the organized process of identifying valid, novel, useful, and understandable patterns from large and complex data sets. Data Mining (DM) is the core of the KDD process [10], involving the inferring of algorithms that explore the data, develop the model and discover previously unknown patterns. The model is used for understanding phenomena from the data, analysis and prediction.

The scientists have tried to forecast the meteorological characteristics using a large set of methods, some of them more accurate than others. Lately, there has been discovered that data mining, a method developed recently, can be successfully applied in this domain. Data mining is about solving problems by analyzing data already present in databases. Data mining is defined as the process of discovering patterns in data. The process must be automatic or (more usually) semiautomatic. The patterns discovered must be meaningful in that they lead to some advantage, usually an economic advantage. The data is invariably present in substantial quantities [10].

In contrast to standard statistical methods, data mining techniques search for interesting information without demanding a priori hypotheses, the kind of patterns that can be discovered depend upon the data mining tasks employed. There are two types of data mining tasks: descriptive data mining tasks that describe the general properties of the existing data and predictive data mining tasks that attempt to do predictions based on inference on available data. This techniques are often more powerful, flexible, and efficient for exploratory analysis than the statistical.

Data mining is aggregation of many disciplines which include machine learning, statistics, data base technology, information science and visualization. Techniques like neural networks, support vector machines, fuzzy and rough set theory from other disciplines are often used depending upon the data mining approach used.

Meteorological data mining is a form of Data mining concerned with finding hidden patterns inside largely available meteorological data, so that the information retrieved can be transformed into usable knowledge. Useful knowledge can play important role in understanding the climate variability and climate prediction. In turn, this understanding can be used to support many important sectors that are affected by climate like agriculture, vegetation, water resources and tourism.

Data Mining is the search for the relationships and global patterns that exist in large databases but are hidden among vast amount of data, such as the relationship between patient data and their medical diagnosis [11]. This relationship represents valuable knowledge about the database, and the objects in the database, if the hidden database is a faithful mirror of the real world registered by the database. Data Mining refers to using a variety of techniques to identify nuggets of information or decision making knowledge in the database and extracting these in such a way that they can be put to use in areas such as decision support, prediction, forecasting and estimation. The data is often voluminous, but it has low value and no direct use can be made of it. It is the hidden information in the data that is useful [11].

Zhao and Wang [113] developed a neural network technique, support vector regression (SVR), to monthly rainfall forecasting. Authors used particle swarm optimization (PSO) algorithms, which searches for SVR's optimal parameters, and then adopts the optimal parameters to construct the SVR models. The monthly rainfalls in Guangxi of China during 1985–2001 were employed as the data set. Authors compared the new neural network technique with back–propagation neural networks (BPNN) and the autoregressive integrated moving average (ARIMA) model. The experimental results demonstrated that SVR outperformed the BPNN and ARIMA models based on the normalized mean square error (NMSE) and mean absolute percentage error (MAPE).

Lee et al. [49] studied the predicting the daily rainfall at 367 locations based on the daily rainfall at nearby 100 locations in Switzerland. The whole area is divided into four sub-areas and each is modeled with a different way. Predictions in two larger areas were prepared by RBF networks based on the location information only. Predictions in two smaller were made using a simple linear regression model based on the elevation information only. They have concluded that the RBF networks produced good prediction while the linear models poor prediction.

Shamseldin [67] examined the effectiveness of rainfall-runoff modeling with ANNs by comparing their results with the Simple Linear Model (SLM), the seasonally based Linear Perturbation Model (LPM) and the Nearest Neighbor Linear Perturbation Model (NNLPM) and concluded that ANNs could provide more accurate discharge forecasts than some of the traditional models.

Isa et al. [114] tried to predict daily weather conditions based on various measured parameters gained from the Photovoltaic (PV) system. In that work, Multiple Multilayer Perceptron (MMLP) network with majority voting technique was used and trained using Levenberg Marquardt (LM) algorithm. Different techniques of voting are used such as majority rules, decision-making, consensus democracy, consensus government and supermajority. The way of the voting technique is different depending on the problem involved. Majority voting technique was applied in the study so that the performance of MMLP can be approved as compared to single MLP network. The proposed work has been used to classify four weather conditions; rain, cloudy, dry day and storm. The system can be used to represent a warning system for likely adverse conditions. Experimental results demonstrate that the applied technique gives better performance than the conventional ANN concept of choosing an MLP with least number of hidden neurons.

On the other hand Lunagariya et al. [115] made an effort to verify the weather forecast from NCMRWF. Analysis was carried out weekly, seasonal as well as yearly basis using various numerical verification techniques like ratio score, usability analysis and correlation approach during July 2006 and September 2008-2009. The forecasts were found within usability range for some parameters but for other parameter improvement is still possible.

Dawid [116] explained that the purpose of statistical inference is to make sequential probability forecast for future observation rather than to express information about parameters. Therefore, there is a need of an approach, which is better than the statistical inference method. However, Glahn and Lowry [117] proved that Model Output Statistics (MOS) technique is an objective weather forecasting technique, which consists of determining a statistical relationship between a predict and variable forecast by a numerical model at some projection time. It is the determination of the “weather related” statistics of a numerical model. They applied this technique, together with screening regression to the predication of surface wind, probability of precipitation, maximum temperature, cloud amount and conditional probability of frozen precipitation. The obtained results are compared with the national weather system over Teletype and facsimile. Results illustrate that MOS is a useful technique in objective weather forecasting. Therefore, in the proposed research statistical regression as multidimensional response surface tool is applied to forecast local monsoonal precipitation.

Su et al. [118] proposed an approach that can incorporate both types of prediction (global prediction and local prediction.) to increase prediction accuracy. The proposed Markov–Fourier gray model (MFGM) prediction approach uses a gray model to roughly predict the next datum from a set of the most recent data and a Fourier series to fit the residual errors produced by the gray model. Finally, Markov state transition matrices are employed to recode the global information generated also by the residual errors. By combining a local predicted value obtained by a Fourier series and a global estimated error obtained by the Markov forecasting method, the approach can predict the future weather more accurately.

Kannan, et al. [119] used some of the data mining functionalities such as: classification, clustering and regression, they classified what is the reason for rainfall fall in the ground level.

They computed values for rainfall using five years input data by using Pearson correlation coefficient and predicted for future years rainfall in ground level by multiple linear regression. However their predicted values are lie below the computed values. So, it did not show an accurate but show an approximate value.

Rupa and Jain [120] employed principal component analysis (PCA) and ANN on monthly rainfall data, and the experimental Results showed that PCA has some more benefits over ANN in analyzing climatic time series such as rainfall, particularly with regards to the interpretability of the extracted signals.

Kusiak et al. [121] studied the effect of rainfall on local water quantity and quality in a watershed basin at Oxford, Iowa, based on radar reflectivity and tipping bucket (TB) data. They used five data-mining algorithms, neural network, random forest, classification and regression tree, support vector machine, and k-nearest neighbor to build rainfall prediction modes. Three Models are selected for all future time horizons. Model I is the baseline model constructed from radar data covering oxford. Model II predicts rainfall from radar and TB data collected at south Amana (16 km west oxford) and Iowa City (25 km east of oxford). Among 5 algorithms MLP neural network has the best performance in comparison to other algorithms. Their computation results indicated that the three models had a similar performance in predicting rainfall at current time, and model II was more than the other models in predicting rainfall at future time horizons. Different lags like t+15, t+30, t+45, t+60, t+75, t+90, t+105, t+120 were considered. The longest acceptable prediction horizon is 120 min.

Sivaramakrishnan and Meganathan [122] attempted to predict spot rainfall for an interior station Trichirappalli (10º48! N/78º41! E) of south India by using association rule mining, The data is filtered using discretization approach based on the best fit ranges and then association mining is performed on dataset using Predictive Apriori algorithm and then the data need be validated using K* classifier approach. The results showed that the overall classification accuracy for occurrence and non-occurrence of the rainfall on wet and dry days using the data mining technique is satisfactory.

Petre [123] tried to predict the average temperature for a future month and developed Classification and Regression Trees (CART) for weather data collection registered over Hong Kong between 2002 and 2005.

Ingsrisawang et al. [124] proposed machine-learning approaches for short-term rain forecasting system. Decision Tree, Artificial Neural Network (ANN), and Support Vector Machine (SVM) were applied to develop classification and prediction models for rainfall forecasts in the northeastern part of Thailand. They used datasets collected during 2004-2006. There are three main parts in their work. Firstly, a decision tree induction algorithm (C4.5) was used to classify the rain status into either rain or no-rain. The overall accuracy of classification tree achieves 94.41% with the five-fold cross validation. The C4.5 algorithm was also used to classify the rain amount into three classes as no-rain (0-0.1 mm.), few-rain (0.1- 10 mm.), and moderate-rain (>10 mm.) and the overall accuracy of classification tree achieves 62.57%. Secondly, an ANN was applied to predict the rainfall amount and the root mean square error (RMSE) were used to measure the training and testing errors of the ANN. It is found that the ANN yields a lower RMSE at 0.171 for daily rainfall estimates, when compared to next-day and next-2-day estimation. Thirdly, the ANN and SVM techniques were also used to classify the rain amount into three classes as no-rain, few-rain, and moderate-rain as above. The results achieved in 68.15% and 69.10% of overall accuracy of same-day prediction for the ANN and SVM models, respectively. The obtained results illustrated the comparison of the predictive power of different methods for rainfall estimation.

Khandelwal and Davey [125] applied data mining techniques regression analysis on the rainfall dataset of Jaipur city. They specifically used multiple regression analysis to predict rainfall of a year by using different 4 climatic factors temperature, humidity, pressure and sea level. For selecting those factors they applied correlation analysis.

Dutta and Tahbilder [126] used Multiple Linear Regression on six years Meteorological data (2007-2012) for Guwahati, Assam, India. The model considered maximum temperature, minimum temperature, wind speed, Mean sea level as Predictors. Experiments results showed that the prediction model based on multiple linear regression achieved 63% accuracy in variation of rainfall.

Wang and Sheng [127] proposed Generalized Regression neural network model for annual rainfall in Zhengzhou. The results of GRNN have more advantage in fitting and prediction compared with BP neural network and stepwise regression analysis methods. The simulation results of GRNN for annual rainfall are better than that of BP neural network. Accuracy predicted using GRNN is better than BP. The stepwise regression method is inferior to both BP and GRNN in accuracy of simulation and prediction results. GRNN network structure is simple and stable.

Olaiya and deyemo [128] investigated the use of data mining techniques in forecasting maximum temperature, rainfall, evaporation and wind speed. That was carried out using Artificial Neural Network and Decision Tree algorithms and meteorological data collected between 2000 and 2009 from the city of Ibadan, Nigeria. A data model for the meteorological data was developed and this was used to train the classifier algorithms. The performances of these algorithms were compared using standard performance metrics, and the algorithm, which gave the best results were used to generate classification rules for the mean weather variables. A predictive Neural Network model was also developed for the weather prediction program and the results compared with actual weather data for the predicted periods. The results show that given enough case data, Data Mining techniques can be used for weather forecasting and climate change studies.

Sethi and Garg [129] used multiple linear regression (MLR) technique for the early rainfall prediction. The model is implemented with the use 30 years (1973-2002) datasets of the climate data such as rainfall precipitation, vapor pressure, average temperature, and cloud cover over Udaipur City, Rajasthan, India. The model forecasts monthly rainfall amount of July (in mm). The experimental results proved that there is a close agreement between the predicted and actual rainfall amount prediction of rainfall.

Singhratina et al. [130] described the development of a statistical forecasting method for SMR over Thailand using multiple linear regression and local polynomial-based nonparametric approaches. SST, sea level pressure (SLP), wind speed, EiNino Southern Oscillation Index (ENSO), and IOD were chosen as predictors. The experiments indicated that the correlation between observed and forecast rainfall was 0.6.

Zaw and Naing [131] presented the MPR technique, an effective way to describe complex nonlinear I/P-O/P relationship for prediction of rainfall and then compared the MPR and MLR technique based on the accuracy.

Kajornrit et al. [132] proposed fuzzy inference system for monthly rainfall prediction in the northeast region of Thailand. The predicted performance of the proposed model was compared to be conventional Box-Jenkins and artificial neural networks model. Accordingly, the experimental results show the modular FIS is good alternative method to predict accurately. The predicted mechanism can be interpreted through fuzzy rules. Auto-regression, Seasonal auto regressive integrated moving average and ANN modular FIS provide better results. The experimental results provide both accurate results and human-understandable prediction mechanism.

Nikam and Meshram [133] proposed Bayesian model for rainfall prediction. Since Bayesian prediction model can easily learn new classes. The accuracy also grows with the increase of learning data. Bayesian model issue is that if the predictor category is not present in the training data, the model assumes that a new record with that category has zero probability. According to this paper, Bayesian model for rainfall prediction provides good accuracy. The features used station level pressure, mean sea level pressure, temperature, relative humidity, vapor pressure, wind speed and rainfall. Some of features is being ignored which are less relevant features in the dataset for model computation.

Ji et al. [134] proposed CART and C4.5 to predict rainfall. To correctly perform rainfall prediction, the chance of rain is first determined. Then, hourly rainfall prediction is performed only if there is any chance of rain. 13 variables are considered, they are wind direction, wind speed, wind gust, outdoor humidity, outdoor temperature, evaporation, solar radiation, wind chill, dew point, pressure altitude, cloud base, air density, vapor pressure. The proposed model would be useful for predicting the chance of rain and estimating hourly rainfall in any geographical regions time-efficiently. CART predicted accurately 99.2% and C4.5 predicted accurately 99.3%. And the average prediction accuracy of estimating hourly rainfall with CART and C4.5 are 92.8% and 93.4% correspondingly. CART and C4.5 both have high accuracy and are efficient algorithms.

2.3.5 Ensemble and Meta Approaches

Meta-learning is a technique that seeks to compute higher-level classifiers (or classification models), called meta-classifiers, that integrate in some principled fashion multiple classifiers computed separately over different databases. Meta-learning means learning from the classifiers produced by the inducers and from the classifications of these classifiers on training data.

Meta-learning improves efficiency by executing in parallel the base-learning processes (each implemented as a distinct serial program) on (possibly disjoint) subsets of the training data set (a data reduction technique). This approach has the advantage, first, of using the same serial code without the time-consuming process of parallelizing it, and second, of learning from small subsets of data that fit in main memory.

Meta-learning improves predictive performance by combining different learning systems each having different inductive bias (e.g. representation, search heuristics, search space) [13]. By combining separately learned concepts, meta-learning is expected to derive a higher level learned model that explains a large database more accurately than any of the individual learners. Furthermore, meta-learning constitutes a scalable machine learning method since it can be generalized to hierarchical multi-level meta-learning.

The idea of ensemble methodology is to build a predictive model by integrating multiple models. It is well-known that ensemble methods can be used for improving prediction performance [14].

Building an ensemble consists of two steps: (1) constructing varied models and (2) combining their estimates. One may generate component models by, for instance, varying case weights, data values, guidance parameters, variable subsets, or partitions of the input space. Combination can be accomplished by voting, but is primarily done through model estimate weights [15].

Diversity is a crucial condition for obtaining accurate ensembles [16-19]. According to [20], diversified classifiers lead to uncorrelated classifications, which in turn improve classification accuracy. However, in the classification context, there is no complete and agreed upon theory to explain why and how diversity between individual models contributes toward overall ensemble accuracy [21].

An important aspect of ensemble methods is to determine how many base classifiers and which classifiers should be included in the final ensemble. Several algorithms, such as bagging, predetermine the ensemble size, by using a controlling parameter such as number of iterations that can be set by the user. Other ensemble algorithms try to determine the best ensemble size while training. When new members are added to the ensemble, we check if the performance of the ensemble has improved. If it is not, the procedure stops and no new base classifier are trained. Usually these algorithms also have a controlling parameter, which bounds the number of base classifiers in the ensemble. An algorithm that decides when a sufficient number of classification trees have been created was proposed by Robert et al. [22].

Ensemble methodology imitates our second nature to seek several opinions before making a crucial decision. The core principle is to weigh several individual pattern classifiers, and combine them in order to reach a classification that is better than the one obtained by each of them separately. Researchers from various disciplines such as pattern recognition, statistics, and machine learning have explored the use of ensemble methods since the late seventies. Given the growing interest in the field, it is not surprising that researchers and practitioners have a wide variety of methods at their disposal.

An ensemble is largely characterized by the diversity generation mechanism and the choice of its combination procedure. While ensemble approaches to classification usually make use of non-linear combination methods like majority voting; regression problems are naturally tackled by linearly weighted ensembles. These types of ensembles have a much clearer framework for explaining the role of diversity than voting methods, in particular the ambiguity decomposition [17].

The ensemble idea in supervised learning has been investigated since the late seventies. Tukey [135] suggested combining two linear regression models. The main progress in the field was achieved during the Nineties. Hansen and Salamon [136] suggested an ensemble of similarly configured neural networks to improve the predictive performance of a single one. At the same time Schapire [137] laid the foundations for the award winning AdaBoost Freund and Schapire [138] algorithm by showing that a strong classifier in the probably approximately correct (PAC) sense can be generated by combining “weak” classifiers (that is, simple classifiers whose classification performance is only slightly better than random classification).

After that, researchers from various disciplines such as statistics and AI considered the use of ensemble methodology; Merler et al. [139] developed the P-AdaBoost algorithm, which is a distributed version of AdaBoost. Instead of updating the “weights” associated with instance in a sequential manner, P-AdaBoost works in two phases. In the first phase, the AdaBoost algorithm runs in its sequential, standard fashion for a limited number of steps. In the second phase the classifiers are trained in parallel using weights that are estimated from the first phase. P-AdaBoost yields approximations to the standard AdaBoost models that can be easily and efficiently distributed over a network of computing nodes.

Zhang and Zhang [140] proposed a new boosting-by-resampling version of Adaboost. In the local Boosting algorithm, a local error is calculated for each training instance, which is then used to update the probability that this instance is chosen for the training set of the next iteration. After each iteration, in AdaBoost, a global error measure is calculated that refers to all instances.

Alhamdoosh and Wang [141] employed the random vector functional link (RVFL) networks as base components, and incorporated with the NCL strategy for building neural network ensembles. The basis functions of the base models are generated randomly and the parameters of the RVFL networks can be determined by solving a linear equation system. An analytical solution is derived for these parameters, where a cost function defined for NCL and the well known least squares method are used. To examine the merits of their proposed algorithm, a comparative study was carried out with nine benchmark datasets. Results indicate that their approach outperforms other ensemble techniques on the testing datasets in terms of both effectiveness and efficiency.

DeWeber and Wagner [142] compared four models with different groups of predictors to determine how well water temperature could be predicted by climatic, landform, and land cover attributes, and used the median prediction from an ensemble of 100 ANNs as their final prediction for each model. The final model included air temperature, landform attributes and forested land cover and predicted mean daily water temperatures with moderate accuracy as determined by root mean squared error (RMSE) at 886 training sites with data from 1980 to 2009 (RMSE = 1.91 _C). Based on validation at 96 sites (RMSE = 1.82) and separately for data from 2010 (RMSE = 1.93), a year with relatively warmer conditions, the model was able to generalize to new stream reaches and years. The most important predictors were mean daily air temperature, prior 7 day mean air temperature, and network catchment area according to sensitivity analyses. Forest land cover at both riparian and catchment extents had relatively weak but clear negative effects. Predicted daily water temperature averaged for the month of July matched expected spatial trends with cooler temperatures in headwaters and at higher elevations and latitudes. Their ANN ensemble is unique in predicting daily temperatures throughout a large region, while other regional efforts have predicted at relatively coarse time steps. The model may prove a useful tool for predicting water temperatures in sampled and un sampled rivers under current conditions and future projections of climate and land use changes, thereby providing information that is valuable to management of river ecosystems and biota such as brook trout.

Li et al. [143] explored the influence of the classification confidence of the base classifiers in ensemble learning and obtain some interesting conclusions. First, they extended the definition of ensemble margin based on the classification confidence of the base classifiers. Then, an optimization objective is designed to compute the weights of the base classifiers by minimizing the margin induced classification loss. Several strategies were tried to utilize the classification confidences and the weights. It is observed that weighted voting based on classification confidence is better than simple voting if all the base classifiers are used. In addition, ensemble pruning can further improve the performance of a weighted voting ensemble. They also have compared the proposed fusion technique with some classical algorithms. The experimental results also show the effectiveness of weighted voting with classification confidence.

Zhang and Suganthan [144] proposed a new method to improve the performance of the Random Forests by increasing the diversity of each tree in the forests and there by improve the overall accuracy. During the training process of each individual tree in the forest, different rotation spaces are concatenated into a higher space at the root node. Then the best split is exhaustively searched within this higher space. The location where the best split lies decides which rotation method to be used for all subsequent nodes. The performance of the proposed method here is evaluated on 42 benchmark data sets from various research fields and compared with the standard Random Forests. The results showed that the proposed method improves the performance of the Random Forests in most cases.

Salih and Abraham [145] proposed a novel ensemble health care decision support for assisting an intelligent health monitoring system, their ensemble method was constructed based of Meta classifier voting combining with three base classifiers J48, Random Forest and Random Tree algorithms. The results obtained from the experiments showed that the proposed Ensemble method achieved better outcomes that are significantly better compared with the outcomes of the other Base and Meta base classifiers.

Li et al. [146] presented a method for improved ensemble learning, by treating the optimization of an ensemble of classifiers as a compressed sensing problem. Ensemble learning methods improve the performance of a learned predictor by integrating a weighted combination of multiple predictive models. Ideally, the number of models needed in the ensemble should be minimized, while optimizing the weights associated with each included model. They solved this problem by treating it as an example of the compressed sensing problem, in which a sparse solution must be reconstructed from an under- determined linear system. Compressed sensing techniques are then employed to find an ensemble, which is both small and effective. The experiments showed that their method gave better accuracy, while being significantly faster than the compared methods

Chen, et al. [147] proposed a unified evolutionary training scheme (UETS) which can either train a generalized feed forward neural network or construct an ANN ensemble. The performance of the UETS was evaluated by applying it to solve the n-bit parity problem and the classification problems on five datasets from the UCI machine-learning repository. By comparing with the previous studies, the experimental results reveal that the neural networks and the ensembles trained by the UETS have very good classification ability for unseen cases.

In the RAndom k-labELsets (RAKEL) algorithm, each member of the ensemble is associated with a small randomly selected subset of k labels. Then, a single label classifier is trained according to each combination of elements in the subset. Rokach et al. [148] adopted a similar approach, however, instead of randomly choosing subsets, they selected the minimum required subsets of k labels that cover all labels and meet additional constraints such as coverage of inter-label correlations. Construction of the cover is achieved by formulating the subset selection as a minimum set covering problem (SCP) and solving it by using approximation algorithms. Every cover needs only to be prepared once by offline algorithms. Once prepared, a cover may be applied to the classification of any given multi-label dataset whose properties conform with those of the cover. The contribution of their work was two-fold. First, they introduced SCP as a general framework for constructing label covers while allowing the user to incorporate cover construction constraints. They demonstrated the effectiveness of this framework by proposing two construction constraints whose enforcement produces covers that improve the prediction performance of random selection. Second, they provided theoretical bounds that quantify the probabilities of random selection to produce covers that meet the proposed construction criteria. The experimental results indicated that the proposed methods improve multi-label classification accuracy and stability compared with the RAKEL algorithm and to other state-of-the-art algorithms.

One of the most important steps in the design of a multi-classifier system (MCS), also known as ensemble, is the choice of the components (classifiers). This step is very important to the overall performance of a MCS since the combination of a set of identical classifiers will not outperform the individual members. The ideal situation would be a set of classifiers with uncorrelated errors – they would be combined in such a way as to minimize the effect of these failures, Canuto et al. [149] presented an extensive evaluation of how the choice of the components (classifiers) can affect the performance of several combination methods (selection-based and fusion-based methods). An analysis of the diversity of the MCSs when varying their components is also performed. As a result of this analysis, it is aimed to help designers in the choice of the individual classifiers and combination methods of an ensemble.

The idea of ensemble is adapted for feature selection. Canedo et al. [150] proposed an ensemble of filters for classification, aimed at achieving a good classification performance together with a reduction in the input dimensionality. With this approach, they tried to overcome the problem of selecting an appropriate method for each problem at hand, as it is overly dependent on the characteristics of the datasets. The adequacy of using an ensemble of filters rather than a single filter was demonstrated on synthetic and real data, paving the way for its final application over a challenging scenario such as DNA microarray classification.

Jin et al. [151] proposed a fuzzy ARTMAP (FAM) ensemble approach based on the improved Bayesian belief method is presented and applied to the fault diagnosis of rolling element bearings. First, by the statistical method, continuous Morlet wavelet analysis method and time series analysis method many features are extracted from the vibration signals to depict the information about the bearings. Second, with the modified distance discriminant technique some salient and sensitive features are selected. Finally, the optimal features are input into a committee of FAMs in different sequence, the output from these FAMs is combined and the combined decision is derived by the improved Bayesian belief method. The experiment results show that the proposed FAMs ensemble can reliably diagnose different fault conditions including different categories and severities, and has a better diagnosis performance compared with single FAM.

Studies have provided theoretical and empirical evidence that diversity is a key factor for yielding satisfactory accuracy-generalization performance with classifier ensembles. Nascimento et al. [152] tried to empirically assess the impact of using, in a sequential manner, three complementary approaches for enhancing diversity in classifier ensembles. For this purpose, simulations were conducted on 15 well-known classification problems with ensemble models composed of up to 10 different types of classifiers. Overall, the results evidence the usefulness of the proposed integrative strategy in incrementing the levels of diversity progressively.

Hu et al. [153] proposed a novel ensemble learning algorithm named Double Rotation Margin Forest (DRMF) that aims to improve the margin distribution of the combined system over the training set. They utilized random rotation to produce diverse base classifiers, and optimize the margin distribution to exploit the diversity for producing an optimal ensemble. They demonstrated that diverse base classifiers are beneficial in deriving large-margin ensembles, and that therefore their proposed technique will lead to good generalization performance. They examined their method on an extensive set of benchmark classification tasks. The experimental results confirm that DRMF outperforms other classical ensemble algorithms such as Bagging, AdaBoostM1 and Rotation Forest. The success of DRMF is explained from the viewpoints of margin distribution and diversity.

Timms et al. [154] developed three novel voting methods are presented for combining classifiers trained on regions with available examples for predicting rare events in new regions ;specifically the closure of shellfish farms. The ensemble methods introduced are consistently more accurate at predicting closures. Approximately 63% of locations were successfully learned with Class Balance aggregation compared with 37% for the Expert guidelines, and 0% for One Class Classification.

Kourentzes et al. [155] proposed a mode ensemble operator based on kernel density estimation, which unlike the mean operator is insensitive to outliers and deviations from normality, and unlike the median operator does not require symmetric distributions. The three operators are compared empirically and the proposed mode ensemble operator is found to produce the most accurate forecasts, followed by the median, while the mean has relatively poor performance. The findings suggested that the mode operator should be considered as an alternative to the mean and median operators in forecasting applications. Experiments indicated that mode ensembles are useful in automating neural network models across a large number of time series, overcoming issues of uncertainty associated with data sampling, the stochasticity of neural network training, and the distribution of the forecasts.

Yin et al. [156] formulated the classifier ensemble problem with the sparsity and diversity learning in a general mathematical framework, which proves beneficial for grouping classifiers. In particular, derived from the error-ambiguity decomposition, they designed a convex ensemble diversity measure. Consequently, accuracy loss, sparseness regularization, and diversity measure can be balanced and combined in a convex quadratic programming problem. They proved that the final convex optimization leads to a closed-form solution, making it very appealing for real ensemble learning problems. They compared their proposed novel method with other conventional ensemble methods such as Bagging, least squares combination, sparsity learning, and AdaBoost, extensively on a variety of UCI benchmark data sets and the Pascal Large Scale Learning Challenge 2008 web spam data. Experimental results confirmed that their approach has very promising performance.

Díez-Pastor et al. [157] presented two new methods for tree ensemble constructions are: G-Forest and GAR-Forest. In a similar way to Random Forest, the tree construction process entails a degree of randomness. The same strategy used in the GRASP Meta heuristic for generating random and adaptive solutions is used at each node of the trees. The source of diversity of the ensemble is the randomness of the solution generation method of GRASP. A further key feature of the tree construction method for GAR-Forest is a decreasing level of randomness during the process of constructing the tree: maximum randomness at the root and minimum randomness at the leaves. The method is therefore named ‘‘GAR’’, GRASP with annealed randomness. The results conclusively demonstrate that G-Forest and GAR-Forest outperform Bagging, AdaBoost, MultiBoost, Random Forest and Random Subspaces. The results are even more convincing in the presence of noise, demonstrating the robustness of the method.

Own and Abraham [158] proposed a novel weighted rough set as a Meta classifier framework for 14 classifiers to find the smallest and optimal ensemble, which maximize the overall ensemble accuracy. They proposed a new entropy-based method to compute the weight of each classifier. Each classifier assigns a weight based on its contribution in classification accuracy. Thanks to the powerful reduct technique in rough set, which guarantee high diversity of the produced reduct ensembles. The higher diversity between the core classifiers has a positive impact on the performance of minority class as well as in the overall system performance. Experimental results with ozone dataset demonstrated the advantages of weighted rough set Meta classifier framework over the well-known Meta classifiers like bagging, boosting and random forest as well as any individual classifiers.

Many researchers have worked on the ensemble of multiple algorithms to improve the performance of classification or prediction in data mining or machine learning. In our study we seek to develop a novel ensemble model for long term rainfall prediction by using Meta classifier

Vote combining with three base classifiers IBK, K-star and M5P, for increasing not only the accuracy of the prediction, but also to lead to greater confidence in the results.

2.3.6 Adaptive Neuro-Fuzzy Inference System (ANFIS) Technique

A hybrid intelligent system is one that combines at least two intelligent technologies [9], For example, combining a neural network with a fuzzy system results in a hybrid neuro-fuzzy system.

Neuro-fuzzy or fuzzy-neural structures [23], has largely extended the capabilities of both technologies in hybrid intelligent systems. The advantages of neural networks in learning and adaptation and those of fuzzy logic systems in dealing with the issues of human-like reasoning on a linguistic level, transparency and interpretability of the generated model, and handling of uncertain or imprecise data, enable building of higher level intelligent systems. The synergism of integrating neural networks with fuzzy logic technology into a hybrid functional system with low-level learning and high-level reasoning transforms the burden of the tedious design problems of the fuzzy logic decision systems to the learning of connectionist neural networks. In this way the approximation capability and the overall performance of the resulting system are enhanced.

A neuro-fuzzy system is, in fact, a neural network that is functionally equivalent to a fuzzy inference model. It can be trained to develop *IF-THEN* fuzzy rules and determine membership functions for input and output variables of the system. Expert knowledge can be easily incorporated into the structure of the neuro-fuzzy system. At the same time, the connectionist structure avoids fuzzy inference, which entails a substantial computational burden.

A number of different schemes and architectures of this hybrid system have been proposed, such as fuzzy-logic-based neurons [159], fuzzy neurons [160], neural networks with fuzzy weights [161], neuro-fuzzy adaptive models [162], etc. The proposed architectures have been successful in solving various engineering and real-world problems, such as in applications like system identification and modeling, process control, systems diagnosis, cognitive simulation, classification, pattern recognition, image processing, engineering design, financial trading, signal processing, time series prediction and forecasting, etc.

The adaptive neuro-fuzzy inference system is a common approach in which the two techniques such as a neural network and a fuzzy logic get combined [24] to create a complete shell. Basically the system of ANFIS applies the technique of the artificial neural network learning rules to determine and tune the fizzy inference systems' structure and parameters. A number of important features of ANFIS can help the system accomplish a task brilliantly; these features are characterized as easy to implement, fast and accurate learning, strong generalization abilities, excellent explanation facilities through fuzzy rules, and easy to incorporate both linguistic and numeric knowledge for problem solving [24-29].

In Neuro-Fuzzy technique a neural network is introduced to devise the fuzzy system so that the structure and parameters which identify the fuzzy rules are accomplished by adopting and optimizing the topology and the parameters of corresponding neuro fuzzy network based on data sets. The system is considered to be an adaptive fuzzy inference system with the capability of learning fuzzy rules from data and as a connectionist architecture provided with linguistic meaning. Jang had developed one type of hybrid neuro-fuzzy inference expert system that works in Takagi-Sugeno type fuzzy inference system [30-32]. This is called ANFIS that is one of the most successful schemes which combine the benefits of these two powerful paradigms into a single capsule.

There are several features of the ANFIS which enable it to achieve great success in a wide range of scientific applications. The attractive features of an ANFIS include: easy to implement, fast and accurate learning, strong generalization abilities, excellent explanation facilities through fuzzy rules, and easy to incorporate both linguistic and numeric knowledge for problem solving [27].

The Adaptive Neuro-Fuzzy Inference System (ANFIS), first introduced by Jang [25], is a universal approximator and, as such, is capable of approximating any real continuous function on a compact set to any degree of accuracy [27]. Thus, in parameter estimation, where the given data are such that the system associates measurable system variables with an internal system parameter, a functional mapping may be constructed by ANFIS that approximates the process of estimation of the internal system parameter. Since it integrates both neural networks and fuzzy logic principles, it has potential to capture the benefits of both in a single framework. Its inference system corresponds to a set of fuzzy *IF–THEN* rules that have learning capability to approximate nonlinear functions. Jang [25] introduced architecture and a learning procedure for the fuzzy inference systems (FIS) that uses a neural network learning algorithm for constructing a set of fuzzy if-then rules with appropriate membership functions (MFs) from the specified input-output pairs. There are two approaches for FIS, namely Mamdani [163] and Sugeno [24].

The differences between these two approaches arise from the consequent part. Mamdani's approach uses fuzzy MFs, whereas Sugeno's approach uses linear or constant functions.

Many researchers employed ANFIS approach in whether forecasting, Castellanos and James [164] explored Adaptive Neuro-Fuzzy Inference Systems (ANFIS) to forecast the average hourly wind speed. To determine the characteristics of ANFIS that best suited the target wind speed forecasting system, several ANFIS models were trained, tested and compared. Different types and number of inputs, training and checking sizes, type and number of membership functions were analyzed. Comparisons of the different models were performed and the results showed that the 4 inputs models generated by grid partitioning and the 6 inputs models generated by subtractive clustering provided the smallest errors with the models using wind speed and air pressure as inputs having the best forecasting accuracy. Sharma et al. [165] used ANFIS and Multiple linear regression model to analyze metrological data sets The data covers a five year period (2008-2012) were for the monthly means of minimum and maximum temperature, wind speed, and relative humidity and mean sea level pressure (MSLP). The performance evaluation of the two models that was carried out on the basis of root mean square error (RMSE) showed that the ANFIS model yielded better results than the multiple linear regression (MLR) model with a lower prediction error. Babu et al. [166] built Auto-Regressive Integrated Moving and Average (ARIMA) and Adaptive Network Based Fuzzy Inference System (ANFIS) models for weather forecasting. The climate determining is taken from University of Waterloo. The information was taken as Relative Humidity, Ambient Air Temperature, Barometric Pressure and Wind Direction. The results showed that ARIMA is most effective methods for weather forecasting when compared with ANFIS but, it took more time. Sharma et al. [165] tried to analyze metrological data sets by using Adaptive Neuro-Fuzzy Inference System (ANFIS) and Multi-Layer Perceptron (MLP) Artificial Neural Network (ANN) models. The data covered a five-year period (2008-2012) for the monthly means of minimum and maximum temperature, rainfall, wind run, and relative humidity and mean sea level pressure (MSLP). The results showed that both models could be applied to weather prediction problems. The performance evaluation of the two models that was carried out on the basis of root mean square error (RMSE) showed that the ANFIS model yielded better results than the MLP ANN model with a lower prediction error. Ramesh et al. [167] used an adaptive neuro-fuzzy inference system (ANFIS) to retrieve profiles of temperature and humidity up to 10 km over the tropical station Gadanki

(13.5_ N, 79.2_ E), India. ANFIS is trained by using observations of temperature and humidity measurements by co-located Meisei GPS radiosonde (henceforth referred to as radiosonde) and microwave brightness temperatures observed by radiometrics multichannel microwave radiometer MP3000 (MWR). ANFIS is trained by considering these observations during rainy and non-rainy days (ANFIS (RDCNRD)) and during non-rainy days only (ANFIS (NRD)). The comparison of ANFIS(RDCNRD) and ANFIS(NRD) profiles with independent radiosonde observations and profiles retrieved using multivariate linear regression (MVLR: RDCNRD and NRD) and artificial neural network (ANN) indicated that the errors in the ANFIS(RDCNRD) are less compared to other retrieval methods.

In the field of rainfall prediction and prediction of groundwater level, Mayilvaganan and Naidu [168] built ANFIS models for predicting groundwater level in Thurinjapuram watershed, Tamilnadu, India. The results were compared with different type of membership functions. The model with Gaussian membership functions gave the best performance among all given models. Bisht and Jangid [169] investigated the best model to forecast river discharge. They developed ANFIS and Multiple Linear Regression (MLR). The developed models were trained, tested & validated on the data of Godavari river at Rajahmundry, Dhawalaishwaram Barrage site in Andhra Pradesh. The results proved that the developed ANFIS models predicted better results the traditional models, like MLR. Alipour et al. [170] used artificial neural networks, Adaptive Neuro-Fuzzy Inference System and Time Series to find the best way to predict ground water levels in North Mahyar plain, Isfahan. The rainfall, temperature, relative humidity, the operation wells and aquifer fed by the near aquifer are considered as input data, and groundwater levels of 14 observed wells were considered as output. The results showed that the Adaptive Neuro-Fuzzy Inference System can give more accuracy for predicting groundwater level than Time Series analysis and artificial neural network.

The adaptive neuro-fuzzy inference system (ANFIS) has been widely used for modeling different kinds of nonlinear systems including rainfall forecasting. Akrami et al. [171] developed modified ANFlS (MANFlS) to making ANFIS technique more efficient regarding to Root Mean Square Error, Correlation Coefficient, Root Mean Absolute Error, Signal to Noise Ratio and computing epoch. In their study, two scenarios were introduced; in the first scenario, monthly rainfall was used solely as an input in different time delays from the time (t) to the time (t-4) to conventional ANFIS, second scenario used the modified ANFIS to improve the rainfall forecasting efficiency. The result showed that the model based Modified ANFIS performed higher rainfall forecasting accuracy; low errors and lower computational complexity (total number of fitting parameters and convergence epochs) compared with the conventional ANFIS model. Ytoui [172] studied the rainfall-runoff relationship modeling at monthly and daily time by comparing the performance of neuro-fuzzy inference system with Conceptual models GR2M and GR4J. The results showed the neuro-fuzzy inference system was more accurate than the Conceptual models. Aldrian and Djamil [173] investigated the use of multi variable Adaptive Neuro Fuzzy Inference System (ANFIS) in predicting daily rainfall using several surface weather parameters as predictors. The data used in that study came from automatic weather station data collected in Timika airport from January until July 2005 with 15-minute time interval. They found out that relative humidity is the best predictor with a stable performance regardless of training data size and low RMSE amount especially in comparison to those from other predictors. Other predictors showed no consistent performances with different training data size. Performances of ANFIS reached a slightly above 0.6 in correlation values for daily rainfall data without any filtering for up to 100 data in a time series. Mayya [174] developed ANFIS models for rainfall runoff modeling. The ANFIS models used different membership functions Triangular, Trapezoidal, Bell-shaped, Sigmoid and Gaussian method. West flowing river Nethravathi located in Karnataka, India has been selected as study area. It was observed that adaptive neuro fuzzy inference system using Triangular membership function showed a good performance compared to other models developed. Vafakhah et al. [175] used artificial neural networks (ANNs) and adaptive neuro-fuzzy interface system (ANFIS) for rainfall-runoff modeling. Rainfall, temperature and snow water Equivalent (SWE) were used as inputs for ANN and ANFIS. Root mean square error (RMSE), Nash–Sutcliffe efficiency coefficient (NS) and determination coefficient (R2) statistics are employed to evaluate the performance of the ANN and ANFIS models for forecasting runoff. Based on the results of test stage ANN was very good and superior to rainfall-runoff modeling than the ANFIS. Solgi et al. [176] used wavelet analysis combined with artificial neural network and then compared with adaptive neuro fuzzy system to predict the precipitation in Verayneh station, Nahavand, Hamedan, Iran. For this purpose, the original time series using wavelet theory decomposed to multiple sub time series. Then, these subseries were applied as input data for artificial neural network, to predict daily precipitation, and compared with results of adaptive neuro fuzzy system. The results showed that the combination of wavelet models and neural networks has a better performance than adaptive neuro fuzzy system, and can be applied to predict both short-term and long-term precipitations. Dastorani et al. [177] evaluated the applicability of artificial neural networks (ANN) and adaptive neoro-fuzzy systems (ANFIS) in prediction of precipitation yazd meteorological station in central Iran. Different architectures of ANN and ANFIS models have been done. Precipitation moving average, maximum temperature, relative humidity, mean wind speed, maximum wind direction and evaporation have been used as inputs of the models. Final results showed that the efficiency of time lagged recurrent network (TLRN) and ANFIS were almost the same. Charaniya and Dudul [178] designed adaptive neuro fuzzy inference system (ANFIS), linear and nonlinear regressive models to predict the rainfall in future for the next year based on the rainfall pattern for past four years. The experimental results showed that ANFIS model has better prediction capability due to combined power of fuzzy logic and neural network. But the execution time taken is more. In [179] rainfall has been predicted using Adaptive Neural Fuzzy Inference System (ANFIS) and the best input combination has been identified using Gamma Test (GT) for the rainfall prediction. Then, runoff was simulated by a conceptual hydrological MIKE11/NAM model and the results were compared together. The study area is Qaleh Shahrokh basin located in Iran. The ability of ANFIS and MIKE11/NAM models were evaluated based on Root Mean Square Error (RMSE), correlation coefficient (R2) and Efficiency Index (EI). The results showed that both models (NAM and ANFIS) had good capabilities in simulating discharge during calibration and verification periods. Using the predicted rainfall instead of the observed rainfall caused lower efficiency in the NAM model and runoff simulation. Panchal et al. [180] have used Adaptive Neuro-Fuzzy Inference System (ANFIS) for rainfall-runoff modeling for the Dharoi sub-basin, India. Different combinations of rainfall were considered as the inputs to the model, and runoff was considered as the output. Input space partitioning for model structure identification was done by grid partitioning. A hybrid learning algorithm consisting of back-propagation and least-squares estimation was used to train the model for runoff estimation. The optimal learning parameters were determined by trial and error using Triangular membership function. Root mean square error (RMSE) and correlation coefficient (r) were used for selecting the best performing model. The results showed the best Rainfall-Runoff model for the Hadad, Khedbrhama and Dharoi rain gauge stations had 7 triangular type membership functions with the input and output training and testing ratio of 80-20%. In [181], an adaptive neuro-fuzzy inference system (ANFIS) model has been proposed to forecast the rainfall for Klang River in Malaysia on monthly basis. To be able to train and test the ANFIS and ANN models, the statistical data from 1997 to 2008, was obtained from Klang gates dam data. The optimum structure and optimum input pattern of model was determined through trial and error. Different combinations of rainfall were produced as inputs and five different criteria were used in order to evaluate the effectiveness of each network and its ability to make precise prediction. The performance of the ANFIS model is compared to artificial neural network (ANN) model. The five criteria are root mean square error, Correlation Coefficient, Nash Sutcliffe coefficient, gamma coefficient and Spearman correlation coefficient. The result indicate that the ANFIS model showed higher rainfall forecasting accuracy and low error compared to the ANN model. Furthermore, the rainfall estimated by this technique was closer to actual data than the other one. Vafakhah et al. [182] proposed adaptive neuro-fuzzy inference system (ANFIS), artificial neural network (ANN) and wavelet-artificial neural network (Wavelet-ANN) models for modeling rainfall-runoff (RR) relationship. daily stream flow time series of hydrometric station of Hajighoshan on Gorgan River and the daily rainfall time series belonging to five meteorological stations (Houtan, Maravehtapeh, Tamar, Cheshmehkhan and Tangrah climatologic stations) were used for period of 1983-2007. Root mean square error and correlation coefficient statistics were employed to evaluate the performance of the ANN, ANFIS, ARX and ARMAX models for rainfall-runoff modeling. The results showed that ANFIS models outperformed the system identification, ANN and Wavelet-ANN models. Fallah-Ghalhary et al. [183] tried to analyze 33 years of rainfall data in khorasan state, the northeastern part of Iran by using 3 soft computing based prediction models with 33 years of rainfall data. For performance evaluation, network predicted outputs were compared with the actual rainfall data. Simulation results reveal that soft computing techniques are promising and efficient. the test results using by ANN, FIS and ANFIS learning algorithms showed that the lowest RMSE was obtained using ANN, ANFIS and FIS, it was 41, 52 and 58 millimeter respectively. For modeling suspended sediment load (SSL), researchers [184] compared three different soft computing methods, namely, artificial neural networks (ANNs), adaptive neuro-fuzzy inference system (ANFIS), coupled wavelet and neural network (WANN), and conventional sediment rating curve (SRC) approaches for estimating the daily SSL in two gauging stations in the USA. The performances of these models were measured by the coefficient of correlation, Nash-Sutcliffe efficiency coefficient, root mean-square error, and mean absolute percentage error (MAPE) to choose the best fit model. Obtained results demonstrated that applied soft computing models were in good agreement with the observed SSL values, while they depicted better results than the conventional SRC method. The comparison of estimation accuracies of various models illustrated that the WANN was the most accurate model in SSL estimation in comparison to other models.

Bekuretsion and Beshah [185] employed neural network, mamdani and sugeno adaptive neuro fuzzy models to predict rainfall of Ethiopian Upper Blue Nile Basin with different time lag. The result shows the soft computing models perform the prediction with relatively small error. The developed soft computing models show better skill than techniques used by Ethiopian National Meteorological Service Agency (ENMSA) and other previous studies which used statistical techniques.

Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference Systems (ANFIS) have been investigated for estimation of effective rainfall [186]. Tabriz synoptic station located in the northwest of Iran was considered as studying area and USDA-SCS method as the most widely used method for calculation of effective rainfall was applied to determine the monthly effective rainfalls, for the period of 1982 to 2010. The results showed the accuracies of both models are satisfied, and ANNs showed relatively more accurate results. Abraham et al. [187] compared the performance of several Soft Computing (SC) models: Evolving Fuzzy Neural Network (EFuNN), Artificial Neural Network using Scaled Conjugate Gradient Algorithm (ANNSCGA), Adaptive Basis Function Neural, Network (ABFNN) and General Regression Neural Network (GRNN) with Multivariate Adaptive Regression Splines (MARS) which is a regression technique that uses a specific class of basis functions as predictors in place of the original data. 87 years of rainfall data in Kerala state have been used; SC and MARS models were trained with 40 years of rainfall data and tested with 47 years of rainfall data. Simulation results revealed that MARS is a good forecasting tool and performed better than the considered SC models.

Patel and Parekh [188] investigated the development of an efficient model to forecast monthly monsoon rainfall for Gandhinagar station using Adaptive Neuro Fuzzy Inference System (ANFIS). Eight models were developed using various membership functions and climatic parameters as inputs. In their study, the generalized bell-shaped built-in membership function (gbell) has been used as a membership function in both Hybrid and Back propagation method for ANFIS. The four-evaluation parameters Root mean square error, Correlation Coefficient, Coefficient of Determination and Discrepancy ratio are used to evaluate the developed model. The study showed that hybrid Model with seven membership functions and using three inputs, temperature, relative humidity and wind speed gives best result to forecast rainfall for study area. Niksaz and Latif [189] applied ANFIS for rainfall events evaluation. Four parameters: Temperature, relative humidity, total cloud cover and due point were the input variables for ANFIS model, each has 121 membership functions. The data is six years METAR data for Mashhad city (2007-2012). Different models for Mashhad city stations were constructed depending on the available data sets. Among the overall 25 possibilities one model with one hundred twenty one fuzzy IF-THEN rules has chosen. The output variable is 0 (no rainfall event) or 1 (rainfall event). Results showed a high agreement with the actual data.

Faulina and Suhartono [190] proposed hybrid and ensemble model of forecasting method for ten-daily rainfall prediction based on ARIMA (Autoregressive Integrated Moving Average) and ANFIS (Adaptive Neuro Fuzzy Inference System) at six certain areas in Indonesia. To find an ensemble forecast from ARIMA and ANFIS models, the averaging and stacking method was implemented. In this study, Triangular, Gaussian, and Gbell function are used as membership function in ANFIS. The best model is measured by the smallest root of mean square errors (RMSE) at testing datasets. The results show that an individual ARIMA method yields more accurate forecast in five rainfall data, whereas ensemble averaging multi model yields better forecast in one rainfall data. In general, these results indicated that more complicated model not always yield better forecast than simpler one.

2.4 Geographical Information System (GIS) and Heat Maps

Many studies used GIS technology and produced heat maps have been carried out to many disciplines such as risk assessment [191-193], weather and climate [194-196], agriculture [197-199], Health and Disease [200] and environmental studies [201-204].

In the field of Geostatistical modeling, Seng et al. [205] used GIS to model, analysis and map Epidemiology of Dengue Fever in Johor State, Malaysia. In the same context Haarman et al. [206] proposed feature-expression heat maps as a new visual method to explore complex associations between two variable sets. Their results showed the feature-expression heat map is a useful graphical instrument to explore associations in complex biological systems where one-way direction is assumed, such as genotype-phenotype pathophysiological models. For mining the electronic medical record (EMR) to deliver new medical knowledge about causal effects, which are hidden in statistical associations between different patient attributes, Toddenroth et al. [207] utilized heat maps to visualize tabulated metric datasets as grid-like choropleth maps, and thus present measures of association between numerous attribute pairs clearly arranged. Simulation results demonstrate that the proposed clustering procedure rearranges attributes similar to simulated statistical associations. Thus, heat maps are an excellent tool to indicate whether associations concern the same attributes or different ones, and whether affected attribute sets conform to any preexisting relationship between attributes. Koita et al. [208] examined the seasonality of malaria parasite prevalence in the dry northern part of Mali at the edge of the Sahara desert. Results showed lower prevalence in hot dry than cold dry, Malaria remained stable in the villages with ponds but varied between the 2 seasons in the villages without ponds Malaria was meso-endemic in the study area. Also as application of heat maps in medical field, Sudhakara et al. [209] used a combination of RS and GIS approach to develop landscape predictors of sandfly abundance. Result showed that rural villages surrounded by higher proportion of transitional swamps with soft stemmed edible plants and banana, sugarcane plantations had higher sandfly abundance and would, therefore, be at higher risk prone areas for man-vector contact.

Kawasaki et al. [210] employed GIS for assessing stream water chemistry in a forested watershed. Survey supporting maps were prepared for effective data collection in the field by considering some working hypotheses. Then, factors contributing to water chemistry were analyzed using fine-scale spatial data in an individual catchment. Finally “nitrogen leaching prediction map” was created for decision support concerning the type and location of actions to properly manage future forest ecosystems in the Tanzawa Mountains based on data from 51 sampling points.

In the field of hazard and risks assessment, Pramojanee et al. [211] applied remote sensing interpretation and GIS for mapping of Flood Hazard and Risk Area in Nakorn Sri Thammarat Province, South of Thailand. By means of weighting, a number of causative factors including annual rainfall, size of watershed, side slopes of watershed, gradient of river and stream, drainage density type of soil and land use, communication line and infrastructures, and population density were considered for rating the degree of hazard and risk. Maps of two scales, 1:250,000 of the region and 1:50,000 of two provinces were obtained. Their study discussed only the procedures that were applied for mapping flood hazard and risk areas in one of the two provinces. The study suggested that about 5 percent of the total area of the province is under a high risk of flooding whereas 17 and 22 percent are under moderate and low risk respectively. Comparing the moderate and high risk areas identified in this study with the ground truth data obtained from the field work however, it is found that the risk area obtained from the study is about 20-25 percent higher than the actual one. Dhakal et al. [212] applied GIS for landslide hazard assessment using multivariate statistical analysis, mapping, and the evaluation of the hazard maps. The study area was the Kulekhani watershed (124 km2) located in the central region of Nepal. To determine the factors and classes influencing landsliding, layers of topographic factors derived from a digital elevation model, geology, and land use-land cover were analyzed by quantification scaling type II (discriminant) analysis, and the results were used for hazard mapping. The geology was found to be the most important factor for landslide hazard. The scores of the classes of the factors quantified by the five analyses were used for the hazard mapping in the GIS, with four levels of relative hazard classes: high, moderate, less, and least. The evaluation of five hazard maps indicated higher accuracy for the combinations in which the non-landslide group was generated by the unaligned stratified random sampling method. The agreements in the hazard maps, produced from different sample combinations using unaligned stratified random sampling for selecting non landslide group, were within the acceptable range for the practical use of a hazard map. Chien-Yuan et al. [213] analyzed time-varying rainfall infiltration induced landslide. Results of the analyses showed that under heavy rainfall conditions, the infiltrated slope is unstable and the time of debris masses movement initiated is correlated to the recorded time. While da Silva et al. [214] constructed risk maps for agricultural use from data obtained from remote sensing technology. They used ground temperature, or land surface temperature (LST), data distributed by EUMETSAT/LSASAF (with a spatial resolution of 3×3 km(nadir resolution) and a revisiting time of 15 min) to generate one of the most commonly used parameters in pest modeling and monitoring: “thermal integral over air temperature (accumulated degree-days)”. The results showed a clear association between the accumulated LST values over a threshold and the accumulated values computed from meteorological stations over the same threshold (specific to a particular tomato pest). The results were very promising and enabled the production of risk maps for agricultural pests with a degree of spatial and temporal detail that is difficult to achieve using in-situ meteorological stations.

In the field of the GIS and climate mapping, Daly et al. [215] applied the regression based Parameter-elevation Regressions on Independent Slopes Model (PRISM) to generate maps of mean monthly and annual precipitation and minimum and maximum temperature for the Caribbean islands of Puerto Rico, Vieques and Culebra over the 1963-1995 averaging period. Overall, the full PRISM approach resulted in greatly improved performance over simpler methods for precipitation and January minimum temperature, but only a small improvement for July maximum temperature. Comparisons of PRISM mean annual temperature and precipitation maps to previously-published, hand-drawn maps showed similar overall patterns and magnitudes, but the PRISM maps provided much more spatial detail. While Lindberg et al. [216] studied the Impact of city changes and weather on anthropogenic heat flux in Europe for the interval 1995–2015. They produced heat maps for population density, annual average and daily average of anthropogenic heat flux.

In study of output prediction of large-scale photovoltaics by wind condition analysis using 3D topographic maps Obara et al. [217] developed an analysis method to determine the temperature distribution and electrical conversion efficiency of a photovoltaic module by considering the wind conditions at the installation site. Modular temperature distribution and DC electrical conversion efficiency were obtained by introducing the physical properties of the photovoltaic module, wind conditions, and climatic conditions (plane of array irradiance and ambient temperature) by using a digital three-dimensional topographic map in a heat-transfer calculation. The case analysis resulted suggest a high value for the power-production efficiency for low ambient temperature. However, the difference in the temperature distribution of the photovoltaic module in relation to the difference in the ambient temperature is strongly influenced by wind velocity and wind direction. Moreover, when a large-scale photovoltaic power plant is installed on a complicated mountain slope, the cooling effect is controlled so that the indraft wind velocity on the photovoltaic module decreases. Therefore, in order to maintain high electrical conversion efficiency in the photovoltaic module, the best location for installation is an airy and flat area, as much as possible. According to the case analysis, the electrical conversion efficiency of the photovoltaic module at the time of the analysis under wind condition increased 23% (maximum) compared with that without wind conditions. Wilk and Andersson

[218] determined the areal distribution of precipitation by using a GIS-supported method and other factors (altitude and slope). Also Ye et al. [219] studied glacier variations in the Geladandong mountain region of central Tibet using RS and GIS technics. Data from Landsat images at three different times, 1973–76, 1992 and 2002 were compared with glacier areas digitized from a topographic map based on aerial photographs taken in 1969. Findings showed that there was accelerated glacier retreat in recent years, attributable to increase in summer air temperature. Omuto [220] while tracing the footprint of vegetation dynamics modelled a relationship between Advanced Very High Resolution Radiometer (AVHRR) / Moderate Imaging Spectro-radiometer (MODIS) NDVI and rainfall using regression analysis. Results showed a high correlation between rainfall and NDVI, which proved that vegetation trend monitoring with RS and GIS could give accurate indication of climate change.

In environmental and public health researches Pleil et al. [221] studied the human systems interactions with the environment and they produced heat maps for that purpose. The results showed that visualization of complex measurement data via a heat map approach is a valuable screening tool for quickly testing broad hypotheses regarding relationships among exposure measurements, biomarkers, meta-data, and host factors before computational efforts are expended. Santhi et al. [222] used GIS based hydrological model for estimating canal irrigation demand in the Lower Rio Grande Valley in Texas. Estimated potential water savings were 234.2, 65.9, and 194.0 Mm3 for conservation measures related to on−farm management improvements. It concluded that GIS would be useful for irrigation planning. Yelwa and Eniolorunda [223] simulated the spatial trend of desertification in Sokoto and its environs, Nigeria, using a time series 1-km SPOT Normalized Difference Vegetation Index (NDVI) and GIS. Results showed the direction of desertification movement and that the inter-annual vegetation vigour exhibited a diminishing trend over the time series.

Choo, et al. [224] presented iVisClassifier, which is a system based on supervised linear discriminant analysis that allows the user to iteratively label data and recomputed clusters and projections. By using heat maps, iVisClassifier gives an overview about the cluster relationship in terms of pair wise distances between cluster centroids both in the original space and in the reduced dimensional space.

The monitoring of vegetation degradation processes is an important component in developing appropriate conservation strategies aimed at landscape management for continued human existence [225]. RS and GIS can suitably be used for characterizing vegetation phenology.

2.5 Review Analysis

We presented a review of the use of different computational intelligence tools for weather forecasting and found the unique characteristics of ANNs: adaptability, nonlinearity and arbitrary function mapping ability make them quite suitable and useful for weather forecasting tasks. Overall, ANNs give satisfactory performance in weather forecasting and they surpassed the traditional models. Gorr et al. [226] believed that ANNs can be more appropriate for the following situations:

(1) Large data sets

(2) Problems with nonlinear structure

(3) The multivariate time series forecasting problems

After the review of a wide range of ANN architectures for weather forecasting, it is observed that most of the researchers have used BPN and RBFN techniques for forecasting various weather phenomenon e.g. rainfall, temperature, flood, rainfall-runoff etc, wind, and found significant results using the same architectures. Most of the scientists have concluded that BPNN and RBFN are the appropriate method to predict weather phenomenon. However there are some limitations of neural networks models such as:

1. ANNs are black-box methods. There is no explicit form to explain and analyze the relationship between inputs and outputs. This causes difficulty in interpreting results from the networks. Also no formal statistical testing methods can be used for ANNs.
2. ANNs are prone to have over fitting problems due to their typical, large parameter set to be estimated.
3. There are no structured methods today to identify what network structure can best approximate the function, mapping the inputs to outputs. Hence, the tedious experiments and trial-and-error procedures are often used.
4. ANNs usually require more data and computer time for training.

Liong and He [227] explained that Neural networks offer a number of advantages, including requiring less formal statistical training, ability to implicitly detect complex nonlinear relationships between dependent and independent variables, ability to detect all possible interactions between predictor variables and the availability of multiple training algorithms. Disadvantages include its “black box” nature, greater computational burden, and proneness to over fitting and the empirical nature of model development. Table 1 summarizes the previous related works, it shows the advantages, the limitations and the technologies have been used for weather forecasting.

Table 2.1: Comparative analysis of techniques in Meteorological forecasting.

Technology	**Advantages**	**Limitations**
Neural networks	1- Handle nonlinearity. 2- Does not require pre knowledge about domain. 3- Requires less data preparation. 4- Ability to learn and (generalization). 5-Decrease complexity of mathematical computing and increase accuracy.	1- Does not show the relation between the inputs and the output. 2- Require more data and computer time for training. 3- There are no structured methods to identify what network structure can best approximate the function.
Fuzzy logic	1 -Simplify knowledge acquisition and representation. 2- Solution to nonlinear problems. It allows heuristic decision-making strategies to be formulated by natural language rules rather than mathematical models.	1- Require more fine tuning and simulation before operational. 2- Highly dependent on domain expert’s knowledge, the knowledge extraction process is crucial as the whole fuzzy system is dependent on the domain expert knowledge. 3- Insufficient design standard or methodology the researchers use their own ways to design their applications. They usually use heuristic or trial and error approach in selecting the types of membership functions, inference engine and defuzzification methods. This approach is time-consuming.
Evolutionary algorithms	1- Hybridization with Other Methods: They can be used to	1- No guarantee for finding optimal solutions in a finite amount of time.

	optimize the performance of neural networks, fuzzy systems, production systems, wireless systems. 2- Parallelism, The evaluation of each solution can be handled in parallel. 3- Conceptual Simplicity: The evolutionary algorithm consists of initialization, iterative variation and selection in light of a performance index. Pre-knowledge is not required.	2- Parameter tuning mostly by trial-and-error.
Machine Learning	1- Deal with numerical or categorical variables. 2- Copes with noise. 3- Gives expected error rate. 4- Good predictive power.	1- Can generate large trees that require pruning. 2- Harder to classify more than 2 classes. 3- Poor at handling irrelevant attributes. 4- Can be affected by noise.
Statistical models	1- Good skill for long rage forecasts. 2- Use of multiple predictors. 3- Shows explicit correlations between observations of time series.	1- Not useful to study the highly nonlinear relationships between rainfall and its predictors. 2- There is no ultimate end in finding the best predictors.
Numerical models	1- Suitable to short range weather prediction (used to generate either short term weather forecasts or longer-term climate predictions). 2- Achieve very high resolution simulation	1- The accuracy of the models depends on the initial conditions, which are inherently incomplete. 2- Are not able to achieve satisfactory results in domestic cases. 3- Weak performance in order to predict the long-term

	of severe weather precipitation systems.	seasonal rain even for a large spatial scale. 4- Needs high performance computing and memory space to get more accuracy. 5- Common method to forecast weather which involves a complex of mathematical computing.

.

2.6 Summary

This Chapter presented an overview of using the various computational intelligence tools in weather forecasting, describing the main contributions on this field and providing taxonomy of the existing proposals according to the type of tools used.

Also the Chapter provided review analysis; it focused on the capabilities and limitations of several techniques that used in the prediction of several weather phenomena such as rainfall, temperature, wind, flood and tidal level etc.

Finally, geographical information system and producing heat maps have been discussed in much detail. And up-to-date literature review to many disciplines has been illustrated.

3. Data and Methodologies

3.1 Introduction

This chapter explains the research methodology for long-term rainfall prediction in Sudan. The research methodology includes five phases.

The First phase covers dataset and statistical analysis, followed by data transformation, data preprocessing, data normalization and feature selection to determine the most influencing and important variables that affect on the long term rainfall prediction out of the existing one.

The Second phase describes the base algorithms (Gaussian Processes, Linear Regression, Multilayer Perceptron, IBk, KStar, Decision Table, M5Rules, M5P, REP Tree and User Classifier) and Meta algorithms (Additive Regression, Bagging, Multi Scheme, Random Subset, Regression by Discretization, Stacking, and Vote) and also shows the proposed framework of ensemble method. This proposed ensemble has been constructed based on Meta classifier Vote combining with three base classifiers IBK, K-star and M5P.

The Third phase about ANFIS model includes the framework, structure, membership functions, rules, optimization methods and different dataset ratios for training and testing.

The Fourth phase includes data performance evaluation for each scheme. Finally, the result and discussion for all experiments are available in Chapters 4, 5, and 6.

The Fifth phase explains how it could translate the rainfall prediction into a form of rainfall maps using geographic information system (GIS).

3.2 Meteorological Data set

The meteorological data that used in this research has been brought from Central Bureau of Statistics, Sudan for 13 years from 2000 to 2012 for 24 meteorological stations over the country with 3732 samples. These stations are shown in Table 3.1.

Table 3.1: The names of meteorological stations.

Code	Station	Code	Station
1	Khartoum	13	El Nihood
2	Dongola	14	Kadugli
3	Atbara	15	Nyala
4	Abu Hamad	16	Elgeneina
5	Karima	17	El Fashir
6	Wadi Halfa	18	Kosti
7	Wad Medani	19	El Damazen
8	El Deweim	20	New Halfa
9	Kassala	21	Babanusa
10	Port Sudan	22	Rashad
11	El Gadarif	23	Abu Naama
12	Elobied	24	Sinnar

The dataset had eight (8) attributes containing monthly averages data. Their type and description is presented in Table 3.2, while Table 3.3 shows the analysis of numerical data values.

Table 3.2: The attributes of meteorological dataset.

Attribute	Data type	Description
Station	Nominal	Name of meteorological station
Date	Nominal	Month considered
Max-T	Numerical	Monthly Average maximum temperature in centigrade degrees
Min-T	Numerical	Monthly Average minimum temperature in centigrade degrees
Humidity	Numerical	Relative humidity
Wind D	Nominal	Wind direction
Wind S	Numerical	Wind speed (in knot)
Rainfall	Numerical	Monthly rainfall (in mms)

Table 3.3: Analysis of numeric data values.

Attribute	Max	Min	Average	Standard deviation
Station	-	-	-	-
Date	-	-	-	-
Max-T	47.1	20.6	36.43	4.19
Min-T	42.5	7.3	21.63	4.41
Humidity	88	3	36.21	18.11
Wind D	-	-	-	-
Wind S	70	1	5.78	3.40
Rain	295.3	0	26.16	52.4

Figures (3.1-3.5) show the monthly numerical data of our dataset (maximum temperature, minimum temperature, humidity, wind speed and rainfall) respectively for 13 years (2000-2012).

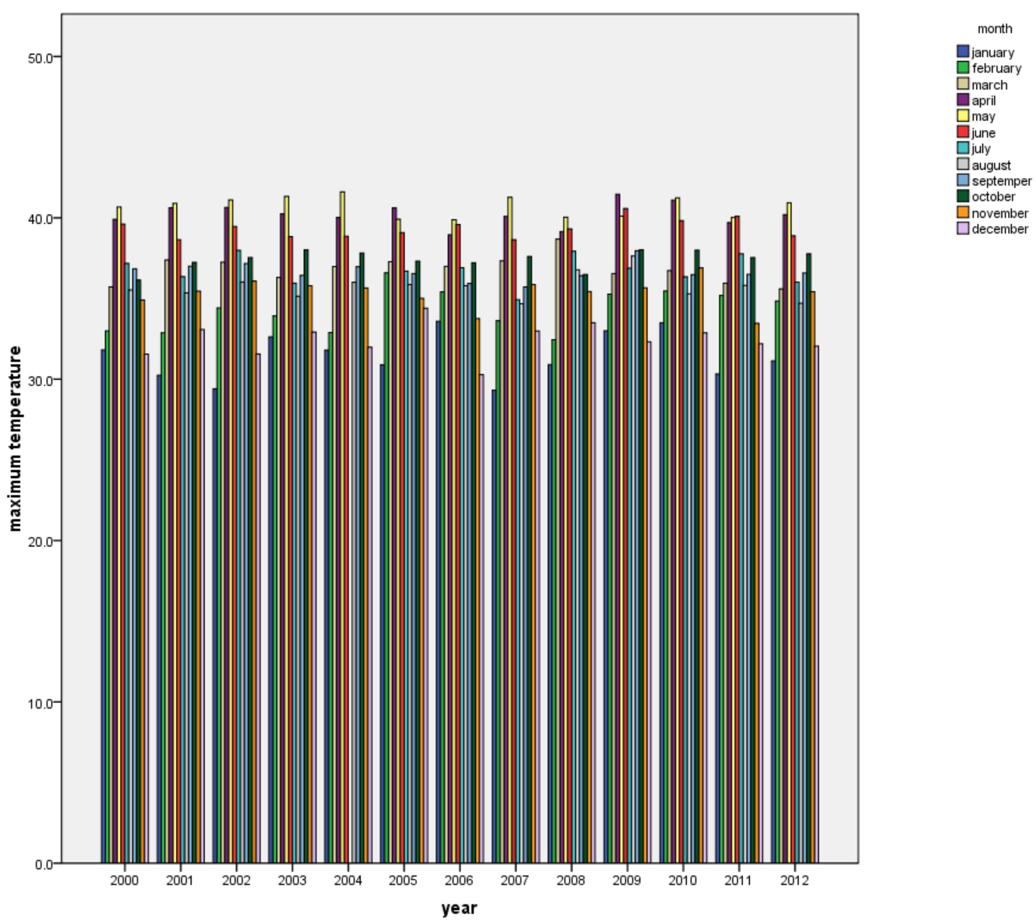


Figure 3.1. Maximum temperature.

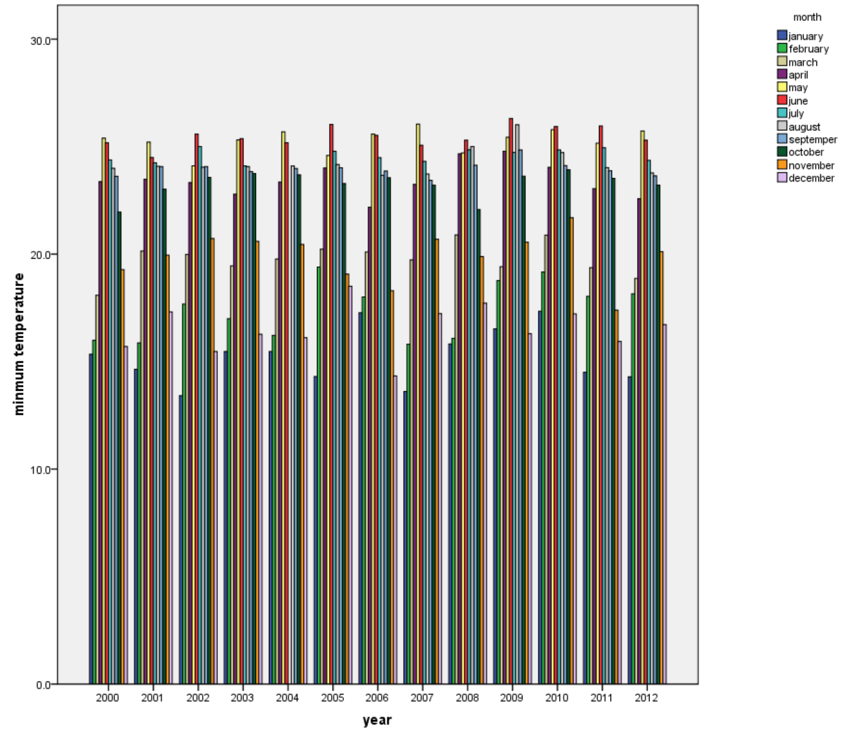


Figure 3.2. Minimum temperature.

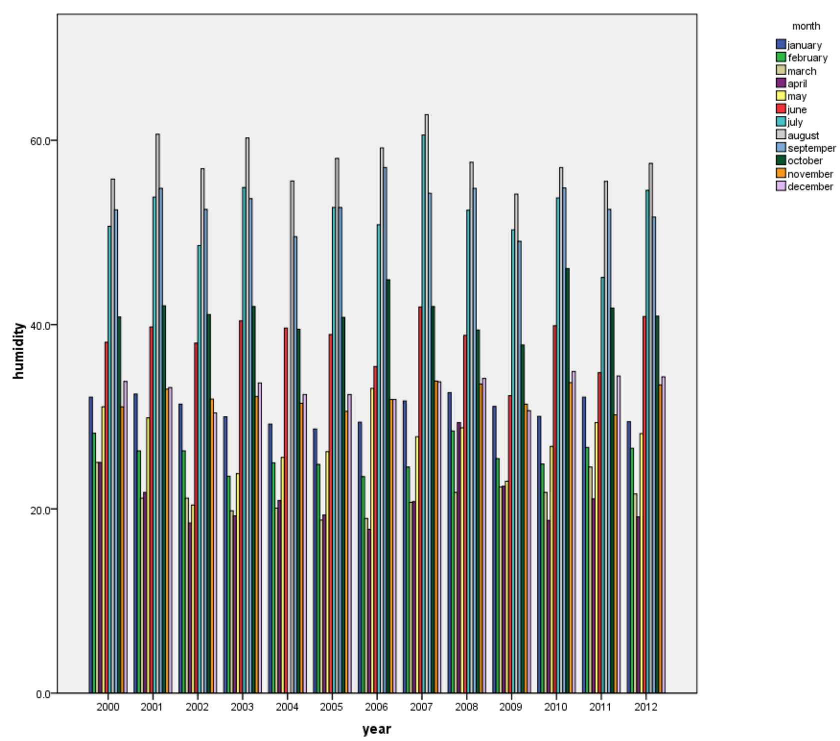


Figure 3.3. Relative humidity.

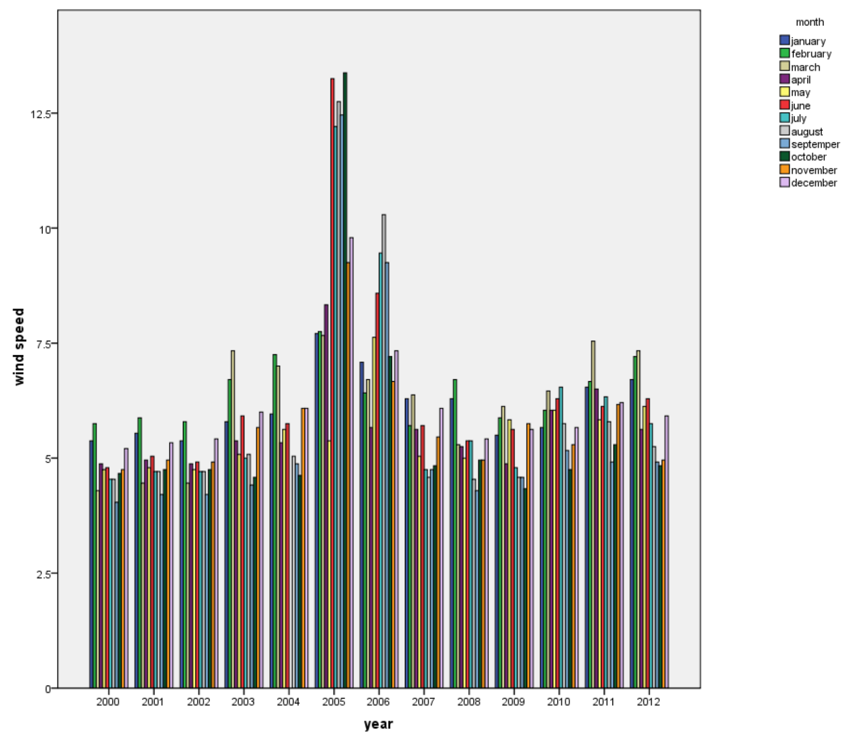


Figure 3.4. Wind speed.

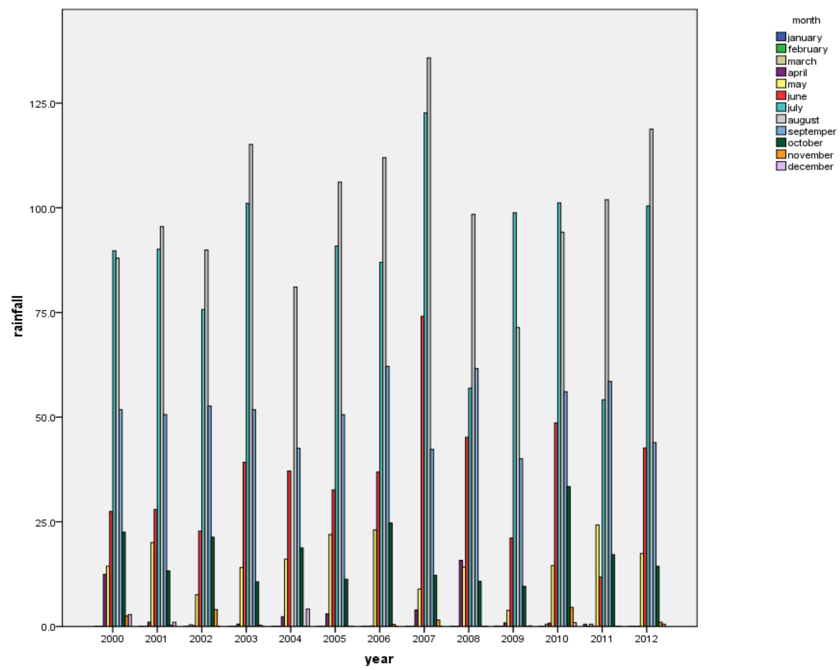


Figure 3.5. Rainfall amount.

3.2.1 Data Transformation

It is the stage data is transformed into forms appropriate for data mining. Firstly we converted the hard copy dataset into soft copy. After that the separated tables for 24 stations have been aggregated in one dataset. Also we expressed of rainfall by 0 and 1, zero if rainfall amount is less than or equal 0.1 mms and one if it's greater than 0.1mms. Then the data file was saved in Commas Separated Value (CVS) file format.

3.2.2 Data Preprocessing

The data obtained till now is noisy and there are some missing values and some unwanted data. We have to clean the data by filling missing values and removing the irrelevant data.

As shown in Table 3.4, we found that all missing values fall in only two attributes Wind-D and Wind-S with total ratio 0.028135. In this stage, a consistent format for the data model was developed which took care of missing data, finding duplicated data, and weeding out of bad data. The unsupervised attribute filter algorithm called Replace Missing Values has been used. That algorithm replaces all missing values for nominal and numeric attributes in a dataset with the modes and means from the training data.

Table 3.4: Analysis of missing values.

Attribute	Missing values
Station	0%
Date	0%
Max-T	0%
Min-T	0%
Humidity	0%
Wind D	0.073955%
Wind S	0.151125%
Rain	0%

3.3 Feature Selection

It is often an essential data processing step prior to applying a learning algorithm. Reduction of the attribute dimensionality leads to a better understandable model and simplifies the usage of different visualization technique and is the process of identifying and removing as much irrelevant and redundant information as possible. Reduces the dimensionality of the data, may allow learning algorithms to operate faster and more effectively and, accuracy can be improved later on future classification. It finds minimum set of attributes such that resulting probability distribution of data classes is as close as possible of original distribution. Methods used for Attribute Selection (AS) can be classified into two types. The filter approach and Wrapper approach. The filter approach actually precedes the actual classification process. The filter approach is independent of the learning algorithm, computationally simple fast and scalable. Using filter method, attribute selection is done once and then can be provided as input to different algorithms [228]. Wrapper approach uses the method of classification itself to measure the importance of attribute set; hence the attribute selection depends on the algorithm model used. Wrapper methods are too expensive for large dimensional database in terms of computational complexity and time since each attribute set considered must be evaluated with the classifier algorithm used. Filter methods are much faster than wrapper methods and therefore are better suited to high dimensional data sets. Some of these filter methods do not perform attribute selection but only attribute ranking hence they are combined with search method when one needs to find out the appropriate number of attributes. Such filters are often used with forward, backward elimination, bi-directional search, best-first search, and other methods [228, 229]. Various attribute selection techniques have been proposed in the literature such as:

a) Correlation-based Feature Selection (CFS)

   CFS is a filter algorithm that ranks feature subsets according to a correlation based heuristic evaluation function.CFS assumes that useful feature subsets contain features that are predictive of the class but uncorrelated with one another. CFS computes a heuristic measure of the "merit" of a feature subset from pair-wise feature correlations and a formula adapted from test theory. Heuristic search is used to traverse the space of feature subsets in reasonable time; the subset with the highest merit found during the search is reported [228].

b) Classifier subset evaluation

   This method of attribute subset evaluation techniques [230] uses a classifier to evaluate the attribute set.

c) Relief Attribute Evaluation

   The main idea of Relief algorithm [231] is to evaluate and estimate the quality of attributes according to distinguishing values between the instances that are near to each other. Both Relief and its extension ReliefF [232] are aware of the content information and can correctly estimate the quality of attributes in classification tasks with strong dependencies between attributes [233].

d) Wrapper attributes Selection

It depends on an induction algorithm to estimate the merit of feature subsets [228].

In this research to determine the most influencing and important variables that affect on the long term rainfall prediction out of the existing one, many attributes evaluator algorithms such as (correlation based feature selection subset evaluator, Classifier subset evaluator, relief attribute evaluator and Wrapper subset evaluator) have been implemented with appropriate different search methods such as (best-first, evolutionary search, exhaustive search, genetic search, greedy stepwise, linear forward selection, PSO search, random search, scatter searchV1, subset size forward selection, Tabu Search and Ranker).

3.4 Data Normalization

One of the steps of data pre-processing is data normalization. Normalization may be applied. For example, normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements [234]. The need to make harmony and balance between data, data must be normalized between 0 and 1. Eq. (1) was used to normalize our dataset.

$$x_n = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{1}$$

Where x is actual data and $\chi_{\min}$ is minimum value of original attribute's values and $\chi_{\max}$ is maximum value of original attribute's values.

3.5 Intelligent Data Analysis: Methodologies Used

For creating the models of the base and Meta algorithms we followed the steps that appear in Figure 3.6

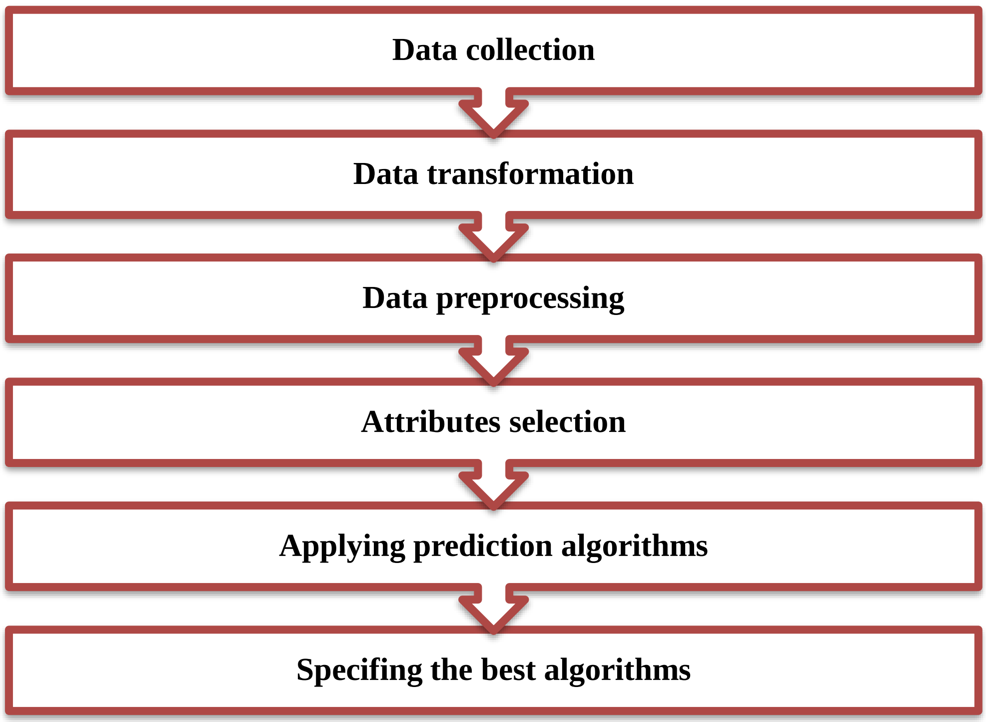


Figure 3.6. The methodology for base and Meta algorithms.

3.5.1 The Base Algorithms

The following 10 methods have used as individual base algorithms to create the rainfall prediction models.

3.5.1.1 Gaussian Processes

GP is based on the assumption that observations follow a normally distributed stochastic process. This leads to the conclusion, that new observations do not change the probability distribution of earlier ones. Based on this simple property Gaussian process regression allows predictions for unknown values [235]. A Gaussian process is stochastic process, any linear functional applied to the sample function Xt will give a normally distributed result. We can write:

$$f \sim GP(m,K) \quad (2)$$

That means the random function f is distributed as a GP with mean function m and covariance function K.

3.5.1.2 Linear Regression

Is an approach for modeling the relationship between a scalar dependent variable y and one or more explanatory variables denoted X. In linear regression, data are modeled using linear predictor functions, and unknown model parameters are estimated from the data. Such models are called linear models [236]. In the case of prediction or forecasting, linear regression can be used to fit a predictive model to an observed data set of y and X values. After developing such a model, if an additional value of X is given without its accompanying value of y, the fitted model can be used to make a prediction of the value of y [237]. If we have a data set $\{y_i, x_{i1}, \cdots, x_{ip}\}_{i=1}^{n}$ of n statistical units, a linear regression model assumes that the relationship between the dependent variable y_i and the p-vector of regressors x_i is linear. This relationship is modeled through a disturbance term or error variable ε_i — an unobserved random variable that adds noise to the linear relationship between the dependent variable and regressors. Thus the model takes the form:

$$y = \beta_1 x_{i1} + \cdots + \beta_p x_{ip} = x_i^T \beta + \varepsilon_i \qquad (3)$$

Where: $i = 1 \cdots n$,T denotes the transpose, so that $x_i^T \beta$ is the inner product between vectors x_i and β. often these n equations are stacked together and written in vector form as:

$$Y = X\beta + \varepsilon \qquad (4)$$

Where:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} x_{11} \cdots x_{1p} \\ x_{21} \cdots x_{2p} \\ \vdots \ \ddots \ \vdots \\ x_{n1} \cdots \ x_{np} \end{pmatrix}, \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{pmatrix}, \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$

3.5.1.3 Multilayer Perceptron (MLP)

MLP is a model that maps sets of input data onto a set of appropriate outputs. A MLP consists of multiple layers of nodes in a directed graph, with each layer fully connected to the next one. Except for the input nodes, each node is a neuron (or processing element) with a nonlinear activation function. MLP utilizes a supervised learning technique called back propagation for training the network. MLP is a modification of the standard linear perceptron and can distinguish data that are not linearly separable. It has the ability to cope with the nonlinearities; the speed of computation, the learning capacity and the accuracy made them valuable tools for Time series prediction [238].

3.5.1.4 IBK

IBK is nearest-neighbour classifier uses the distance metric. The number of nearest neighbours can be specified explicitly in the object editor or determined automatically using leave-one-out cross-validation focus to an upper limit given by the specified value. A kind of different search algorithms can be used to speed up the task of finding the nearest neighbours. A linear search is the default but further options include KD-trees, ball trees, and so-called "cover trees" [239].

3.5.1.5 KStar

K-star algorithm can be defined as a method of cluster analysis which mainly aims at the partition of "n" observation into "k" clusters in which each observation belongs to the cluster with the nearest mean. We can describe K* algorithm as an instance based learner which uses entropy as a distance measure. The benefits are that it provides a consistent approach to handling of real valued attributes, symbolic attributes and missing values. K* is a simple, instance based classifier, similar to K-Nearest Neighbour (K-NN) [239]. New data instances, x, are assigned to the class that occurs most frequently amongst the k-nearest data points, y_j where j = 1, 2…k. Entropic distance is then used to retrieve the most similar instances from the data set. By means of entropic distance as a metric has a number of benefits including handling of real valued attributes and missing values. The K* function can be calculated as:

$$K^*(y_i, x) = -\ln P^*(y_i, x) \quad (5)$$

Where P* is the probability of all transformational paths from instance x to y.

3.5.1.6 Decision Table

This is a precise yet compact way to model complicated logic. Decision tables, like flowcharts and if-then-else and switch-case statements, associate conditions with actions to perform. Each decision corresponds to a variable, relation or predicate whose possible values are listed among the condition alternatives. Each action is a procedure or operation to perform, and the entries specify whether (or in what order) the action is to be performed for the set of condition alternatives the entry corresponds to. Many decision tables include in their condition alternatives the does not care symbol, a hyphen. Using don't cares can simplify decision tables, especially when a given condition has little influence on the actions to be performed. In some cases, entire conditions thought to be important initially are found to be irrelevant when none of the conditions influence which actions are performed.

3.5.1.7 M5 Rules

M5Rules generates a decision list for regression problems using separate-and-conquer. In each iteration, it builds a model tree using M5 and makes the "best" leaf into a rule. The algorithm divides the parameter space into areas (subspaces) and builds in each of them a linear regression model. It is based on M5 algorithm. In each iteration, a M5 Tree is generated and its best rule is extracted according to a given heuristic. The algorithm terminates when all the examples are covered.

3.5.1.8 M5P

M5P a model tree that generated in two stages, the first stage builds an ordinary decision tree, using as splitting criterion the maximization of the intra-subset variation of the target value. The stage second prunes this tree back by replacing subtrees with linear regression functions wherever this seems appropriate. M5rules algorithm produces propositional regression rules in IF-THEN rule format using routines for generating a decision list from M5´Model trees [240]. This model tree is used for numeric prediction and at each leaf it stores a linear regression model that predicts the class value of instances that reach the leaf. In determining which attribute is the best to split the portion T of the training data that reaches a particular node the splitting criterion is used. The standard deviation of the class in T is treated as a measure of the error at that node and calculating the expected reduction in error tests each attribute at that node. The attribute that is chosen for splitting maximizes the expected error reduction at that node. The standard deviation reduction (SDR), which is calculated by (6), is the expected error reduction.

$$SDR = sd(T) - \sum \frac{|T_i|}{|T|} \times sd(T_i) \tag{6}$$

Where Ti corresponds to T1, T2, T3 ... sets that result from splitting the node according to the chosen attribute. The linear regression models at the leaves predict continuous numeric attributes. They are similar to piecewise linear functions and when finally they are combined a non-linear function is formed [241]. The aim is to construct a model that relates a target value of the training cases to the values of their input attributes. The quality of the model will generally be measured by the accuracy with which it predicts the target values of the unseen cases. The splitting process terminates when the standard deviation is only a small fraction less than the standard deviation of the original instance set or when a few instances remain.

In another word we can say that, the algorithm of M5P is based on decision trees, however, instead of having values at tree's nodes, it contains a multivariate linear regression model at each node. The input space is divided into cells using training data and their outcomes, and then a regression model is built in each cell as a leaf of the tree.

3.5.1.9 REPTree

REPTree builds a decision/regression tree using entropy as impurity measure and prunes it. Only sorts values for numeric attributes once [242]. With the help of this method, complexity of decision tree model is decreased by “reduced error pruning method” and the error arising from variance is reduced [10].

Let Y and X be the discrete variables that have the values {y1, …, yn} and {x1, …, xn}. In this case, entropy and conditional entropy of Y are calculated as shown in equation (7) and (8). After that, information gain of X is calculated as shown in equation (9).

$$H(Y) = -\sum_{i=1}^{k} P(Y = y_i) \log P(Y = y_i) \tag{7}$$

$$H(Y \mid \mathrm{X}) = -\sum_{i=1}^{l} P(X = x_i) H(Y \mid \mathrm{X} = \mathrm{x}_i) \tag{8}$$

$$IG(Y; X) = H(Y) - H(Y \mid \mathrm{X}) \tag{9}$$

In decision trees, pruning is done in two ways. These are pre-pruning and post-pruning. If the number of instances that reach a node is lower than the percentage of the training set, that node is not divided. It is considered that variance of the model, which is generated by the training with a small number of instances and accordingly the generalization error will increase. For this reason, if the expansion of the tree is stopped when building the tree, then this is called pre-pruning. Another way of building simple tress is post-pruning. Generally, post-pruning gives better results than pre-pruning in practice [243]. Since the tree does not take steps backward and continues to expand steadily while it is being built, the variance increases. Post-pruning is a way to avoid this situation. In order to do this, firstly, unnecessary sub-trees should be found and pruned.

In post-pruning, the tree is expanded until all the leaves are pure and there is no error in training set. After that, we find the sub-trees that lead to memorizing and prune them. In order to this, we firstly use a major part of training set as growing set and the remaining part as pruning set. Later, we replace each sub-tree with a leaf that is trained by the instances which are covered by the training set of that sub-tree and then we compare these two options on pruning set. If the leaf does not lead to more errors on pruning set, we prune the sub-tree and use the leaf; otherwise we keep the sub-tree [244, 245]. When we compare and contrast pre-pruning and post-pruning, we see that pre-pruning produces faster trees; on the other hand, post-pruning produces more successful trees [243].

3.5.1.10 UserClassifier

User Classifier is special in that it is interactive and lets the user to construct his own decision tree classifier. For the UserClassifier it is best to have numeric attributes because they can be well represented in pixel plots. In the UserClassifier the nodes in the decision tree are not simple tests on attribute values, but are regions the user interactively selects in these plots. So if an instance lies inside the region it follows one branch of the tree, if it lays outside the region it follows the other branch. Therefore each node has only two branches going down from it [246].

3.5.2 Base Meta Classifiers Used

3.5.2.1 Additive Regression

Additive Regression is a kind of algorithm for numerical predictions that can build standard regression model (e.g. tree) and gather residuals, learn model predicting residuals (e.g. tree), and repeat. To predict, it simply sum up individual predictions from all models and also it minimizes squared error of ensemble if base learner minimizes squared error.

Additive regression is another effective ensemble learning method, which uses a set of base learners to achieve greater predictive accuracy. Additive regression implements forward stage wise additive modeling. It starts with an empty ensemble and incorporates new members sequentially. At each stage the model that maximizes the predictive performance of the ensemble as a whole is added, without altering those already in the ensemble. The first regression model for example, a MLP could be used maps the input data to the outputs as usual. Then the residuals between the predicted and observed values are corrected by training a second model e.g., another MLP. Adding the predictions made by the second model to those of the first one yields fewer errors on the training data. The methodology continues with the next model, which learns to predict the residuals of the residuals, and so on [247].

For the additive model Y has been modeled as an additive combination of arbitrary functions of the Xs, which appears in formula (10)

$$Y = A + \sum_{j=1}^{k} f_j(X_j) + \varepsilon \tag{10}$$

Where f_j represent arbitrary functions that can be estimated by lowness or smoothing splines.

Therefore, Additive regression is a form of regression gradient boosting: it enhances performance of basic regression methods [248].

3.5.2.2 Bagging

Bagging is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression. It also reduces variance and helps to avoid over fitting. Although it is usually applied to decision tree methods, it can be used with any type of method. Bagging is a special case of the model averaging approach [59].Bagging is a combination of bootstrapping and averaging used to decrease the variance part of prediction errors [249].

3.5.2.3 MultiScheme

MultiScheme is the simplest technique, as there is no second-level (or meta-) classifier and there is no explicit combination of the individual classifier predictions. MultiScheme, which is also known as Select Best, simply selects a single base-classifier as the predictor for an instance, according to which performed the best on the training data for the instance's class [250].

It selects the best classifier from a set of candidates using cross validation of percentage accuracy (classification) or mean-squared error (regression). The number of folds is a parameter. Performance on training data can be used instead.

3.5.2.4 Random SubSpace

Random subspace Method is combination of random subsets of descriptors and averaging of predictions [251]. Random subspace is an ensemble classifier that consists of several classifiers and outputs the class based on the outputs of these individual classifiers. Random subspace method is a generalization of the random forest algorithm. Whereas random forests are composed of decision trees, a random subspace classifier can be composed from any underlying classifiers. Random subspace method has been used for linear classifiers, support vector machines, nearest neighbours and other types of classifiers. This method is also applicable to one-class classifiers [252].

Each classifier in the ensemble is trained on a random subset of features. The subsets can be intersecting or disjoint. The outputs are aggregated by majority vote. Like Bagging and AdaBoost, the random subspace method is only a 'shell' and can be used with any base classifier. Classifiers that are stable with respect to small changes in the training data may become diverse if trained on different subsets of features [253].

3.5.2.5 Regression by Discretization

Regression by discretization is a meta-learning scheme that applies to regression problems, it based on Random Forest (RD-RF). This is a regression scheme that employs a classifier (random forest, in this case) on a copy of the data which have the property/activity value discretized with equal width. The predicted value is the expected value of the mean class value for each discretized interval (based on the predicted probabilities for each interval) [254].

3.5.2.6 Stacking

Stacking is a meta-classification ensemble introduced by Wolpert [255].The concept of Stacking is to use the predictions of the base-classifiers as attributes in a new training dataset that keeps the original class labels. This new training dataset is learned by a meta-classifier to get the final prediction of the ensemble. Stacking can be viewed as a generalization of Voting [256].

Stacking or stacked generalization is a general method of using the combination of the output from several models in order to achieve a greater predictive accuracy. The final output of the ensemble can be calculated using:

$$y_t = \sum_{K=1}^{N} C_K \hat{Z}_{k,t} + e_t \qquad (11)$$

Where $\hat{Z}_{k,t}$ is output from model K for observation t and the coefficients C_K are estimated in order to construct the final output of the ensemble by minimizing the function G. The function G expressed as:

$$G = \sum_{t=1}^{n} \left[z_t - \sum_{K=1}^{N} C_K \hat{Z}_{k,t} \right]^2 \quad (12)$$

With using constrain $\sum_{K=1}^{N} C_K = 1$ and $0 \leq C_K \leq 1$. In [257] Breiman suggested minimizing the function G that can give better generalization for the model.

In stacking, the result of a set of different base learners at the level-0 is combined by a Meta learner at the level-1. The role of the Meta learner is to discover how best to combine the output of the base learners [258].

3.5.2.7 Vote

Vote is Meta learning scheme, which enables to create an ensemble of multiple base classifies. It provides a baseline method for combining classifiers. The default scheme is to average their probability estimates or numeric predictions, for classification and regression, respectively [250].

3.6 Ensemble Methodology

3.6.1 Combination Methods

There are two main methods for combining the base-classifiers' outputs [239]: weighting methods and meta-learning methods. Weighting methods are useful if the base-classifiers perform the same task and have comparable success. Meta-learning methods are best suited for cases in which certain classifiers consistently correctly classify, or consistently misclassify, certain instances.

a) Weighting methods:

   When combining classifiers with weights, a classifier's classification has strength proportional to its assigned weight. The assigned weight can be fixed or dynamically determined for the specific instance to be classified.

b) Meta-combination methods:

   Meta-learning means learning from the classifiers produced by the inducers and from the classifications of these classifiers on training data. The following sections describe the most well-known meta-combination methods.

   In this research we used vote Meta-combination method to combine the base classifiers.

3.6.2 Structure of Ensemble Classifiers

There are two types for structuring the classifiers of ensembles [15] parallel and Cascading or Hierarchical structure. We used the Parallel Structure of ensemble classifiers. For this kind of structure all the individual classifiers are invoked independently, and their results are fused with a combination rule (e.g., average, weighted voting) or a meta-classifier (e.g., stacked generalization). Figure.3.7 shows the structure of the proposed ensemble classifiers.

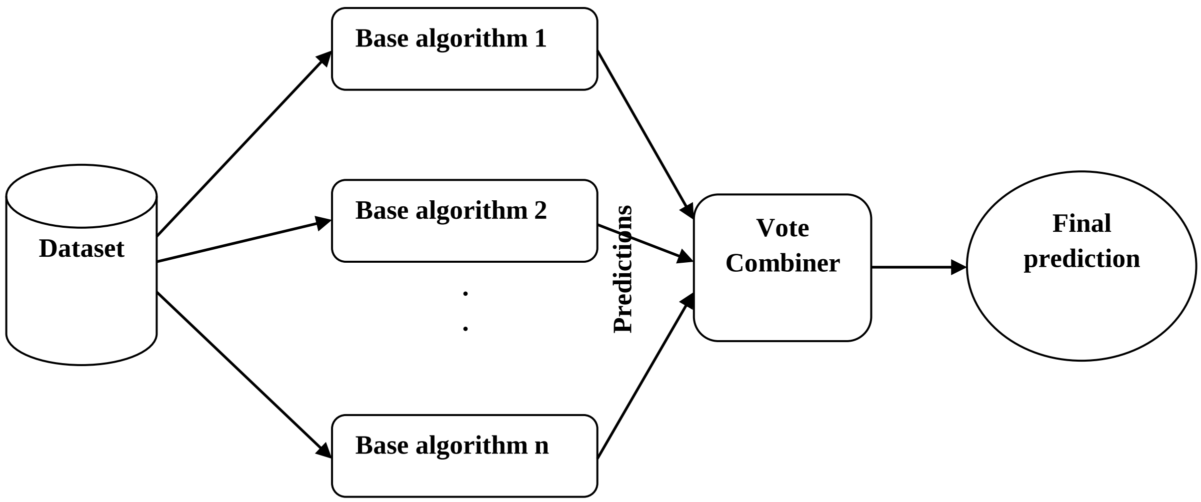


Figure 3.7. The overview of ensemble classifiers framework.

As shown in Figure 3.7, the meteorological dataset are used to train and test the system, each learning algorithm in the system is trained using the training data set, and then give an output. The outputs of all classifiers are combined using median probabilities as combination rule to give the final prediction.

3.6.3 Classifiers Combination Strategy

Combining rules are the simplest combination approach and it is probably the most commonly used in the multiple classifier system [259]. This combination approach is called non-trainable combiner, because combiners are ready to operate as soon as the classifiers are trained and they do not require any further training of the ensemble as a whole [260].

A theoretical framework for fixed rules combination was proposed by Kittler et al. [261] they have discussed many possibilities of combining rule like the sum, product, max, min, average and median rules. In regression problems with vote meta scheme algorithm there are several methods for combination rules such as average of probabilities, minimum probability, maximum probability and median. In this study we have adopted the median probabilities as combination rule method because, it gives the best results for our dataset.

Median Rule

Equation (13) can be used to compute the average a posteriori probability for each prediction over all the classifier outputs, i.e.

assign $Z \rightarrow w_j$ if

$$\frac{1}{R}\sum_{i=1}^{R} P(w_j \mid x_i) = \max_{k=1}^{m} \frac{1}{R}\sum_{i=1}^{R} P(w_k \mid x_i) \tag{13}$$

Where:

Z is the example that has to be predicted.

x_i Is given measurements, i=1,…, R.

R is the number of classifiers.

W_k represent the possible predictions, k= 1,…, m.

Thus, the rule assigns an example to that prediction the average a posteriori probability of which is maximum. If any of the classifiers outputs an a posteriori probability for some prediction which is an outlier, it will affect the average and this in turn could lead to an incorrect decision. It is well known that a robust estimate of the mean is the median. It could therefore be more appropriate to base the combined decision on the median of the a posteriori probabilities. This then leads to the following rule:

assign $Z \rightarrow w_j$ if

$$\operatorname*{med}_{i=1}^{R} P(w_j \mid x_i) = \max_{k=1}^{m} \operatorname*{med}_{i=1}^{R} P(w_k \mid x_i) \tag{14}$$

3.7 Adaptive Neuro-Fuzzy Inference System (ANFIS)

3.7.1 General Framework of the Proposed ANFIS

In order to perform rainfall forecasting using ANFIS and compare the performance criteria with different models, the following steps that shown in Figure 3.8 followed.

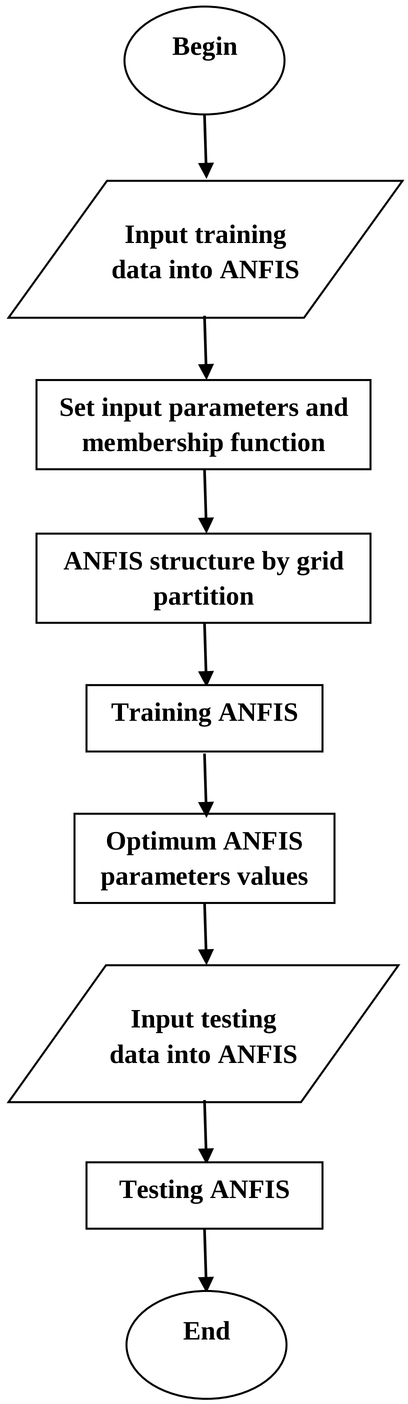


Figure 3.8. The Flowchart for predicting rainfall using ANFIS.

3.7.2 ANFIS Structure

The architecture of Neuro Fuzzy used in this study is ANFIS. Six layered ANFIS Model has been developed with the learning algorithms for training the network are hybridization of forward pass and backward pass using least squares estimate and gradient descent. The total number of nodes for every layer is different for different experiment depending upon the number of membership function of an input variable. In the Figure 3.9, Input 1 (Date) with 3 membership functions (low, medium, high), Input 2 (Average minimum temperature) with 3 membership functions (low, medium, high), Input 3 (Relative Humidity) with 3 membership functions (low, medium, high),Iinput 4 (wind direction) with 3 membership functions (low, medium, high), and a single output as average monthly Precipitation whose degree of membership is Linear.

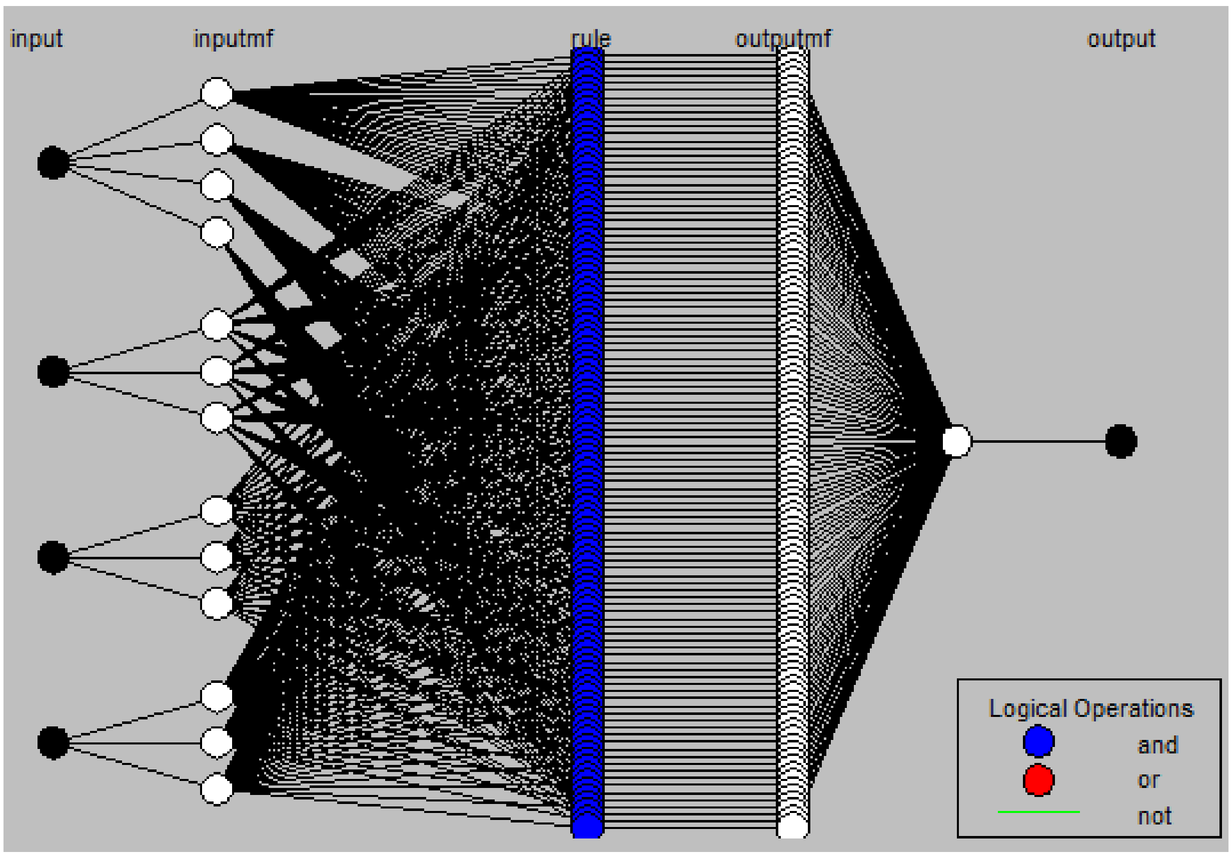


Figure 3.9. Structure of ANFIS used in this Research.

Layer -1(input layer): No computation is done in this layer. Each node, which corresponds to one input variable, only transmits input values to the next layer directly. The link weight in Layer 1 is unity.Layer-2 (fuzzification layer): Each node corresponds to one linguistic label (such as excellent, good) to one of the input variables in Layer 1. In other words, the output link represents the membership value, which specifies the degree to which an input value belongs to a fuzzy set, is calculated in layer 2. The final shapes of the MFs are fine tuned during network learning. Parameters in this layer are referred to as premise parameters. The outputs of this layer are membership values of the premise part. In this study different types of membership functions have been used, these membership functions were:

a) Generalized bell membership function for three different sets of parameters {a, b, c}. This function is defined by (Eq. 14)

$$f(x;a,b,c) = \frac{1}{1 + \left[\frac{x-c}{a}\right]^{2b}} \tag{14}$$

Where {ai, bi, ci} is the parameter set. The parameters a and c represent the width and the center of the bell function, and b represents the slopes at the crossover points. As the values of these parameters change, the bell-shaped function varies accordingly, thus exhibiting various forms of membership functions on linguistic label Ai.

b) Triangular shaped membership function is defined by (Eq. 15)

$$f(x;a,b,c) = \max\left[\min\left[\frac{x-a}{b-a}, \frac{c-x}{c-b}, 0\right]\right] \tag{15}$$

Where a, b and c are parameters of the membership function, a and c set the left and right "feet," or base points, of the triangle. The parameter b sets the location of the triangle peak.

c) Trapezoidal membership function is defined by (Eq. 16)

$$f(x;a,b,c,d) = \max\left[\min\left[\frac{x-a}{b-a}, 1, \frac{d-x}{d-c}, 0\right]\right] \tag{16}$$

Where a, b, c and d are parameters of the membership function, the parameters a and d locate the "feet" of the trapezoid and the parameters b and c locate the shoulders.

d) Gaussian membership function is defined by (Eq. 17)

$$f(x) = \exp^{\frac{-(x-c)^2}{2\sigma^2}} \tag{17}$$

Where σ and c are parameters of the membership function, c is the mean and σ is the variance.

Gaussian2 membership function block implements a membership function based on a combination of two Gaussian functions. The two Gaussian functions are given by (Eq. 18)

$$f_k(x) = \exp^{\frac{-(x-c_k)^2}{2\sigma_k{}^2}} \qquad (18)$$

Where k=1,2. The parameters c1 and σ1 are the mean and variance defining the left-most curve. The parameters c2 and σ2 are the mean and variance defining the right-most curve.

e) Pi membership function this spline-based curve is so named because of its Π shape. The membership function is evaluated at the points determined by the vector x. The parameters a and d locate the "feet" of the curve, while b and c locate its "shoulders." The membership function is a product of smf and zmf membership functions, and is given by (Eq. 19)

$$f\ (x;a,b,c,d) = \begin{cases} 0, x \le a \\ 2\left[\dfrac{x-a}{b-a}\right]^2, a \le x \le \dfrac{a+b}{2} \\ 1-2\left[\dfrac{x-b}{b-a}\right]^2, \dfrac{a+b}{2} \le x \le b \\ 1, b \le x \le c \\ 1-2\left[\dfrac{x-c}{d-c}\right]^2, c \le x \le \dfrac{c+d}{2} \\ 2\left[\dfrac{x-d}{d-c}\right]^2, \dfrac{c+d}{2} \le x \le d \\ 0, x \ge d \end{cases} \qquad (19)$$

f) Dsig membership function composed of difference between two sigmoidal membership functions. The sigmoidal membership function used depends on the two parameters a and c and is given by (Eq. 20)

$$f(x;a,c) = \frac{1}{1+e^{-a(x-c)}} \qquad (20)$$

The membership function dsigmf depends on four parameters, a_1, c_1, a_2, and c_2, and is the difference between two of these sigmoidal functions.

$$f_1(x;\ a_1,\ c_1) - f_2(x;\ a_2,\ c_2) \qquad (21)$$

The parameters are listed in the order: (a_1 c_1 a_2 c_2).

If (Date is low) and (Temperature is high) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is very low)
If (Date is low) and (Temperature is low) and (Humidity is low) and (Wind Direction is low) then (Rainfall is very low)
If (Date is low) and (Temperature is low) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is very low)
If (Date is low) and (Temperature is medium) and (Humidity is low) and (Wind Direction is low) then (Rainfall is low)
If (Date is low) and (Temperature is low) and (Humidity is medium) and (Wind Direction is high) then (Rainfall is low)
If (Date is medium) and (Temperature is low) and (Humidity is medium) and (Wind Direction is high) then (Rainfall is very low)
If (Date is medium) and (Temperature is low) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is low)
If (Date is medium) and (Temperature is low) and (Humidity is high) and (Wind Direction is low) then (Rainfall is very low)
If (Date is medium) and (Temperature is low) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is very low)
If (Date is medium) and (Temperature is medium) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is low)
If (Date is medium) and (Temperature is medium) and (Humidity is medium) and (Wind Direction is high) then (Rainfall is low)
If (Date is medium) and (Temperature is medium) and (Humidity is high) and (Wind Direction is low) then (Rainfall is very low)
If (Date is medium) and (Temperature is medium) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is low)
If (Date is medium) and (Temperature is high) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is medium)
If (Date is medium) and (Temperature is high) and (Humidity is medium) and (Wind Direction is high) then (Rainfall is low)
If (Date is medium) and (Temperature is high) and (Humidity is high) and (Wind Direction is low) then (Rainfall is very low)
If (Date is medium) and (Temperature is high) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is medium)
If (Date is medium) and (Temperature is high) and (Humidity is high) and (Wind Direction is high) then (Rainfall is low)
If (Date is high) and (Temperature is low) and (Humidity is low) and (Wind Direction is medium) then (Rainfall is low)
If (Date is high) and (Temperature is low) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is high)
If (Date is high) and (Temperature is low) and (Humidity is high) and (Wind Direction is low) then (Rainfall is very low)
If (Date is high) and (Temperature is low) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is medium)
If (Date is high) and (Temperature is medium) and (Humidity is low) and (Wind Direction is medium) then (Rainfall is high)
If (Date is high) and (Temperature is medium) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is very high)
If (Date is high) and (Temperature is medium) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is very high)
If (Date is high) and (Temperature is medium) and (Humidity is high) and (Wind Direction is high) then (Rainfall is medium)
If (Date is high) and (Temperature is high) and (Humidity is low) and (Wind Direction is medium) then (Rainfall is high)
If (Date is high) and (Temperature is high) and (Humidity is medium) and (Wind Direction is medium) then (Rainfall is very high)
If (Date is high) and (Temperature is high) and (Humidity is high) and (Wind Direction is medium) then (Rainfall is very high)

g) Psig membership function composed of product of two sigmoidally shaped membership functions. The sigmoid curve plotted for the vector x depends on two parameters a and c as given by (Eq. 20). Psigmf is simply the product of two such curves plotted for the values of the vector x

$$f_1(x; a_1, c_1) \times f_2(x; a_2, c_2) \qquad (22)$$

The parameters are listed in the order $[a_1\ c_1\ a_2\ c_2]$.

Layer-3 (rule antecedent layer): A node represents the antecedent part of a rule. Usually a T-norm operator is used. The output of a Layer 3 node represents the firing strength of the corresponding fuzzy rule.

The fuzzy rule base of the ANFIS model is set up by combining all categories of variables. For example, if there are n inputs and if each input is divided into c categories then there will be c^n rules [262]. For 4 rainfall predictors represented by the inputs date, minimum temperature, humidity and wind direction, having 3 categories namely low, medium, and high, there would be 81 rules in the rule base; the output for each rule is written as a linear combination of input variables. Part of the rule sets can be illustrated as in figure 3.10:

Figure 3.10. Part of ANFIS rules

Each neuron in this layer corresponds to a single Sugeno-type fuzzy rule. A rule neuron receives inputs from the respective fuzzification neurons and calculates the firing strength of the rule it represents. In an ANFIS, the conjunction of the rule antecedents is evaluated by the operator product. Thus, the output of neuron *i* in Layer 3 is obtained as:

$$y_i^{(3)} = \prod_{j=1}^{k} x_{ji}^{(3)} \qquad (23)$$

where $x_{ji}^{(3)}$ are the inputs and $y_i^{(3)}$ is the output of rule neuron *i* in Layer 3.

Layer 4 (rule strength normalization): Every node in this layer calculates the ratio of the *i*-th rule's firing strength to the sum of all rules' firing strength. In other word each neuron in this layer receives inputs from all neurons in the rule layer, and calculates the normalized firing strength of a given rule. The normalized firing strength is the ratio of the firing strength of a given rule to the sum of firing strengths of all rules. It represents the contribution of a given rule to the final result.

$$\overline{w_i} = \frac{w_i}{w_1 + w_2}, i = 1,2,... \qquad (24)$$

Layer-5 (rule consequent layer): Every node *i* in this layer has a node function

$$\overline{w_i} f_i = \overline{w_i}(p_i x_1 + q_i x_2 + r_i) \qquad (25)$$

Where $\overline{w_i}$ is the output of layer 4, and $\{ p_i, q_i, r_i \}$ is the parameter set.

Each neuron in this layer is connected to the respective normalization neuron, and also receives initial inputs, x_1 and x_2. A defuzzification neuron calculates the weighted consequent value of a given rule.

Layer-6 (rule inference layer) the single node in this layer computes the overall output as the summation of all incoming signals:

$$Overalloutput = \sum_i \overline{w_i} f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \qquad (26)$$

3.7.3 Grid Partitioning

For generating the initial fuzzy inference system (FIS), Grid Partitioning has been used. Once the grid partitioning technique is applied at the beginning of training, a uniformly partitioned grid which is defined by membership functions (MFs) with a random set of parameters is taken as the initial state of ANFIS. During training, this grid evolves as the parameters in the MFs change. With the grid partitioning technique, the number of MFs in the premise part of the rules must be determined.

3.7.4 Training of ANFIS Models

Training of ANFIS models has been done using both hybrid optimization method and back propagation algorithm with error tolerance level 0.00001 for 100 epochs.

a) Hybrid learning algorithm combines the least-squares estimator with the gradient descent method.

   1. In the forward pass, a training set of input patterns is presented, neuron outputs are calculated on a layer-by layer basis, and rule consequent parameters are identified by the least-squares estimator.

   2. In the backward pass, the error signals are propagated back (PB) and the rule antecedent parameters are updated according to the chain rule.

b) The back propagation algorithm looks for the minimum of the error function in weight space using the method of gradient descent. The combination of weights which minimizes the error function is considered to be a solution of the learning problem [263].

In this study different choices for dividing the dataset for training and testing have been tried, we used 60% for training and 40% for testing, 70% for training and 30% for testing, 80% for training and 20% for testing and 90% for training and 10% for testing respectively. Once the training is done, ANFIS model is always ready for the prediction. During this training, ANFIS will learn the whole pattern among different input in various years and within each year itself.

3.8 Test Option

The result of applying the chosen classifier will be tested according to a certain test option [264]. There are several test modes such as Use training set, Supplied test set, Cross-validation and Percentage split. In this study, we used supplied test set as the test option, because it provided the best results [12]. In this method the classifier is evaluated on how well it predicts the class of a set of instances loaded from a file. Accordance with that our rainfall dataset, which contains 3732 instances has been divided into two parts with ratio of 70 to 30 for training and testing respectively. The first part contained 2612 examples for training models and the other one contained 1120 examples for testing models.

3.9 Performance Criteria

For evaluating the models performance and comparing between them the following performance metrics have been used:

1) Correlation Coefficient (CC):

This measures the statistical correlation between the predicted and actual values. This method is unique in that it does not change with a scale in values for the test cases [245]. Karl Pearson's correlation coefficient formula is used and it is shown in equation (27).

$$R_{x,y} = \frac{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{\sqrt{\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2}\sqrt{\sum_{i=1}^{n}\left(Y_i - \overline{Y}\right)^2}} \tag{27}$$

Where X_i is the observation data and Y_i computed data and n is the number of data. $\overline{X}$ is the mean of actual data and $\overline{Y}$ is the mean of the computed data.

A higher number means a better model, with a 1 meaning a perfect statistical correlation and a 0 meaning there is no correlation at all.

2) Mean Absolute Error (MAE):

Mean-absolute error is one of the most commonly used measures of success for numeric prediction. This value is computed by taking the average of the differences between each computed value (predicted) and its corresponding correct value (actual) [265]. MAE calculations are shown below in equation (28).

$$MAE = \frac{\left|a_1 - c_1\right| + \cdots + \left|a_n - c_n\right|}{n} \tag{28}$$

Where a is the observation data and c computed data and n is the number of data.

Assuming that the actual output is a, expected output is c. To make regression more robust Minimize absolute error, not squared error.

3) Root mean-squared Error (RMSE):

The Root mean-squared Error is simply the square root of the mean-squared-error. The mean-squared error gives the error value the same dimensionality as the actual and predicted values. Error rate of an estimator arises just because of an arbitrary estimation or lack of information that may provide an accurate estimation [266].RMSE formula is shown in equation (29).

$$RMSE = \sqrt{\frac{(a_1 - c_1)^2 + \cdots + (a_n - c_n)^2}{n}} \quad (29)$$

Where a is the observation data and c computed data and n is the number of data.

If the values of MAE and RMSE rates are closer to zero, the error rates will be lower. In addition, acceptable error values for MSE and RMSE are different for each learning problem.

4) Time

in our study we use the both time taken to build the model and time taken to test model on supplied test set to compare and differentiate among the models.

3.10 GIS and Rainfall Maps

Understanding the spatial distribution of data from phenomena that occur in space constitute today a great challenge to the elucidation of central questions in many areas of knowledge, be it in health, in environment, in geology, in agronomy, among many others. Such studies are becoming more and more common, due to the availability of low cost Geographic Information System (GIS) with user-friendly interfaces. These systems allow the spatial visualization of variables such as individual populations, quality of life indexes or company sales in a region using maps. To achieve that it is enough to have a database and a geographic base (like a map of the municipalities), and the GIS is capable of presenting a colored map that allows the visualization of the spatial pattern of the phenomenon.

Besides the visual perception of the spatial distribution of the phenomenon, it is very useful to translate the existing patterns into objective and measurable considerations

The emphasis of Spatial Analysis is to measure properties and relationships, taking into account the spatial localization of the phenomenon under study in a direct way. That is, the central idea is to incorporate space into the analysis to be made.

Geographic Information System (GIS) is defined as an information system that is used to input, store, retrieve, manipulate, analyze and output geographically referenced data or geospatial data, in order to support decision making for planning and management of land use, natural resources, environment, transportation, urban facilities, health services so on [267]. A GIS is designed for the collection storage, and analysis of objects and phenomena where geographic location is an important characteristic or critical to the analysis. Combination of spatial and attribute data allows users to ask unique spatial questions [268].

Also a geographic information system (GIS) is defined as a computer-based tool for mapping and analyzing geographic phenomenon that exist, and events that occur, on Earth [269]. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies. Map making and geographic analysis are not new, but a GIS performs these tasks faster and with more sophistication than do traditional manual methods.

Spatial analysis the crux of GIS because it includes all of the transformations, manipulations, and methods that can be applied to geographic data to add value to them, to support decisions, and to reveal patterns and anomalies that are not immediately obvious [270]. Spatial analysis is the process, by which we turn raw data into useful information.

The term analytical cartography is sometimes used to refer to methods of analysis that can be applied to maps to make them more useful and informative [271].

The human eye and brain are also very sophisticated processors of geographic data and excellent detectors of patterns and anomalies in maps and images [272]. So the approach taken here is to regard spatial analysis as spread out along a continuum of sophistication, ranging from the simplest types that occur very quickly and intuitively when the eye and brain look at a map, to the types that require complex software and sophisticated mathematical understanding.

Spatial analysis can be [273]:

- Inductive, to examine empirical evidence in the search for patterns that might support new theories or general principles.
- Deductive, focusing on the testing of known theories or principles against data.
- Normative, using spatial analysis to develop or prescribe new or better designs.

There is a range of techniques for visualizing the data recorded by an eye tracker such as simple plots, fixation maps and heat maps [33].

A heat map is a graphical representation of data where the individual values contained in a matrix are represented as colors. Fractal maps and tree maps both often use a similar system of color-coding to represent the values taken by a variable in a hierarchy [34].

Heat maps uses differences in shading or color saturation to encode quantities, which means they are not suited for giving accurate readings, but they allow a lot of data points to be displayed side by side and can be good for spotting patterns and for giving an overview of the data.

Heat maps can simply be colored rectangles or circles in a row or on a grid, but heat mapping can also be combined with other chart types, giving them an extra quantitative axis. A common use case is heat mapping on geographical maps [35].

In the heat maps colors should be used consistently throughout a design; the same colors should not be used to represent different things. When using color to represent categories it's good to think about what people associate with different colors. Using a color the viewer naturally links with the category can reduce their cognitive load [36]. Also keep in mind that some colors are difficult to distinguish for colorblind persons, red and green being the most common.

The number of colors should be kept limited, and the colors used should mainly be soft and muted. Stronger more saturated colors can be used sparingly to highlighting important information. Too many and bright colors makes it hard to focus on the information and bright colors can also affect how the size of areas are perceived [37].

When using colors to represent ordinal or quantitative data such as in a heat map a “perceptually ordered" color scheme should be used. The strongest cue for perceived order is a color's lightness, so a good choice of scale is using a single color hue and gradually altering the lightness, going from light to dark. It's also possible to use a scale with changing hue, as long as the lightness also changes. If the data has a natural midpoint, such as a mean or zero value, a diverging color scheme, with two colors meeting in the middle, can be used.

Using light colors for smaller values and dark colors for larger values is a common cartographic convention [38]. This convention is perhaps confusing when using the common yellow-red color scale for people familiar with black-body radiation and color temperature, since a color gets darker and redder as the temperature decreases.

A rainfall map like heat map, it is a geographical representation of data in which values are represented by colors. Rainfall maps allow users to understand and analyze complex data sets. Rainfall maps make prediction rainfall data more realistic because, they connected rainfall predictions with their geographical locations. They leverage the human visual system to help users gain deeper and faster insights than other visualizations. Users can visually aggregate, determine relevance and detect micro-patterns in their data in ways other visualizations can’t match.

3.10.1 General Framework

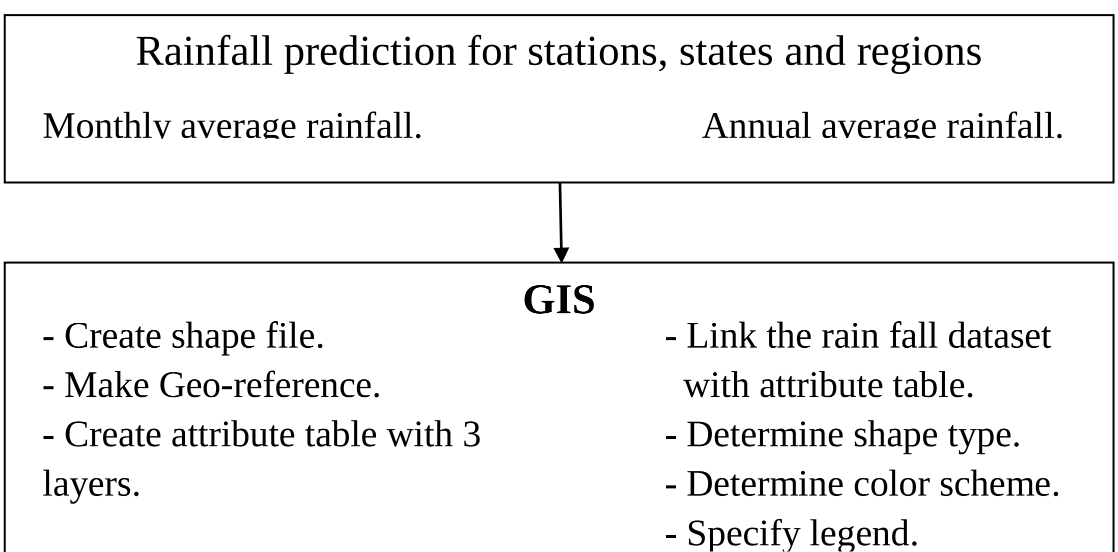
The spatial data for the study area which represented in a form of position and boundaries of regions, states and meteorological stations in towns has been collected by (CBS) using Global Position System (GPS) device of type GPSMAP Garmin60cs with 0-10 meters expected error. Next, we downloaded the spatial data from GPS device into computer device and create a shape file contained three layers for regions, states and meteorological stations in towns.

The polygon has been used to represent regions, states and meteorological stations at town level in the map.

Adjustment process (Georeference) has been made by using GIS to ensure that positions, coordinates and boundaries for towns, states and regions are correct.

After that we created attribute file contained monthly and annual rain dataset which consist of the rainfall predictions and the effect numerical factors that influence on rainfall such as date, minimum temperature and relative humidity.

Finally, advanced GIS software has been used to visualize, compare and analyze the distribution of rainfall and other predictors among town, state and region levels by using different graduated mapping methods such as graduation by size and graduation by color. Figure 3.11 shows the mechanism which has been followed to produce the rainfall maps.



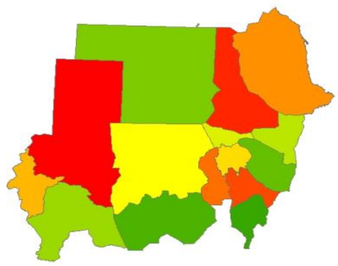


Figure 3.11. The methodology to produce rainfall maps.

3.10.2 Spatial Data

The term spatial data means according to GIS dictionary "the Information about the locations and shapes of geographic features and the relationships between them, usually stored as coordinates and topology". Thus it represents data that are inherently geographic; that is, they describe geographic features.

Also it defined as the data or information that identifies the geographic location of features and boundaries on Earth, such as natural or constructed features, oceans, etc [274]. Spatial data is usually stored as coordinates and topology, and is data that can be mapped. Spatial data is often accessed, manipulated or analyzed through Geographic Information Systems (GIS).

In this study we obtained the coordinates from CBS Sudan. The coordinates such as latitude and longitude, which determined any position on the earth have been taken by using GPS device of type GPSMAP Garmin60cs with 0-10 meters expected error for 24 meteorological stations, 15 states and 6 regions in Sudan. Table 3.5 shows the meteorological stations with their corresponding states and regions.

Table 3.5: The meteorological stations, states and regions.

Regions	States	Stations
Northern	Northern and River Nile.	Dongola, Atbara, Abu Hamad, Karima and Wadi Halfa.
Eastern	Red Sea, Kassala and El Gadarif.	Kassala, Port Sudan, El Gadarif and New Halfa.

Khartoum	Khartoum.	Khartoum.
Central	Al Jazeera, White Nile, Sinnar and Blue Nile.	Wad Medani, El Deweim, Kosti, El damazen, Abu Naama and Sinnar.
Kordofan	North Kordofan and South Kordofan.	Elobied, El Nihood, Kadugli, Babanusa and Rashad.
Darfur	North Darfur, West Darfur and South Darfur.	Nyala, Elgeneina and El Fashir.

3.10.3 Data representation

Real world data are represented by one of four features in a GIS. They include point, line, polygon, and image features [275].

Point feature

A point feature is a discrete location that is usually depicted by a symbol or label.

Line feature

A line feature is a geographic feature that can be represented by a line or set of lines.

Polygon feature

A polygon feature is a multisided figure represented by a closed set of lines.

Image feature

An image feature is a vertical photo taken from a satellite or a plane that is digitized and placed within the geographic information system coordinate system so that there are -x and -y coordinates associated with it.

Each type of feature has “attributes” or a table of data that describe it. All the attributes for three of the four types of features (point, line, and polygon) are stored in a GIS as a data table (Note that a digital orthophotograph has an –x and –y coordinate but does not have an associated data table worthy of analysis). The ability to view, query, relate, and manipulate data behind these features is the true power of a GIS. A manual pin map and a computer map depict points, lines, and polygons but do not have data associated with the features and are not easily manipulated. In a GIS, simply clicking on a point, line, or polygon can produce the data table associated with that particular feature.

In this study the polygon has been used to represent regions, states and meteorological stations at town level in the map.

3.10.4 Data Projection and registration

To georeference means to associate something with locations in physical space. The term is commonly used in the geographic information systems field to describe the process of associating a physical map or raster image of a map with spatial locations. Georeferencing may be applied to any kind of object or structure that can be related to a geographical location, such as points of interest, roads, places, bridges, or buildings [276].

Geographic projections and their parameters (datums, geoids etc) are ways to model the earth's curved surface to a flat plane [276].

Registration is necessary to "tie" geographic data to specific points on the Earth's surface to allow accurate mapping and analysis between different GIS layers

Geospatial data should be geographically referenced (called georeferenced or geocoded) in a common coordinate system. The reference points are called tic masks or ground control points. One of the most convenient way of locating points is to use plane orthogonal coordinates with x (horizontal) and y (vertical) axis.

In this research adjustment process (Georeference) has been made by using GIS to ensure that positions, coordinates and boundaries for towns, states and regions are correct.

3.10.5 Spatial Analysis Techniques

There are two main spatial analysis techniques:

a) **Single Symbol Mapping**

Single symbol mapping refers to the use of individual symbols to represent point, line, and polygon features [277]. The utility of single symbol maps is that they allow for a detailed analysis of small amounts of data.

A drawback of single symbol mapping is that if two incidents have the same address, they are placed exactly on top of one another and cannot be differentiated by looking at the map. However, to avoid confusion since the points may still be placed on top of one another, one should list the number of incidents in the legend. Single symbol maps are more useful for small amounts of data.

b) Graduated Mapping

Graduated mapping consists of aggregating data into groupings that are displayed on the map [277].

These groupings can be graduated by size or by color and can be classified statistically in various ways.

Graduation by Size

Graduated size mapping is the process by which data are summarized so that symbols (point or line features) are altered in size to reflect the frequencies in the data.

In other words, in this type of map, more than one incident at a given point or line is represented with a larger symbol or a thicker line. One drawback is that oftentimes, the size of the symbol or line is difficult to distinguish and the actual value associated with that symbol is not clearly displayed. In addition, similar to single symbol mapping, this type of map is most helpful with smaller amounts of data, since too many incidents make the map unclear and difficult to read.

Graduation by Color

In graduated color mapping, symbols (point, line, or polygon features) are altered in color to reflect a particular value of the feature.

Each location must only represent one value since it would be impossible to shade one point with data of two values.

As with single symbol mapping, this map is most helpful when examining a small number of data.

In this study GIS software has been used to visualize, compare and analyze the distribution of rainfall and other predictors among town, state and region levels by using different graduated mapping methods such as graduation by size and graduation by color.

3.10.6 Classification Methods

There are several different statistical methods for classifying numeric data in a GIS when creating both graduated size and graduated color maps. In this study the natural breaks has been used as classification method.

Natural breaks

This is the default classification in most GIS programs and identifies the natural break points within the data using a statistical formula [278].

The software examines the selected data and their distribution, identifies natural break points, and creates the categories based on the best fit to the data. With each data set, the natural breaks classification would result in different ranges of categories; thus, the classification is data dependent.

3.11 Hardware and Software Requirements

In this research, several hardware and software requirements were employed in order to experiment on the proposed algorithms. A personal computer with Intel ® Pentium CPU B960, 2.2GHz, 2.00GB memory and 300GB hard drive was used, which ran on Microsoft Windows 7 Ultimate.

Microsoft ®Excel® 2007 (12.0.4518.1014) was used for statistical analysis (calculating minimum, maximum, average and standard deviation) at the stage of dataset analysis. SPSS 16 for Windows © 2007 was installed and used to visualize and represent graphically the numerical attributes (minimum temperature, maximum temperature, relative humidity, wind speed and rainfall) in meteorological dataset.

Waikato Environment for Knowledge Analysis (WEKA) Version 3.6.9 and Version 3.7.10 © 1999-2013 were utilized to model the three groups of machine learning approaches for rain forecasting, including basic, Meta and ensemble algorithms and compared these models performance.

Moreover, MATLAB® Version 7.8.0.347 (R2009a) 32-bit (win32) was explored to develop the Adaptive Neuro-Fuzzy Inference System (ANFIS) model for long term rainfall prediction.

Finally, ESRI®Arc-GIS 9.3was used to create the rainfall maps and performed different types of spatial analysis.

3.12 Summary

In this chapter the meteorological data used in this research is presented and analyzed. The dataset were obtained from central Bureau of Statistics, Sudan. Feature selection was performed to determine the most influencing and important variables that affect on the long term rainfall prediction out of the existing one.

Secondly, it presented many machine learning techniques for long term rainfall prediction such as ANFIS, base algorithms (Gaussian Processes, Linear Regression, Multilayer Perceptron, IBk, KStar, Decision Table, M5Rules, M5P, REP Tree and User Classifier), Meta algorithms(Additive Regression, Bagging, Multi Scheme, Random Subset, Regression by Discretization, Stacking, and Vote) and the proposed ensemble method which has been constructed based of Meta classifier Vote combining with three base classifiers IBK, K-star and M5P.

Next, the performance measurements of each proposed schemes are presented. At the end of this chapter, the hardware and software requirements to develop and implement the proposed models framework are discussed. In the next chapters, the results and discussion of ANFIS, basic, Meta and ensemble models to optimize long term rainfall prediction will be presented.

Finally, concepts such as GIS and rainfall maps were defined. Then our general framework for producing rainfall maps has been introduced.

4. Results of Features Selection and Rainfall Prediction Models

4.1 Introduction

This Chapter presents the results of attribute selection process to determine the most influencing variables that affect on the long-term rainfall prediction in Sudan out of the 7 available variables. Different attributes evaluators such as (Correlation based Feature Selection subset evaluator, Classifier subset evaluator, Relief attribute evaluator and Wrapper subset evaluator) has been applied with their appropriate search methods such as (Best-first, Evolutionary Search, Exhaustive Search, Genetic Search, Greedy Stepwise, Linear Forward Selection, PSO Search, Random Search, Scatter SearchV1, Subset Size Forward Selection, Tabu Search and Ranker).The number and order of the selected attributes has been determined for every attribute evaluator with its own search method.

The rest of this Chapter is organized as follows: Section 4.2 compares between the experimental results of the 10 base algorithms (Gaussian Processes, Linear Regression, Multilayer Perceptron, IBK, Kstar, Decision Table, M5Rules, M5P, REPTree and User Classifier) according to several performance evaluation criteria such as correlation coefficient, mean absolute error, root mean squared error time taken to build model and time taken to test model on supplied test set. Comparison process designed to determine the best models to predict the rainfall among basic models.

Section 4.3 compares between the experimental results of the individual Meta algorithms (Additive Regression, Bagging, Multi_scheme, Random SubSpace, Regression by Discretization, Staking and Vote) according to the same performance evaluation criteria: correlation coefficient, mean absolute error, root mean squared error time taken to build model and time taken to test model on supplied test set. Comparison process designed to determine the best models to predict the rainfall among individual Meta models.

4.2 Attribute Selection Results

Table 4.1 shows the different attributes evaluators with their appropriate search methods beside the number and order of the selected attributes.

Table 4.1: The results of attributes selection.

Attributes evaluator	Search method	No of selected Attributes	Selected attributes
Correlation based Feature Selection subset evaluator	Best-first	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Evolutionary Search	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Exhaustive Search	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Genetic Search	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Greedy Stepwise	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Linear Forward Selection	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	PSO Search	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Random Search	3	Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Scatter SearchV1	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Subset Size Forward Selection	4	Date, Min-T, Humidity, Wind D.
Correlation based Feature Selection subset evaluator	Tabu Search	4	Date, Min-T, Humidity, Wind D.

Classifier subset evaluator	Genetic Search	1	Wind D.
Classifier subset evaluator	Random Search	4	Date, Min-T, Humidity, Wind D.
Relief attribute evaluator	Ranker	7	station, Wind D, Date, Humidity, Min-T, Max-T, Wind S.
Wrapper subset evaluator	Genetic Search	1	Wind D.
Wrapper subset evaluator	Random Search	4	Date, Min-T, Humidity, Wind D.

As shown in Table 4.1, the result of applying different attributes evaluators with different search methods, we have obtained 4 choices of dataset's attributes (1, 3, 4 and 7), the dominant choice was four attributes (Date, Minimum Temperature, Humidity and Wind Direction).

In this study, only the most influencing variables (Date, Minimum Temperature, Humidity and Wind Direction) that affect on the long term rainfall prediction out of the 7 variables [12] have been used.

4.3 Experimental Results for the Base Algorithms

Table 4.2 shows the performance of the base classifier models according to correlation coefficient, mean absolute error and root mean squared error, while Table 4.3 displays time taken to build model and time taken to test model on supplied test set for the base algorithm.

Table 4.2: The performance of the base algorithms.

Base algorithm	CC	MAE	RMSE
Gaussian Processes (GP)	0.8656	0.1638	0.2512
Linear Regression (LR)	0.8642	0.1643	0.2527
Multilayer Perceptron (MLP)	0.8594	0.1327	0.2654
IBK	0.8192	0.0905	0.3005
KStar	0.8901	0.1091	0.2285
Decision Table (DT)	0.8351	0.1219	0.2775
M5Rules(M5R)	0.8642	0.1113	0.2529
M5P	0.8863	0.1047	0.2322
REPTree(RT)	0.8262	0.1286	0.2841
User Classifier (UC)	0.8801	0.2352	0.32

According to the Experimental results in Table 4.2, we find that KStar algorithm has the maximum correlation coefficient 0.8901, the minimum root mean squared error 0.2285 and the third lower mean absolute error 0.1091. M5P algorithm comes in second place after KStar as the second highest correlation coefficient 0.8863; the second less mean absolute error 0.1047 and second less root mean squared error 0.2322. User Classifier algorithm comes in third place in terms of the standard correlation coefficient 0.8801, but it’s the worst on both levels of mean absolute error 0.2352 and root mean squared error 0.32.IBK algorithm has the minimum mean absolute error 0.0905, but at the same time it has the second biggest root mean squared error 0.3005 and unsatisfactory correlation coefficient 0.8192 compared with the other base algorithms.

Table 4.3: The Base algorithms training and testing time.

Base algorithm	Training Time (Sec)	Testing Time (Sec)
Gaussian Processes (GP)	87.2	1.97
Linear Regression (LR)	0.1	0.01
Multilayer Perceptron (MLP)	27.04	0.02
IBK	0.02	0.34
KStar	0.01	4.59
Decision Table (DT)	0.1	0.02
M5Rules (M5R)	0.3	0.02
M5P	0.1	0.01
REPTree (RT)	0.5	0.01
User Classifier (UC)	0.6	0.01

Figure 4.1 compared between the base algorithms in terms of correlation coefficient, mean absolute error and root mean squared error. We can observe that the most accurate base algorithm in the term of correlation coefficient and root mean squared error is Kstar, while IBK algorithm has the lowest mean absolute error.

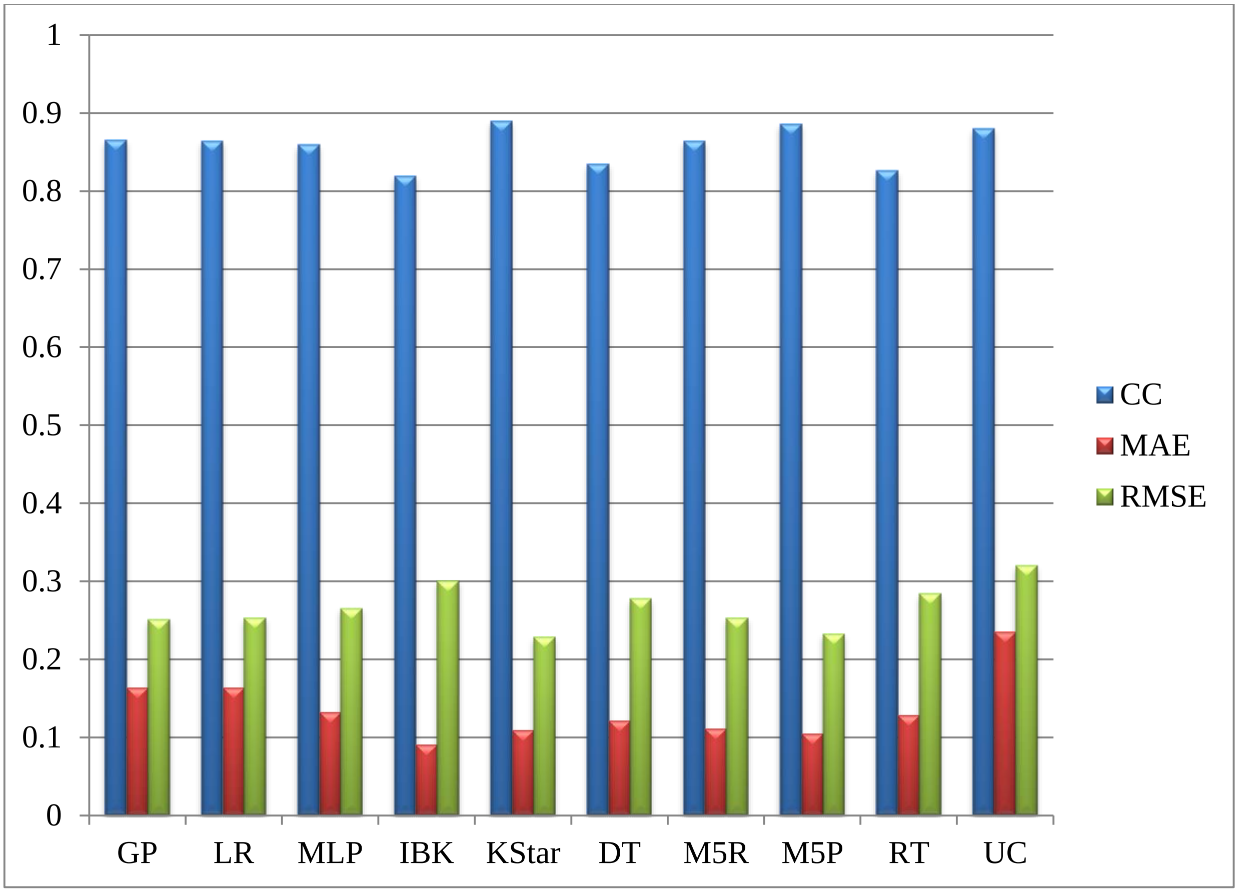


Figure 4.1. Performance comparison between the base algorithms.

Figures 4.2 and 4.3 compare between the base algorithms in terms of time taken to build model on supplied test set test option. We can easily find that the longest time to build a model is 87.2seconds belong to Gaussian Processes (GP) followed by Multilayer Perceptron (MLP) with 27.04 sec, while the shortest times are 0.01and 0.02 second belong to Kstar and IBK respectively.

Figures 4.4 and 4.5 compare between the base algorithms in terms of time taken to test model on supplied test set test option test set, we conclude that the longest test time 4.59 sec belong to Kstar followed by Gaussian Processes and IBK with testing time 1.97 sec and 0.34 sec respectively. The shortest test time is 0.01 sec belonging to Linear Regression, M5P, REPTree and User Classifier. Multilayer Perceptron, Decision Table and M5Rules come in the second order with 0.02 sec as testing time for each one.

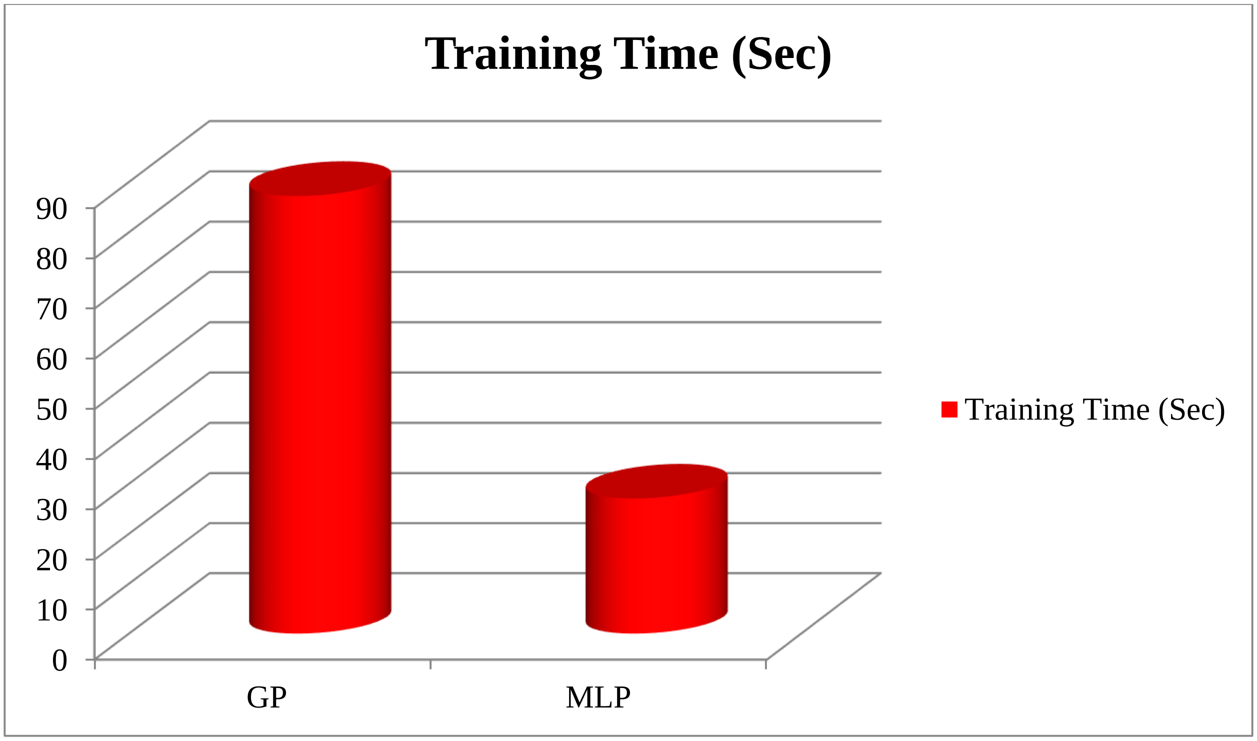


Figure 4.2. Comparison between Gaussian Processes and Multilayer Perceptron according to training time.

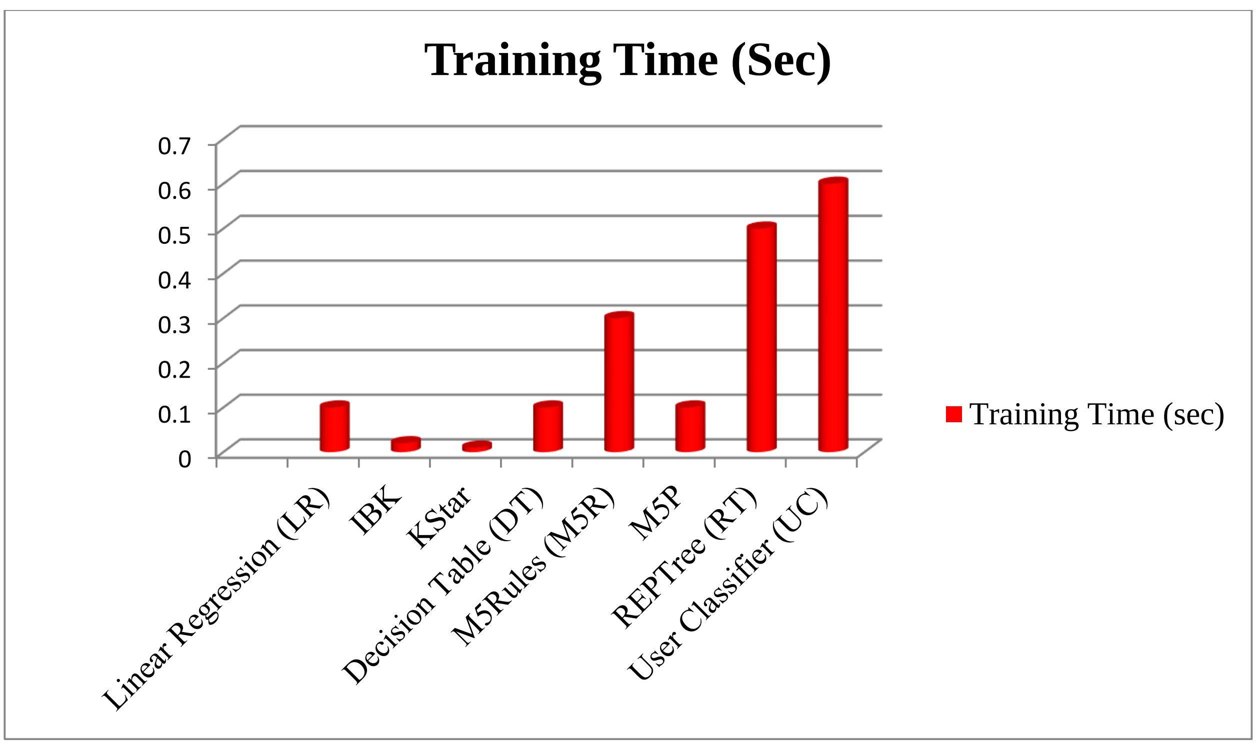


Figure 4.3. Comparison between the rest of the base algorithms according to training time.

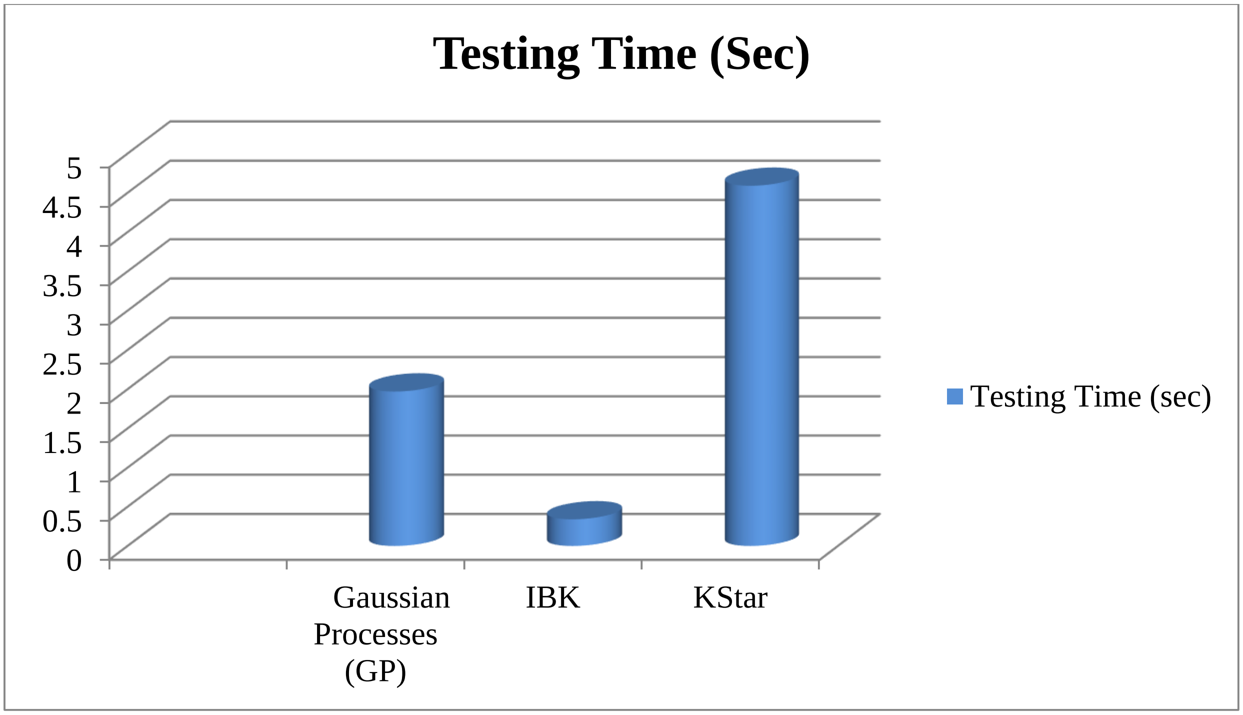


Figure 4.4. Comparison between Gaussian Processes IBK and K-Star according to test time.

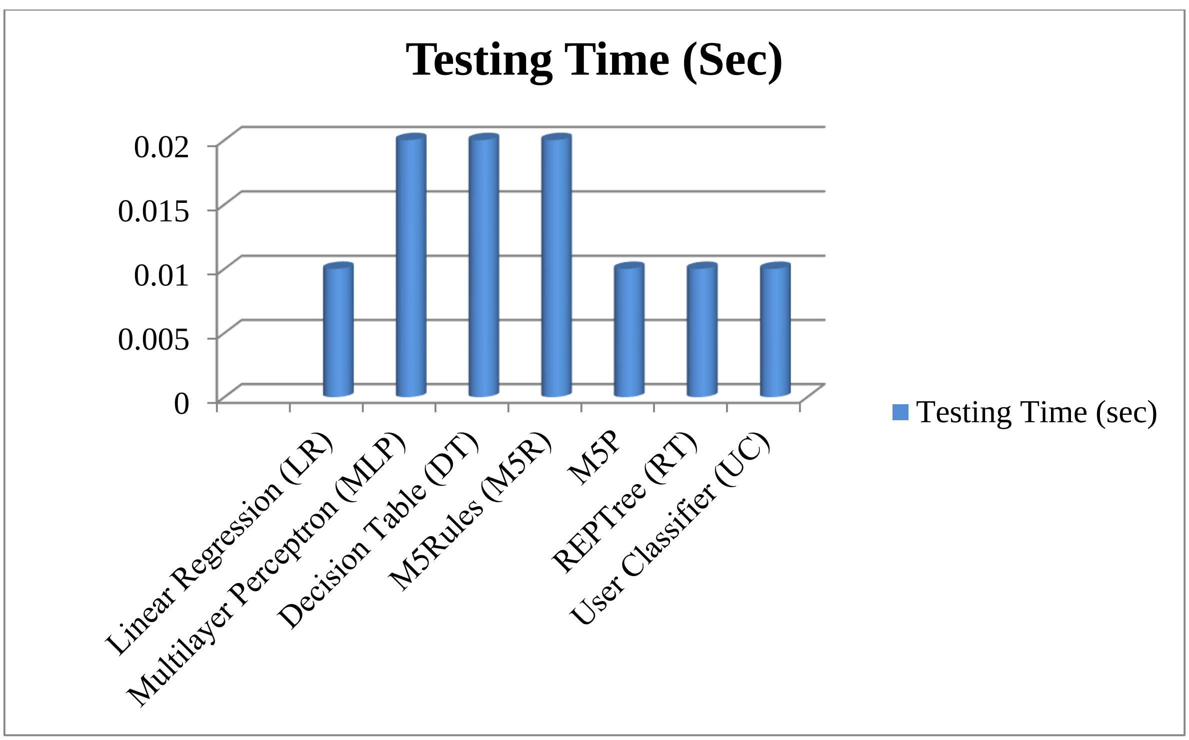


Figure 4.5. Comparison between the rest of the base algorithms according to testing time.

4.4 Experimental Results for the Meta Algorithms

Table 4.4 shows the performance of the individual Meta classifiers according to correlation coefficient, mean absolute error, root mean squared error. The best correlation coefficient 0.8651 belongs to Random SubSpace Meta classifier, bagging Comes second has 0.8529 correlation coefficient and vote Meta classifier in third place has 0.8463 correlation coefficient. The worst correlation coefficient 0 resulted from both multi scheme and staking Meta methods. In terms of the Mean absolute error Bagging and vote have the lowest 0.1237 and 0.1287 respectively, while multi scheme and staking have the highest 0.4783 and 0.504 respectively. Most of the tested Meta methods their Root mean squared error in the range of 0.2563-0.2965 except multi scheme and staking these have 0.4892 and 0.5273 respectively.

Table 4.4: Performance of the individual Meta algorithms.

Meta Classifier	**CC**	**MAE**	**RMSE**
Additive Regression	0.8052	0.196	0.2965
Bagging	0.8529	0.1237	0.2614
Multi scheme	0	0.4783	0.4892
Random SubSpace	0.8651	0.1636	0.2563
Regression by Discretization	0.8154	0.1397	0.2914
Staking	0	0.504	0.5273
Vote	0.8463	0.1287	0.267

Table 4.5 Shows time taken to build model and time taken to test model on supplied test set for the individual Meta methods.

Table 4.5: Individual Meta algorithms training and testing time.

Meta Classifier	**Training Time (Sec)**	**Testing Time (Sec)**
Additive Regression	0.08	0.01
Bagging	0.16	0.01
Multi scheme	0.03	0.01
Random SubSpace	0.11	0.02
Regression by Discretization	0.15	0.03
Staking	0.02	0.01
Vote	0.01	0.03

Figure 4.6 compared between the Meta algorithms in terms of correlation coefficient, mean absolute error and root mean squared error. We can infer that Random Sub-Space is the best Meta method in both Correlation coefficient 0.8651and Root mean squared error 0.2563, while Bagging has the lowest Mean absolute error 0.1237. On the other hand multi scheme and staking Meta models deal very bad with our data and give the worst results 0 Correlation coefficient, and the highest Mean absolute error and Root mean squared error comparing with other proposed Meta methods.

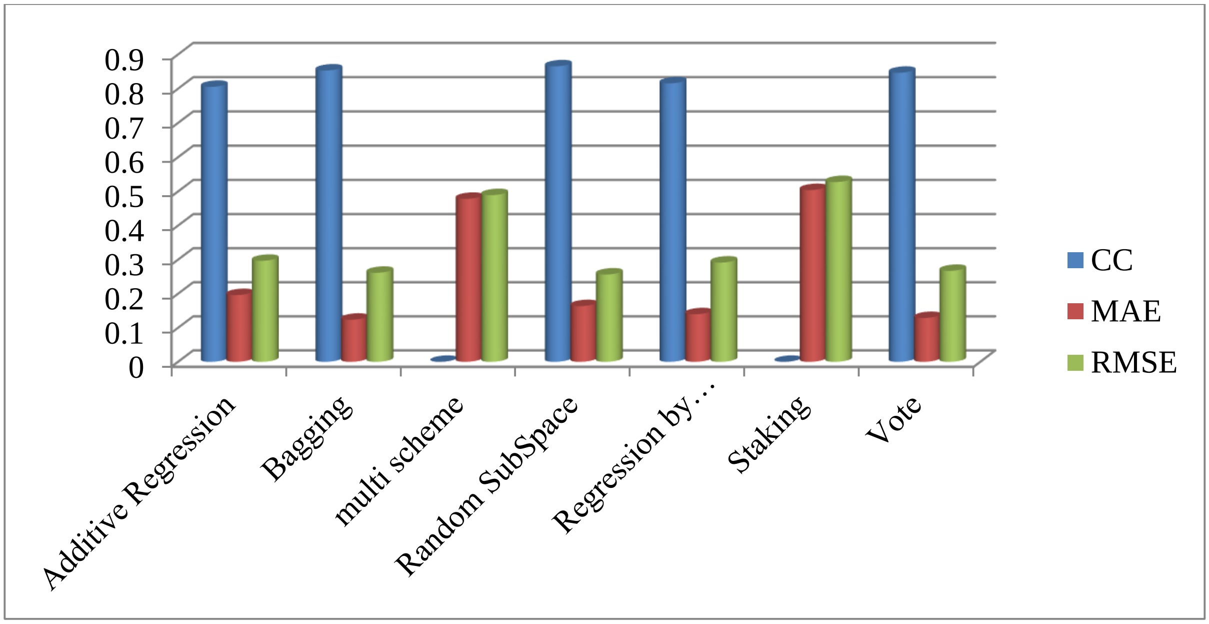


Figure 4.6. Performance Comparison between the Meta algorithms.

Figure 4.7 Compared between the Meta algorithms in terms of time taken to build model and time taken to test model on supplied test set. We can note that the longest time to build individual meta model is 0.16seconds belong to Bagging, while the shortest time are 0.01second belong to Vote. If we look to the time taken to test model on supplied test set we conclude that the longest test time 0.03belong to both Vote and Regression by Discretization, while the shortest test time 0.01 belong to Additive Regression, Bagging and multi scheme.

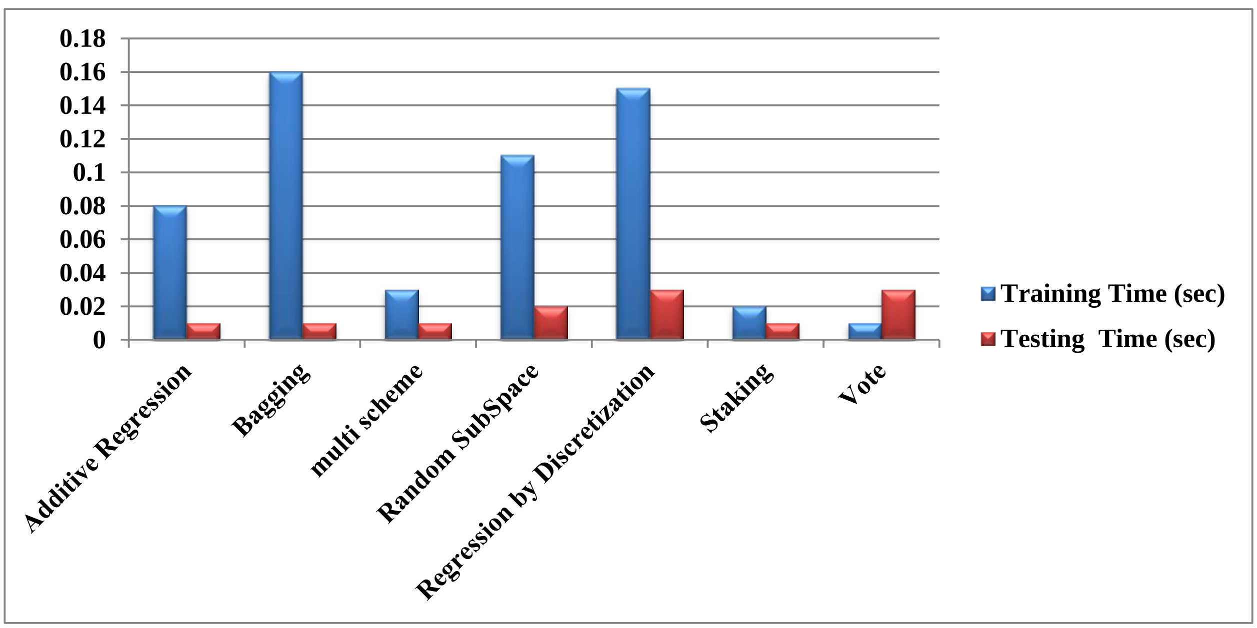


Figure 4.7. Comparison between the Meta algorithms by training and testing times.

4.5 Summary

This chapter presented the Results of features selection, base algorithms and Meta algorithms. The results led us to:

Applying different attributes evaluators with different search methods resulted 4 choices of dataset's attributes (1, 3, 4 and 7); the dominant choice was four attributes (Date, Minimum Temperature, Humidity and Wind Direction). So we concluded that the most influencing variables are (Date, Minimum Temperature, Humidity and Wind Direction) that affect on the long term rainfall prediction out of the 7 variables.

K-Star algorithm has the maximum correlation coefficient 0.8901, the minimum root mean squared error 0.2285 and the third lower mean absolute error 0.1091. M5P algorithm comes in second place after K-Star as the second highest correlation coefficient 0.8863; the second less mean absolute error 0.1047 and second less root mean squared error 0.2322. User Classifier algorithm came in third place in terms of the standard correlation coefficient 0.8801, but it's the worst on both levels of mean absolute error 0.2352 and root mean squared error 0.32.IBK algorithm has the minimum mean absolute error 0.0905, but at the same time it has the second biggest root mean squared error 0.3005 and unsatisfactory correlation coefficient 0.8192 compared with the other base algorithms. Thus the most accurate base algorithm in the term of correlation coefficient and root mean squared error is Kstar, while IBK algorithm has the lowest mean absolute error.

Comparison between the base algorithms in terms of time taken to build model and time taken to test model on supplied test set test option implied that the longest time to build a model is 87.2seconds belong to Gaussian Processes (GP), while the shortest times are 0.01and 0.02 second belong to Kstar and IBK respectively. If we look to the time taken to test model on supplied test set we concluded that the longest test time 4.59 belong to Kstar and the shortest test time 0.01 belong to Linear Regression, M5P, REPTree and User Classifier.

The performance of the individual Meta classifiers according to correlation coefficient, mean absolute error, root mean squared error. The best correlation coefficient 0.8651 belongs to Random SubSpace Meta classifier, bagging Comes second has 0.8529 correlation coefficient and vote Meta classifier in third place has 0.8463 correlation coefficient. The worst correlation coefficient 0 resulted from both multi scheme and staking Meta methods. In terms of the Mean absolute error Bagging and vote have the lowest 0.1237 and 0.1287 respectively, while multi scheme and staking have the highest 0.4783 and 0.504 respectively. Most of the tested Meta methods their Root mean squared error in the range of 0.2563-0.2965 except multi scheme and staking these have 0.4892 and 0.5273 respectively. We inferred that Random Sub-Space is the best Meta method in both Correlation coefficient and Root mean squared error, while Bagging has the lowest Mean absolute error. On the other hand multi scheme and staking Meta models deal very bad with our data and give the worst results 0 Correlation coefficient, and the highest Mean absolute error and Root mean squared error comparing with other proposed Meta methods.

From comparison between the Meta algorithms in terms of time taken to build model and time taken to test model on supplied test set. We concluded that the longest time to build individual meta model is 0.16seconds belong to Bagging, while the shortest time are 0.01second belong to Vote. If we look to the time taken to test model on supplied test set we conclude that the longest test time 0.03belong to both Vote and Regression by Discretization, while the shortest test time 0.01 belong to Additive Regression, Bagging and multi scheme.

5. Ensemble and ANFIS Results

5.1 Introduction

This Chapter presents the results of the proposed ensemble algorithm and Adaptive Nero-fuzzy Inference System (ANFIS) model to predict long-term rainfall in Sudan. The chapter is organized as follows:

Section 5.2 shows the results of the proposed ensemble models of Meta Vote method combining with various base classifiers.Vote+2 algorithms (Kstar and linear regression), Vote+3 algorithms (IBK, Kstar and M5P), Vote+4 algorithms (IBk, Kstar, M5P, and linear regression), Vote+5 algorithms (IBk, Kstar, M5P, REPTree and User Classifier), Vote+6 algorithms (Linear Regression, IBk, Kstar, M5P, REPTree and User Classifier), Vote+7 algorithms (Linear Regression, IBk, Kstar, M5Rules, M5P, REPTree and User Classifier), Vote+8 algorithms (Linear Regression, IBk, Kstar, Decision Table, M5Rules, M5P, REPTree and User Classifier),Vote+9 algorithms (Linear Regression, Multilayer Perceptron, IBk, Kstar, Decision Table, M5Rules, M5P, REPTree and User Classifier). The performance of the Ensemble models has been compared according to correlation coefficient, mean absolute error, root mean squared error, time taken to build model and time taken to test model.

Section 5.3 compares between the experimental results of several ANFIS models using different types of membership functions, different optimization methods and different dataset ratios for training and testing. The proposed models have been evaluated and compared by using correlation coefficient, mean absolute error and root mean-squared error as performance metrics.

Section 5.4 compares between the experimental results of the proposed ensemble Vote+3 algorithms with its base algorithms. Also it shows comparison between the proposed ensemble Vote+3 algorithms and the best ANFIS model. Finally it appears the results of the performance measure criteria of our proposed ensemble Vote+3 algorithms and ANFIS comparing with other related models form literature. Comparison processes designed to determine the best models to predict the rainfall among different models.

5.2 Experimental Results for the Ensemble Algorithms

Table 5.1 shows the performance of the Ensemble models according to correlation coefficient, mean absolute error, root mean squared error. We construct ensemble model of Meta Vote method combining with various base classifiers.Vote+2 algorithms (Kstar and linear regression), Vote+3 algorithms (IBK, Kstar and M5P), Vote+4 algorithms (IBk, Kstar, M5P, and linear regression), Vote+5 algorithms (IBk, Kstar, M5P, REPTree and User Classifier), Vote+6 algorithms (Linear Regression, IBk, Kstar, M5P, REPTree and User Classifier), Vote+7 algorithms (Linear Regression, IBk, Kstar, M5Rules, M5P, REPTree and User Classifier), Vote+8 algorithms (Linear Regression, IBk, Kstar, Decision Table, M5Rules, M5P, REPTree and User Classifier),Vote+9 algorithms (Linear Regression, Multilayer Perceptron, IBk, Kstar, Decision Table, M5Rules, M5P, REPTree and User Classifier).

Table 5.1: Ensemble methods training and testing time.

Ensemble method	**CC**	**MAE**	**RMSE**
Vote+2 algorithms	0.8861	0.1311	0.2319
Vote+3 algorithms	**0.8986**	**0.0888**	**0.1092**
Vote+4 algorithms	0.8803	0.1376	0.2728
Vote+5 algorithms	0.884	0.1328	0.2379
Vote+6 algorithms	0.8753	0.1235	0.2418
Vote+7 algorithms	0.8852	0.1378	0.2375
Vote+8 algorithms	0.8835	0.1327	0.2383
Vote+9 algorithms	0.8832	0.1369	0.2386

According to experimental results in Table 5.1, we can conclude that the proposed ensemble method achieved the best performance overall other ensemble methods. Ensemble Vote+3 algorithm has the highest correlation coefficient 0.8986, the lowest of both mean absolute error and root mean squared error 0.0888 and 0.1092 respectively. Ensemble Vote+2 algorithms come second in terms of both correlation coefficient 0.8861and root mean squared error 0.2319 but it has 0.1311mean absolute error. From other point of view Ensemble Vote+6 algorithms comes second in terms of mean absolute error 0.1235 however it has the worst correlation coefficient 0.8753 and second worst root mean squared error 0.2418.

Table 5.2 Shows the time taken to build model and time taken to test model on supplied test set for ensemble methods.

Table 5.2: Ensemble methods training and testing time.

Ensemble method	Training Time (sec)	Testing Time (sec)
Vote+2 algorithms	0.08	4.44
Vote+3 algorithms	0.09	4.71
Vote+4 algorithms	0.78	4.77
Vote+5 algorithms	3.95	4.83
Vote+6 algorithms	8.41	4.87
Vote+7 algorithms	34.81	4.54
Vote+8 algorithms	35.17	4.81
Vote+9 algorithms	67.42	4.34

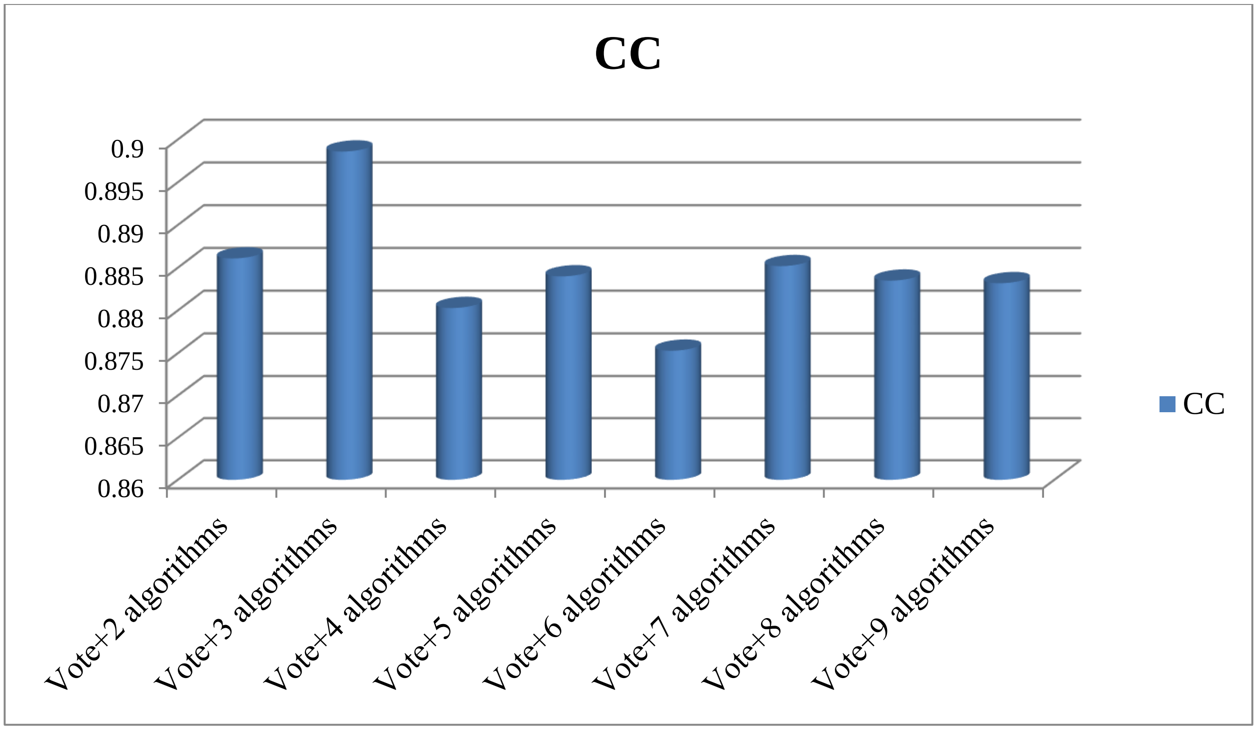


Figure 5.1. Comparison between the Ensemble models according to correlation coefficient.

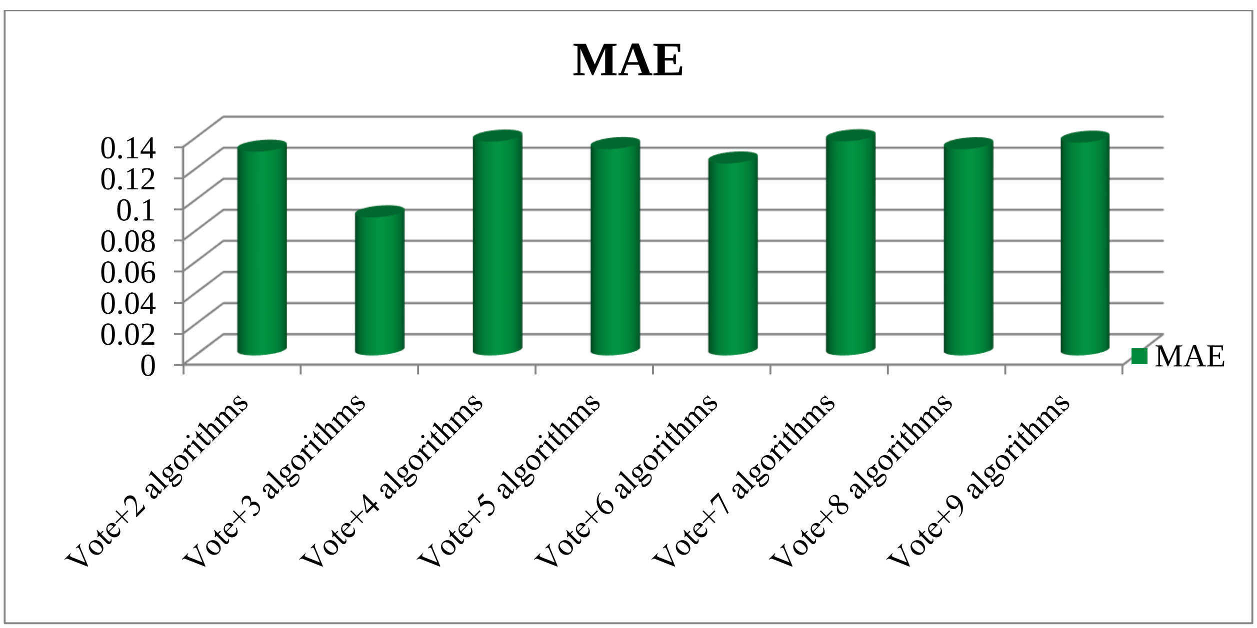


Figure 5.2. Comparison between the Ensemble models according to mean absolute error.

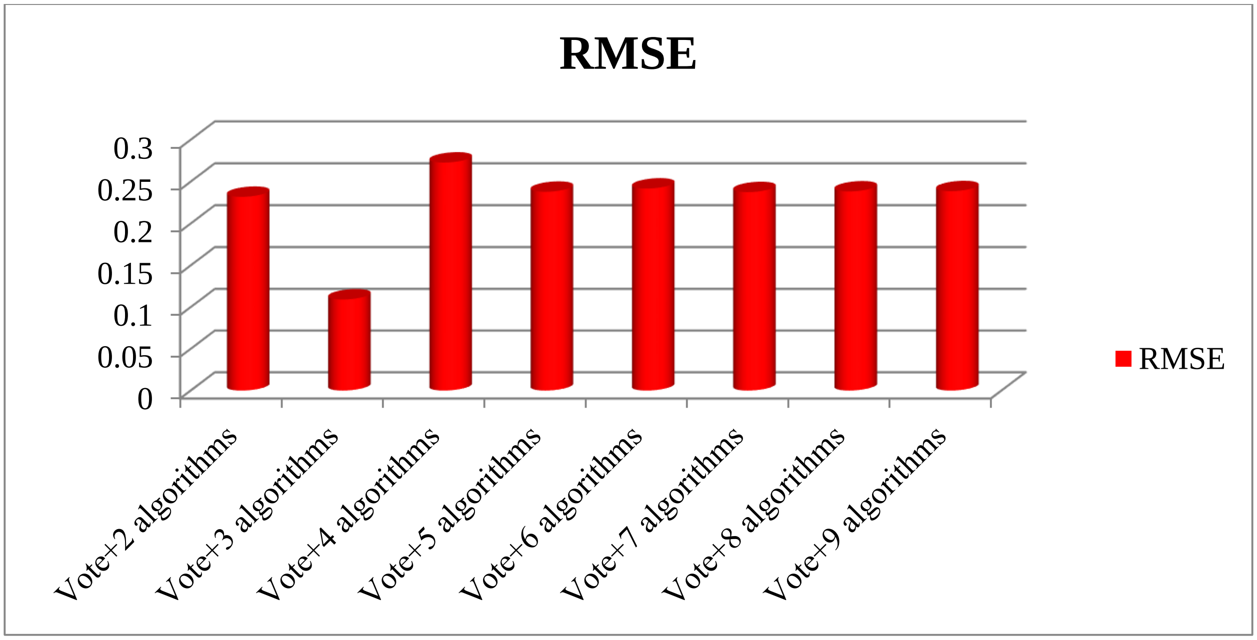


Figure 5.3. Comparison between the Ensemble models according to root mean squared error.

Figures 5.1-5.3 compared between the Ensemble models in terms of correlation coefficient, mean absolute error and root mean squared error respectively. As shown in Figures above, we found that:

The results of the proposed ensemble methods performance were very close, but also we noticed ensemble Vote+3 algorithms outperformed the rest of the other ensemble methods, because it had the highest correlation coefficient (0.8986) and the lowest of both mean absolute error (0.0888) and root mean squared error (0.1092).

The minimum correlation coefficient (0.8753) resulting from the ensemble Vote+6 algorithms, followed by ensemble Vote+4 algorithms (0.8803) Which is considered the second worst correlation coefficient.

From other side the ensemble Vote+7 algorithms recorded the highest mean absolute error (0.1378) followed by the ensemble Vote+4 algorithms (0.1376) and the ensemble Vote+9 algorithms (0.1369).

In the term of the root mean squared error, the ensemble Vote+4 algorithms considered the worst with (0.2728) root mean squared error followed by the ensemble Vote+6 algorithms (0.2418) and the ensemble Vote+9 algorithms (0.2386).

While the ensemble Vote+3 algorithms recorded the best results according to the three performance measurement criteria (correlation coefficient, mean absolute error and root mean squared error) and outperformed others ensemble methods. Ensemble Vote+6 algorithms, ensemble Vote+7 algorithms, ensemble Vote+4 algorithms and ensemble Vote+9 algorithms considered the worst in terms of results.

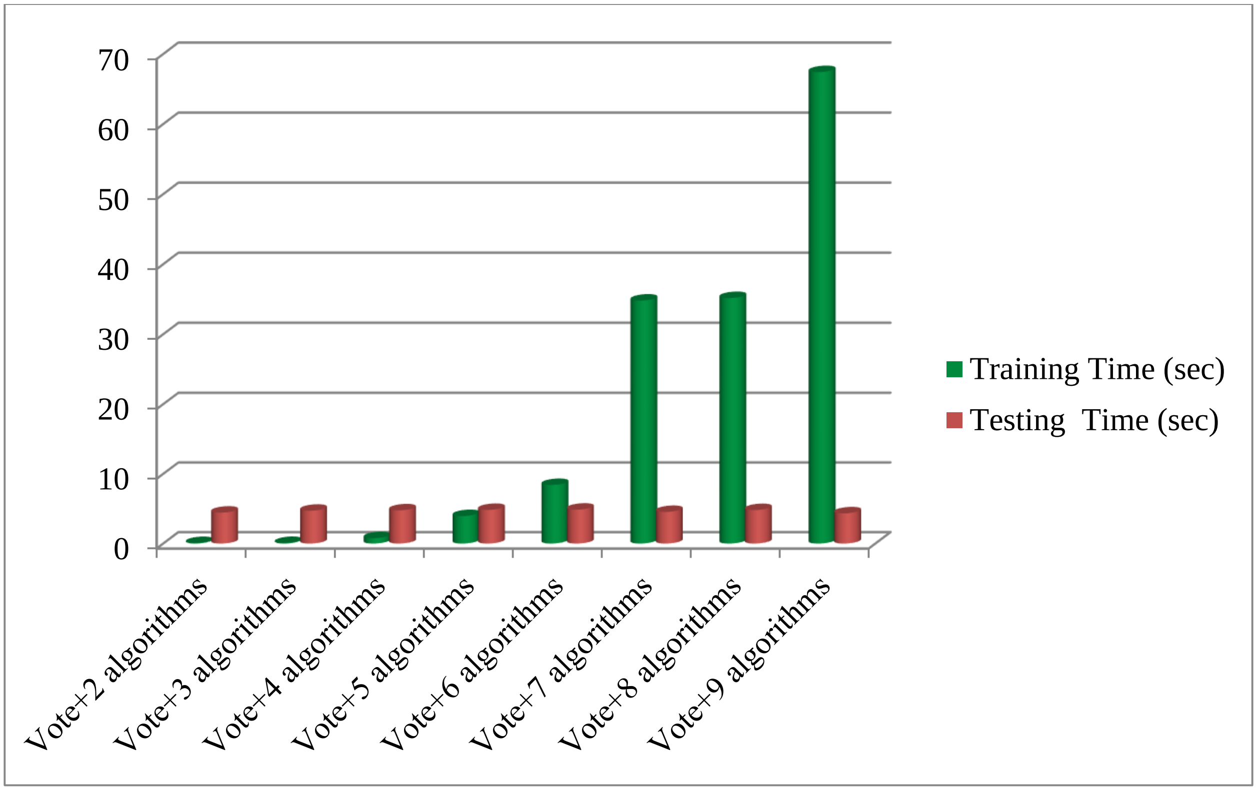


Figure 5.4. Comparison between the Ensemble models according to training and testing time.

Figure 5.4 compared between the Ensemble models in term of time taken to build model and time taken to test model on supplied test set. We deduced the longest time to build ensemble model is 67.42 seconds belong to ensemble Vote+9 algorithms, while the shortest time is 0.08 second belong to ensemble Vote+2 algorithms. If we look to the time taken to test model on supplied test set the results showed that the longest test time 4.87 belong to ensemble Vote+6 algorithms, while the shortest test time 4.34belong to ensemble Vote+9 algorithms.

Table 5.3 displays a comparison between the best Ensemble model and its base algorithms according to correlation coefficient, mean absolute error, root mean squared error, time taken to build model and time taken to test model.

Table 5.3: Comparison between the best Ensemble model and its base algorithms

Algorithm	Correlation coefficient	Mean absolute error	Root mean squared error	time taken to build model in sec	time taken to test model in sec
IBk	0.8192	0.0905	0.3005	0.02	0.34
KStar	0.8901	0.1091	0.2285	**0.01**	4.59
M5P	0.8863	0.1047	0.2322	0.1	**0.01**
Ensemble Vote+3	**0.8986**	**0.0888**	**0.1092**	0.09	4.71

As shown in Table 5.3, ensemble Vote+3 algorithms outperformed its own basic algorithms and produced the best results. It obtained the highest correlation coefficient 0.8986, the lowest of both mean absolute error 0.0888 and root mean squared error 0.1092. Figure 5.5 shows the comparison between the best ensemble method Vote+3 and its basic algorithms according to correlation coefficient, mean absolute error and root mean squared error.

Also Vote+3 achieved the longest time for both building model 0.09 sec and testing model 4.71 sec. Figure 5.6 displays a comparison between the best ensemble method Vote+3 and its basic algorithms according to time taken to build model and time taken to test model in seconds.

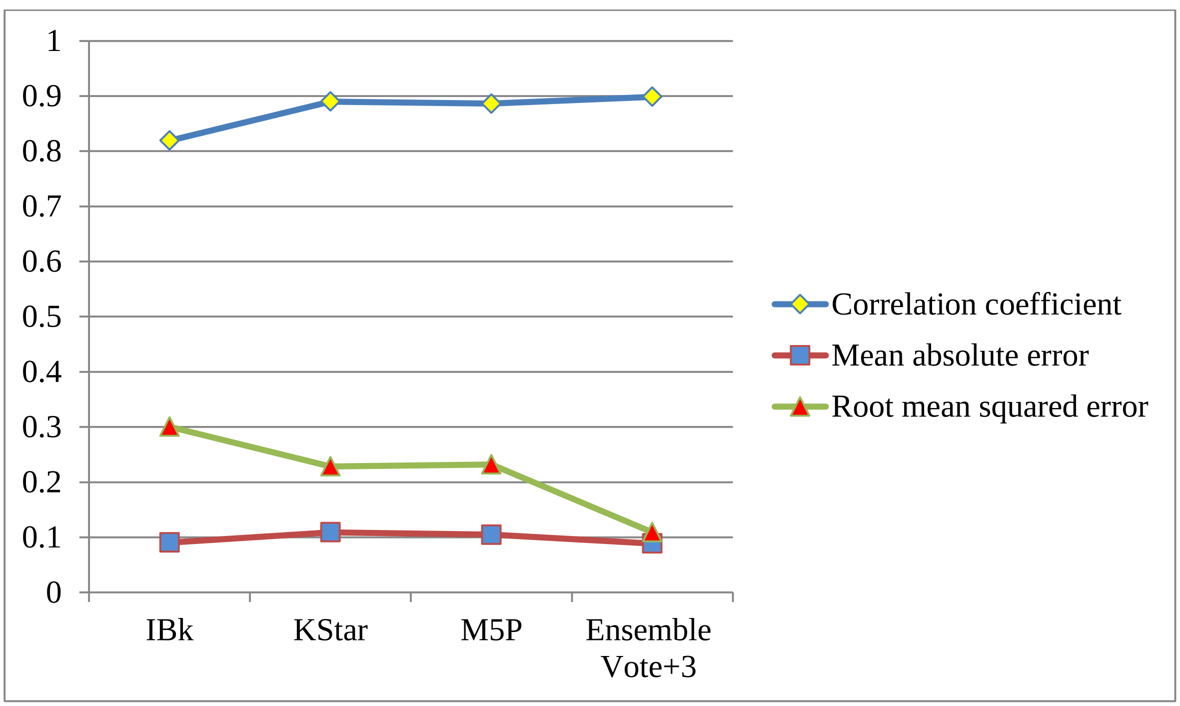


Figure 5.5. Comparison between the proposed Ensemble Vote+3 and its basic algorithms.

As shown in Figure 5.5, the ensemble Vote+3 surpassed its base algorithms, because it achieved the highest correlation coefficient 0.8986, the lowest of both mean absolute error 0.0888 and root mean squared error 0.1092. KStar came in the second order in term of both correlation coefficient 0.8901and root mean squared error 0.2285, but it had the highest mean absolute error0.1091. Thirdly in term of correlation coefficient came M5P 0.8863. IBK obtained the worst of both correlation coefficient 0.8192 and root mean squared error 0.3005, but at the same time it achieved the second best mean absolute error 0.0905.

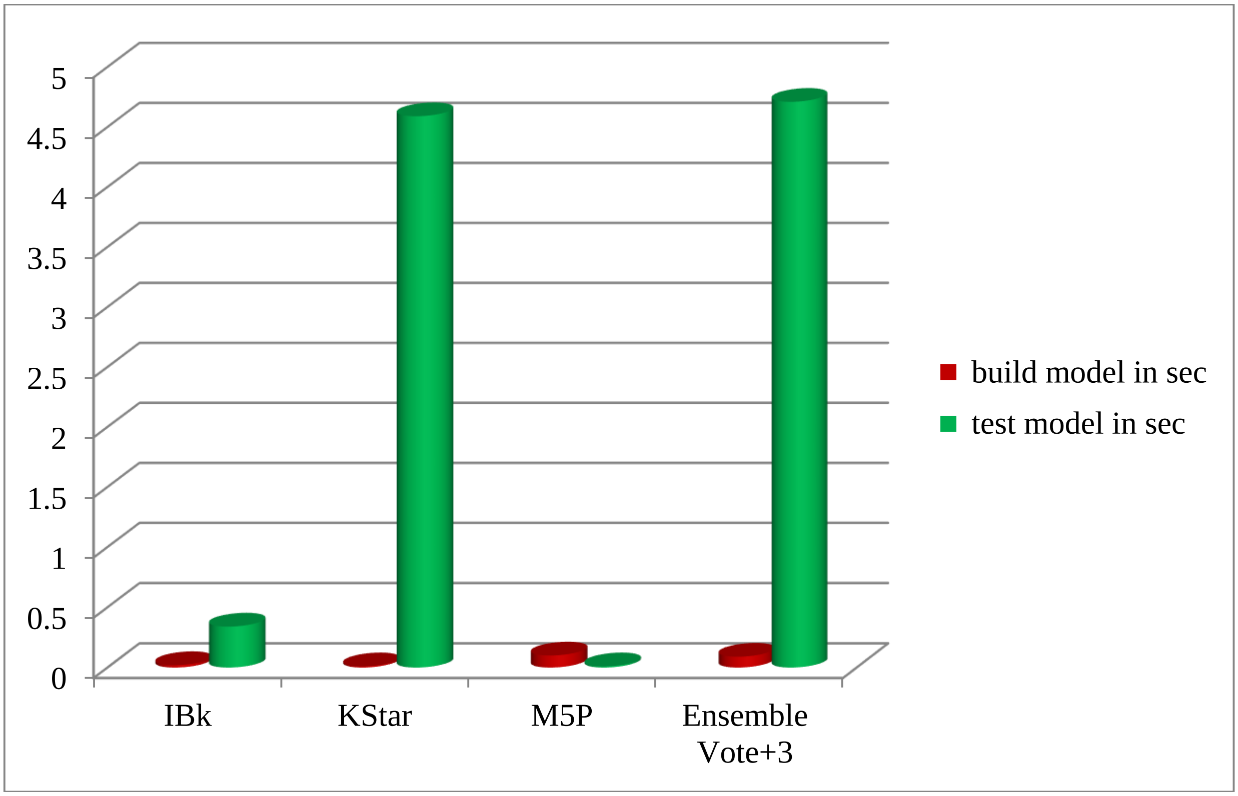


Figure 5.6. Comparison between the proposed ensemble method Vote+3 and its basic algorithms.

As shown in Figure 5.6 ensemble Vote+3 obtained the second worse time to build model 0.09 sec and the longest time for testing 4.71sec. The shortest building time 0.01sec was obtained by KStar algorithm. IBK came in the second order for both building time 0.02 sec and testing time 0.34sec. The longest building time 0.1sec has been acquired by M5P algorithm, but at the same time M5P achieved the shortest testing time 0.01sec. KStar obtained the second worse test time 4.59 sec.

5.3 Experimental Results for ANFIS

ANFIS system is sensitive to number of membership function (MF). Giving additional number of membership function to the system did not always improve the result. Negnevitsky et al. [279] stated that a larger number of MFs better represents a complex system and therefore should produce better results. However, a large number of inputs or MFs in the premise part of the rules can produce a large number of fuzzy rules which can cause the learning complexity of ANFIS to suffer an exponential explosion. This is called the curse of dimensionality which can adversely affect the performance of ANFIS [27, 280, 281]. Not many literature papers have alluded to what is considered to be a large number of fuzzy rules. However, drawing on the experience of [27, 280] as a guide to choosing the number of MFs per input and since the largest number of inputs to be used was 7 then the smallest number of MFs that could produce overlapping while not invoking the curse of dimensionality is 2. The number of MFs was increased with one of the ANFIS models to get a greater understanding of the impact on the performance of ANFIS with this change.

In this study, by increasing number of membership function from 2 to 3 the results are better. But by increasing number of membership function from 3 to 4, in most cases the performances are decreasing. Applying 3 number of membership function gave the best results. Different types of membership functions have been used, Table 5.4 compared between the results of ANFIS models with different membership functions using the measure performance.

Table 5.4: ANFIS results with different membership function.

Membership function	**Correlation coefficient**	**MAE**	**RMSE**
trapmf	0.85	0.0297	0.1724
trimf	0.47	3.6442	1.9090
gaussmf	0.83	0.0228	0.1512
gauss2mf	0.75	0.0378	0.1946
gbellmf	**0.90**	**0.0074**	**0.0861**

dsigmf	0.59	3.5171	1.8754
psigmf	0.59	3.5167	1.8753
pimf	0.61	3.24828	1.8023

As shown in Table 5.4 the model which used the bell-shaped membership function outperform the others, since it has the lowest both mean absolute error and root mean squared error and the highest correlation coefficient, close to unity. Other functions such as Gaussian and Trapezoidal were used as well to evaluate the performance with different types of MFs, and they come in second order after the bell-shaped. The results of the both ANFIS models that used Gaussian and Trapezoidal were close to each other and they performed better than the rest models. The ANFIS model which used triangular gave the worst performance.

The bell-shaped MF was favored over the other types since it offered more parameters which provided a greater number of degrees of freedom. The generalized bell-shaped MF is standard for ANFIS because of its smoothness and concise notation [27, 279, 280]. Figures 5.7 and 5.8 shows the changes in generalized bell shaped membership functions before and after training stage.

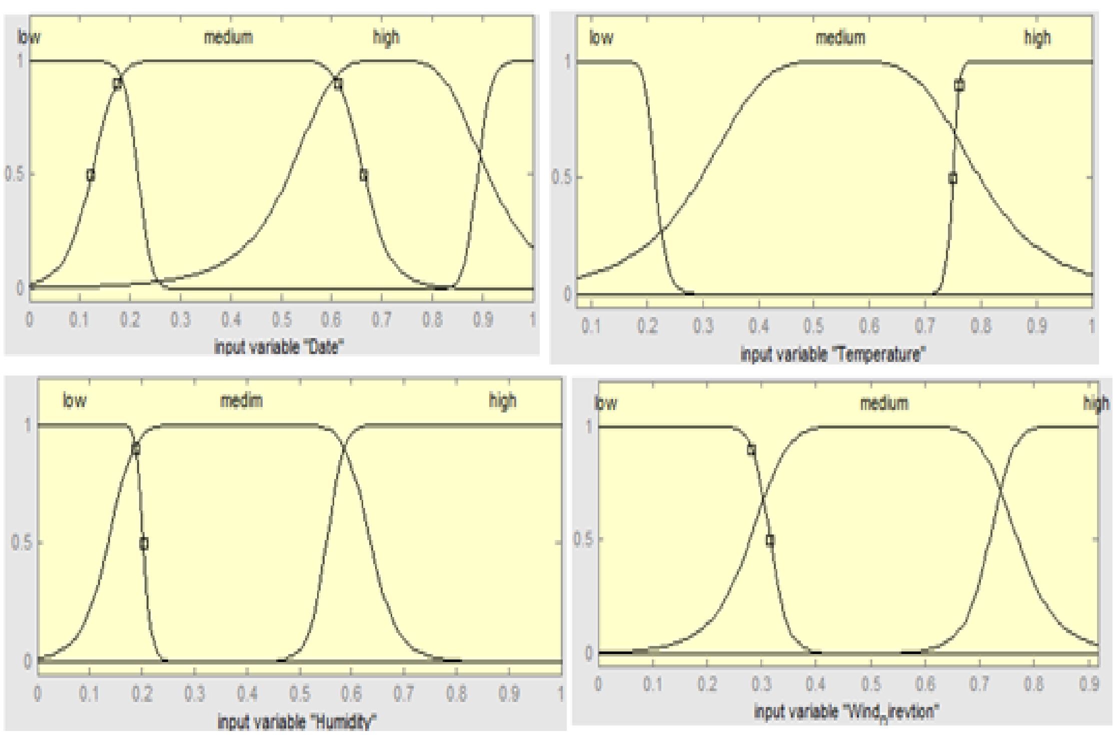


Figure 5.7. Initial membership functions for inputs date, minimum temperature, humidity and wind direction.

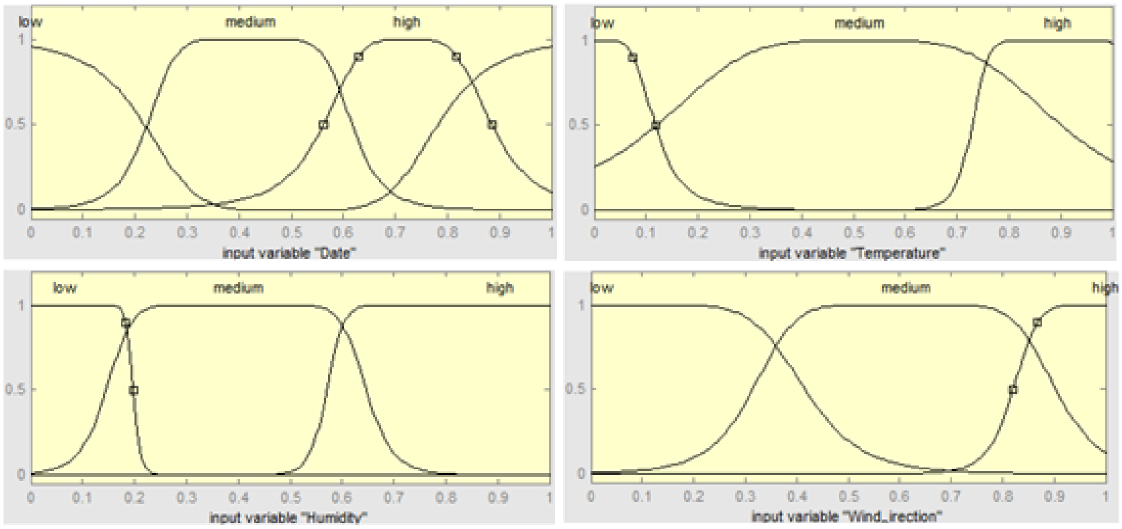


Figure 5.8. Final membership functions for inputs date, minimum temperature, humidity and wind direction.

The models were developed using 12 numbers of membership functions of type generalized bell with 81 If-then rules for all different sets of training and testing dataset for rainfall prediction in ANFIS. After obtaining the results the best model for the long term rainfall prediction was selected and highlighted by means of the evaluation parameters that are mean absolute error (MAE) root mean squared error (RMSE) and correlation coefficient (CC) values given in table 5.5.

Table 5.5: ANFIS results with different ratio of training and testing dataset.

Dataset Ratio % (Train – Test)	**Training**			**Testing**		
	CC	**RMSE**	**MAE**	**CC**	**RMSE**	**MAE**
90-10	0.91	0.0815	0.0069	0.81	0.1847	0.0341
80-20	0.93	0.0080	0.0064	0.84	0.1293	0.0167
70-30	**0.96**	**0.0056**	**0.0031**	**0.90**	**0.0861**	**0.0074**
60-40	0.808	1.3951	1.9460	0.79	1.7470	2.0520

From the comparison of the four options of dataset as depicted in Ttable 5.5, the best results have been obtained when we applied ANFIS model with ratio 70-30 of dataset for training and testing. It achieved the highest correlation coefficient and the lowest of both mean absolute error and root mean squared error in training and testing phases. When applying the ANFIS model with dataset ratios 80-20 and 90-10 results are close and still acceptable, but when the dataset ratio 60-40 has been used for training and testing we obtained the worst results.

There are two optimization methods: hybrid and back propagation. To develop the ANFIS rainfall prediction models, both of them were used. The hybrid technique is more popularly used with ANFIS than the back propagation [25, 279]. In addition, it is regarded as the faster of the two techniques [25].

Average errors of monthly rainfall as predicted by using the back-propagation method and by using the hybrid learning algorithm are shown in Table 5.6

Table 5.6: Average Errors of Back-propagation and the Hybrid Algorithm.

Method	**Training error**	**Testing error**
Hybrid	**0.005697**	**0.086139**
Back Propagation (BP)	0.012586	0.114682

From this table, the capability of the ANFIS model in predicting the rainfall using the hybrid learning algorithm is much better than when using the back propagation method. This is because the hybrid method comprises of back propagation and Least-Square methods.

Figures 5.9 and 5.10 show training and testing errors for ANFIS model with bell membership function and hybrid learning algorithm.

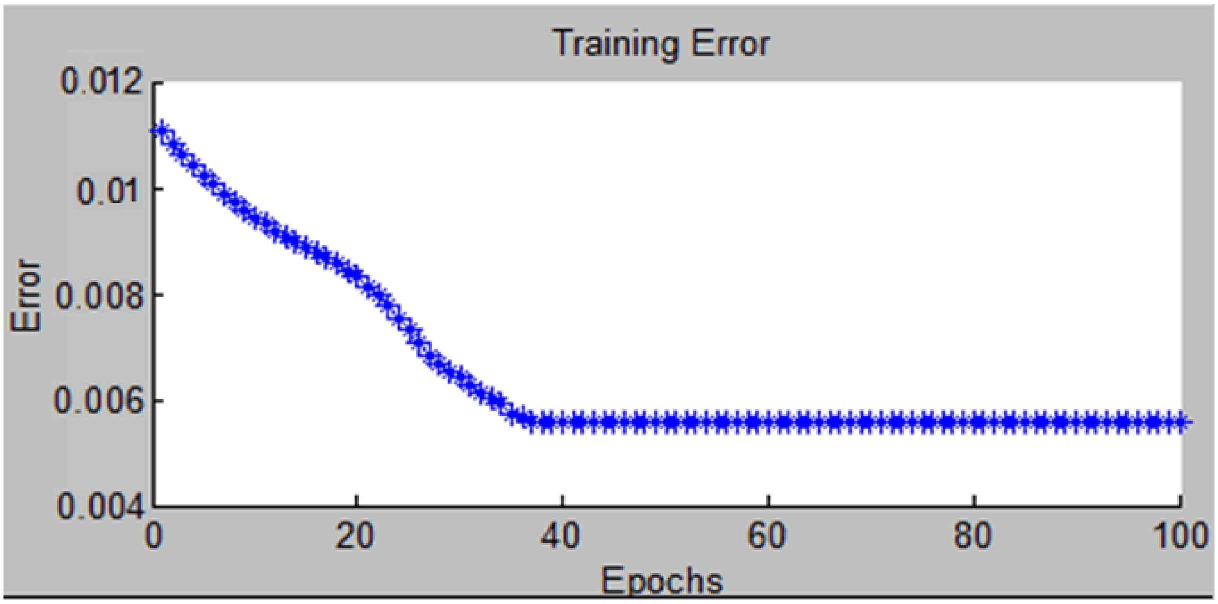


Figure 5.9. Training error for 70% of dataset at 100 epochs.

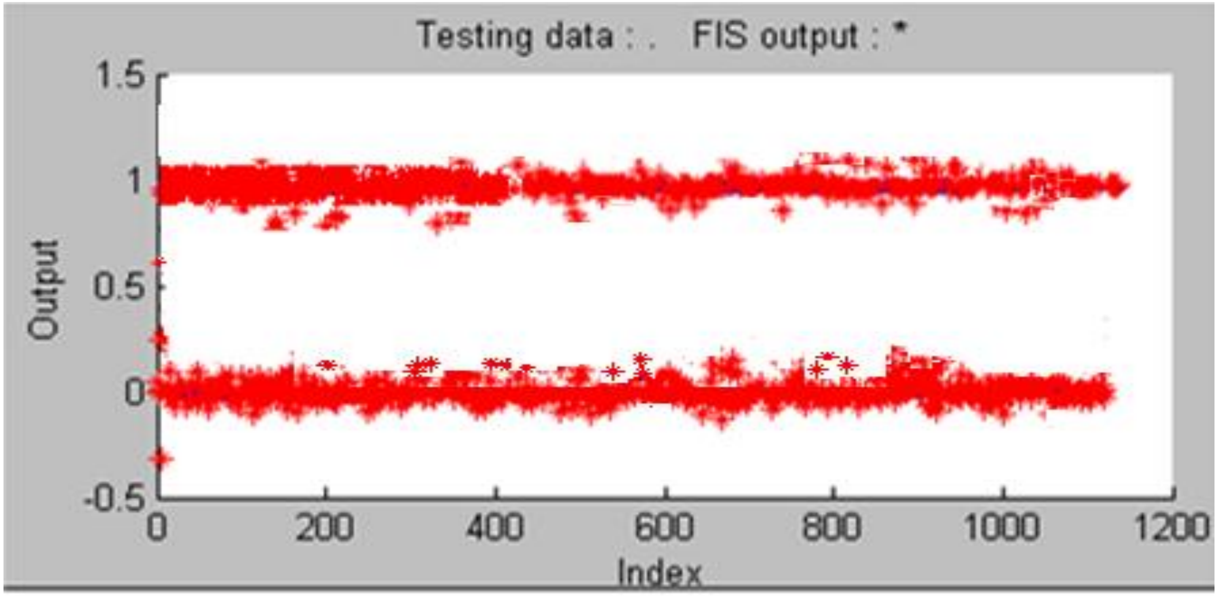


Figure 5.10. Testing error for 30% of dataset.

Figures 5.11-5.16 display Surface viewer for (a) Date with temperature and output rainfall, (b) Temperature with wind direction and output rainfall, (c) Date with humidity and output rainfall, (d) Date with wind direction and rainfall, (e) Humidity with wind direction and output rainfall and (f) Humidity with temperature and output rainfall respectively.

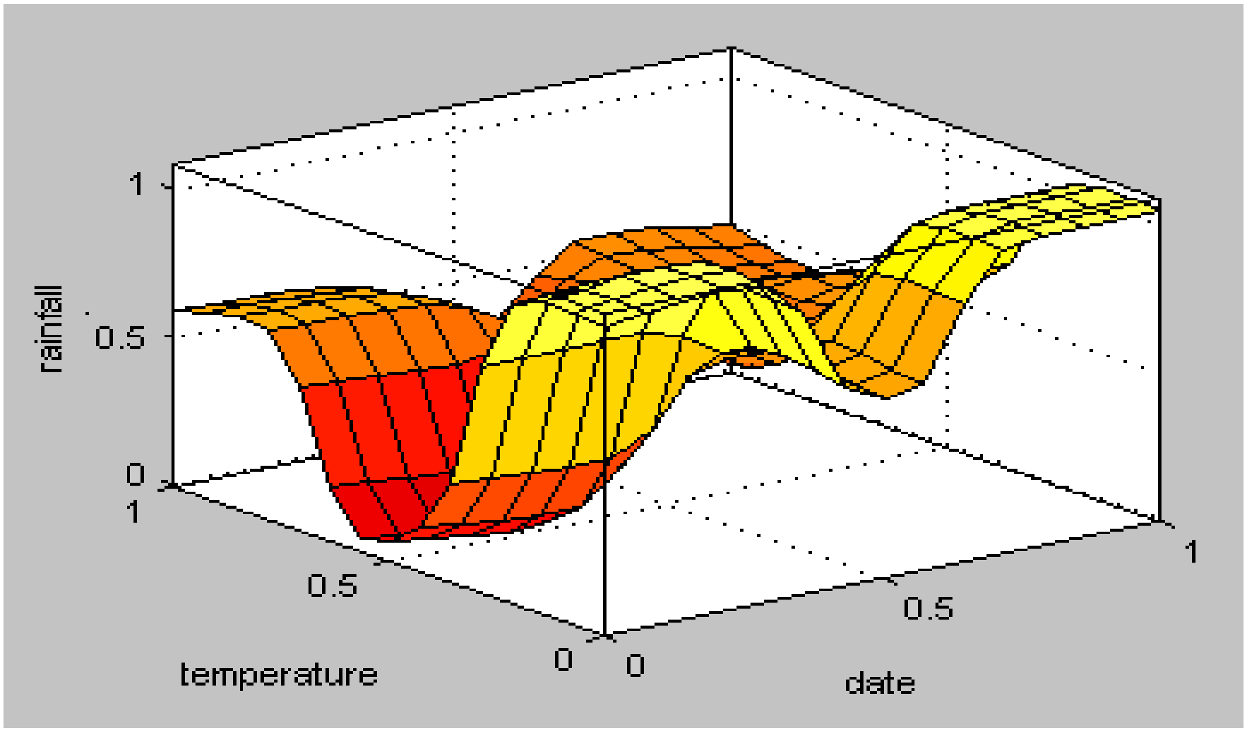


Figure 5.11. Surface viewer for date with temperature and output rainfall.

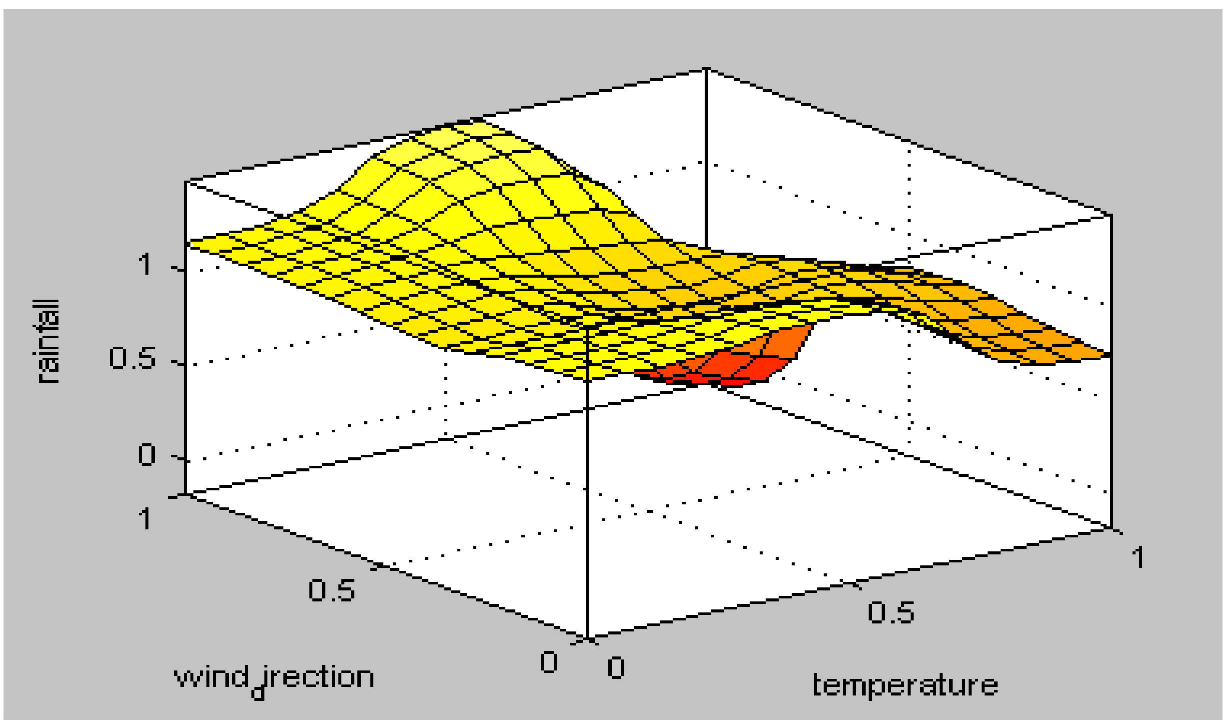


Figure 5.12. Surface viewer for temperature with wind direction and output rainfall.

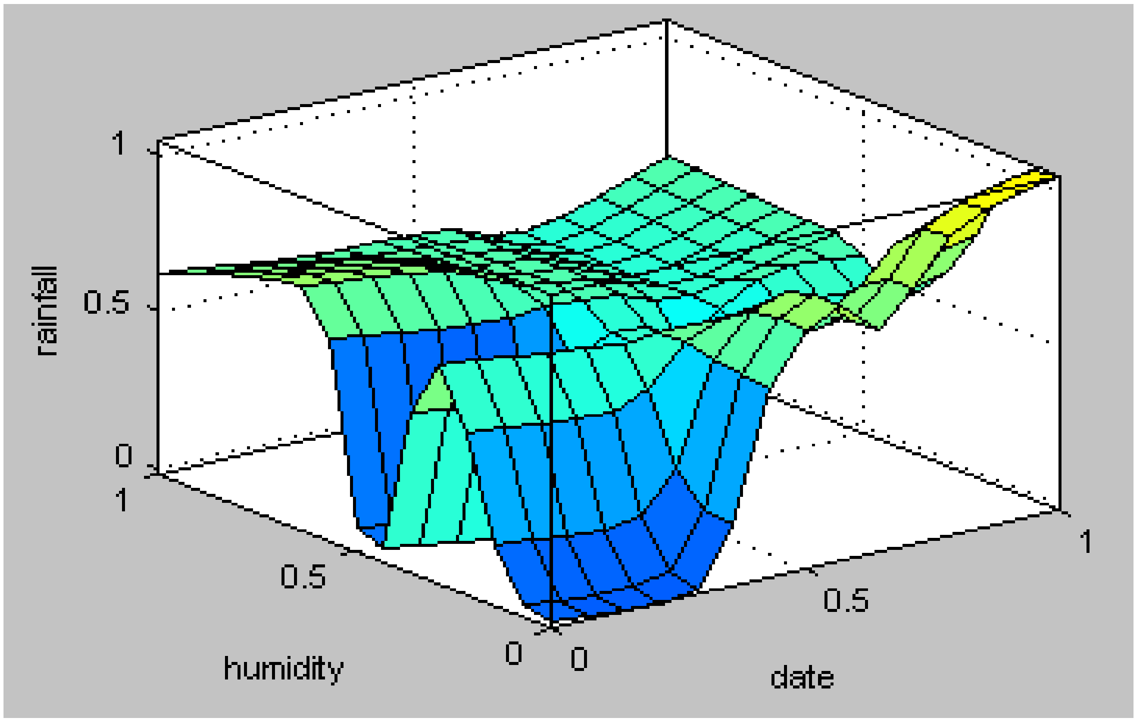


Figure 5.13. Surface viewer for date with humidity and output rainfall.

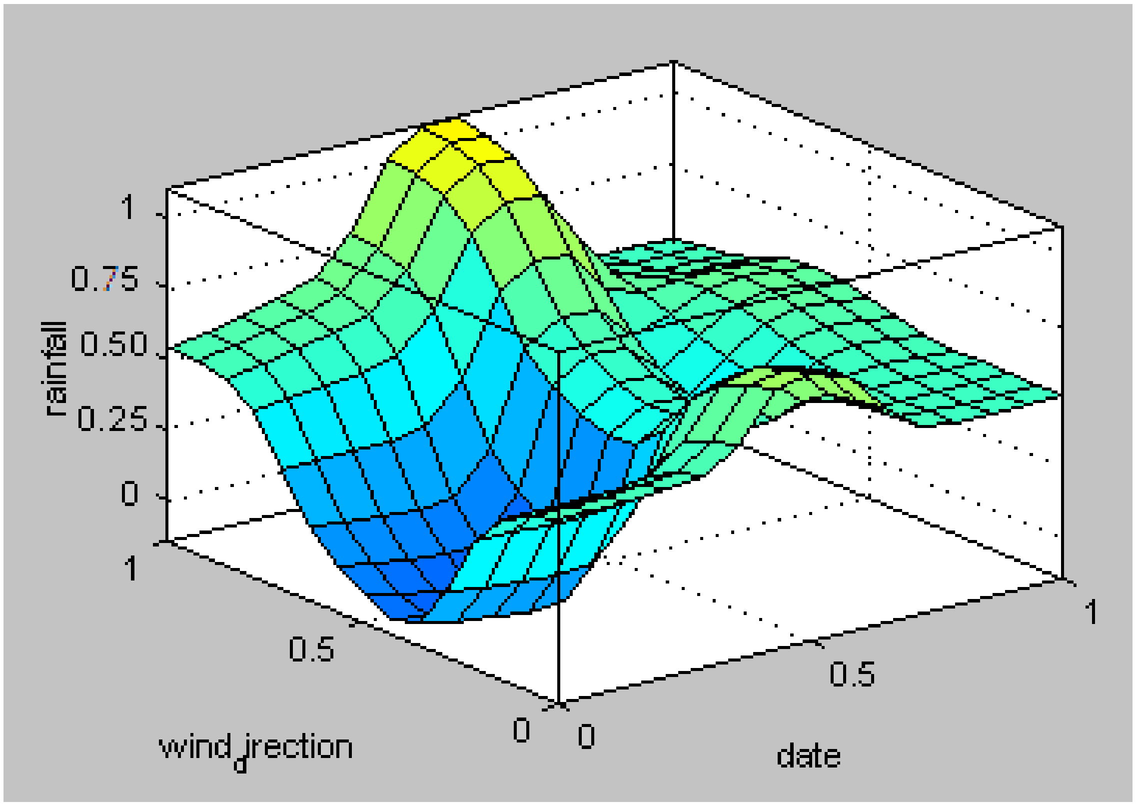


Figure 5.14. Surface viewer for date with wind direction and rainfall.

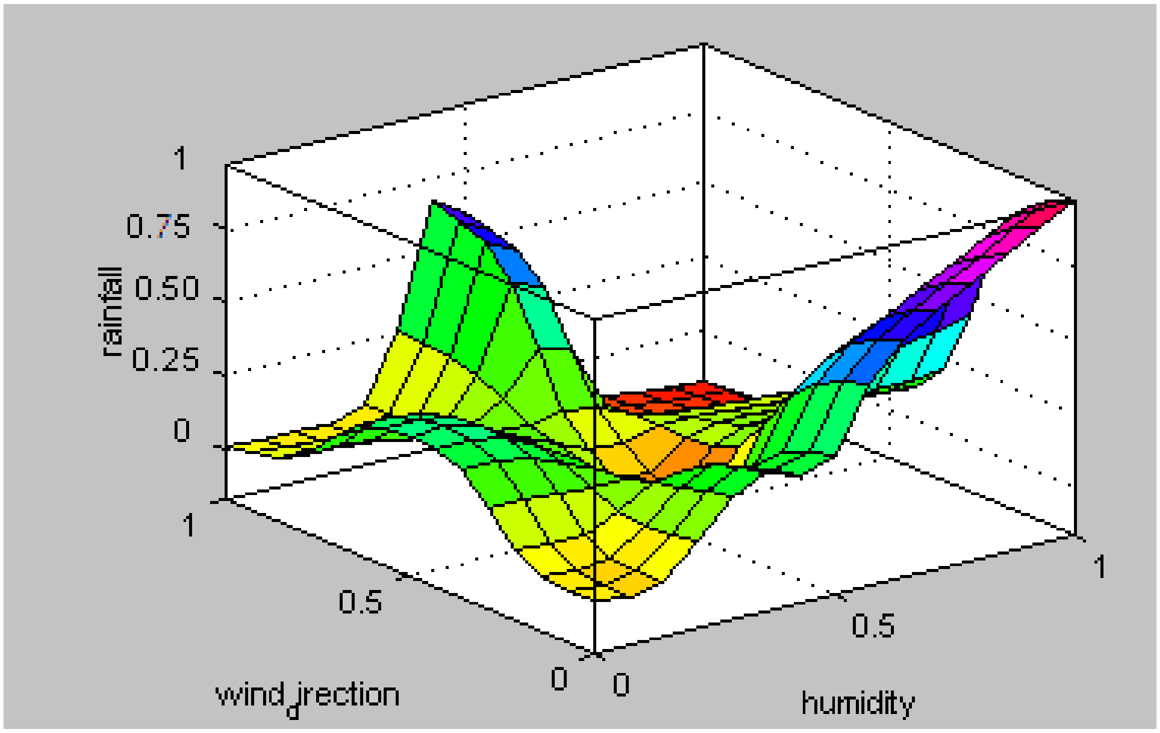


Figure 5.15. Surface viewer for humidity with wind direction and output rainfall.

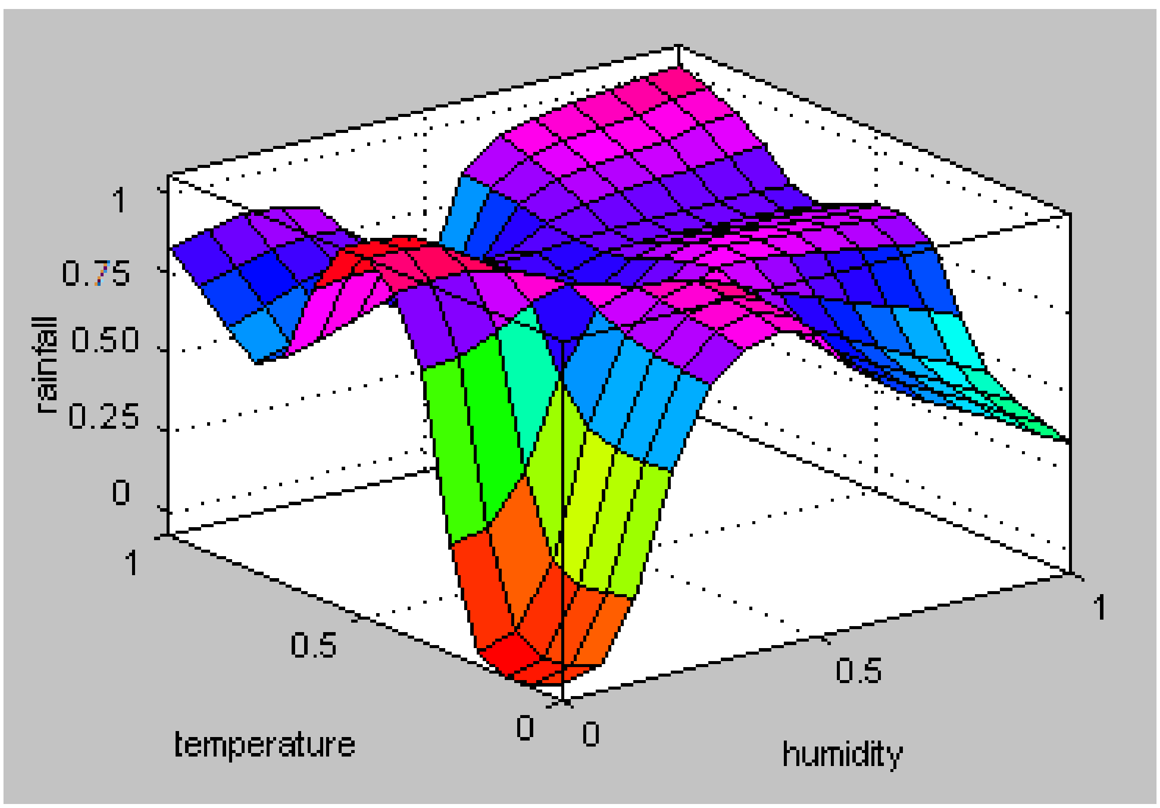


Figure 5.16. Surface viewer for humidity with temperature and output rainfall.

5.4 Proposed Ensemble, ANFIS and Other Models from Literature

Table 5.7 displays results of both Vote+3 and ANFIS model according to correlation coefficient, mean absolute error and root mean squared error.

Table 5.7: Comparison between the proposed ensemble Vote+3 and ANFIS model

Algorithm	Correlation coefficient	Mean absolute error	Root mean squared error
ANFIS	**0.90**	**0.0074**	**0.0861**
Vote+3	0.8986	0.0888	0.1092

As Table 5.7 shows, ANFIS model outperformed the ensemble Vote+3 model in term of all performance criteria, it achieved the highest correlation coefficient 0.90, lowest of both mean absolute error 0.0074 and root mean squared error 0.0861.

Figure 5.17 shows a comparison between the ensemble Vote+3 and ANFIS model according to correlation coefficient, mean absolute error and root mean squared error.

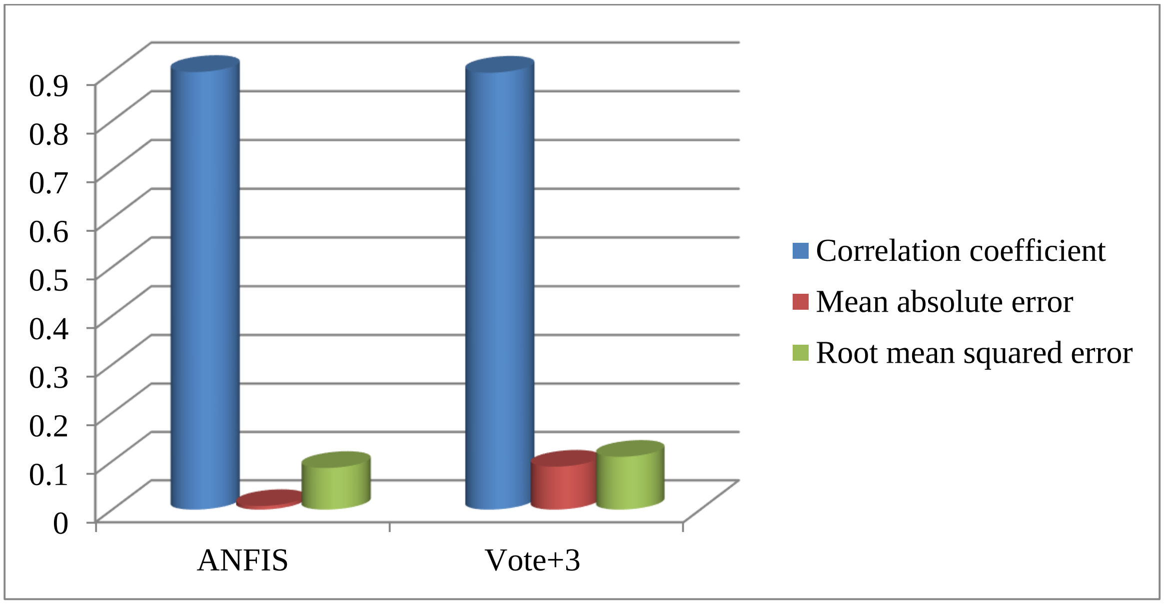


Figure 5.17. Comparison between ensemble Vote+3 and ANFIS model according to correlation coefficient, mean absolute error and root mean squared error.

As shown in Figure 5.17, results of ANFIS and ensemble Vote+3 are too closed, but the results of ANFIS model are more accurate. ANFIS model provided higher correlation coefficient and lower of both mean absolute error and root mean squared error comparing with ensemble Vote+3.

Table 5.8 shows a comparison between ensemble Vote+3 and ANFIS model with models proposed in the literature for related applications.

Table 5.8: Comparison between the proposed Vote+3, ANFIS and models in the literature.

Reference	Technology used	Application	Performance
[61]	ANN architecture, with Genetic Optimizer (GO)	Seasonal rainfall forecasting	CC = 0.8951
[4]	Evolving Fuzzy Neural Network (EFuNN)	Forecast the monthly rainfall	RMSE = 0.0901
[187]	Multivariate Adaptive Regression Splines (MARS)	Forecast monthly rainfall	RMSE = 0.0780
[82]	multilayered artificial neural network	Seasonal rainfall prediction	MSE = 0.42
[245]	Bagging classifier using REPTree classifier as a base-learner.	Daily wind speed forecasting	CC = 0,8154 RMSE = 0,8774
ANFIS	**ANFIS using grid partition and hybrid algorithm for learning**	**Monthly rainfall prediction**	**CC = 0.90 RMSE = 0.086139**
Proposed Ensemble Vote+3	**Ensemble IBK, Kstar, M5P**	**Monthly rainfall prediction**	**CC = 0.8986 RMSE = 0.1092**

4.5 Summary

It is important to have reliable and accurate techniques to forecast the rainfall in the long term. In this study we proposed models based on different soft computing technologies namely ensemble technology and ANFIS. Finally we compared them until a model that produced satisfactory results was obtained.

Different evaluation measures such as correlation coefficient, mean absolute error, root mean squared error, time taken to build model and time taken to test model in seconds have been used for comparing between different models to determine which one is the highest performance.

Dividing dataset into 70-30 for training and testing respectively considered the best choice for dividing our dataset, since it provided the best results comparing with the other choices for both training and testing phases.

Different types of membership functions have been used to decide which one is the most appropriate for the target, the results showed that ANFIS model which used generalized bell shaped membership outperforms the others and achieve the highest performance accuracy. Also our experiments ensured that the hybrid learning method is much better than back propagation method as learning algorithm for ANFIS model and it give better and more accurate results.

In this study, we have used date, minimum temperatures, humidity and wind direction as predictors for long term rainfall prediction in Sudan. The empirical results indicate that ANFIS neuro-fuzzy and the proposed ensemble Vote+3 models are able to capture the dynamic behavior of the rainfall data and it may be useful in long term rainfall prediction.

ANFIS model outperformed the ensemble Vote+3model and produces better results. Also ANFIS model has ability to interpret and explain its results by using rules, and this feature is not available for other models.

6. Spatial Analysis and Rainfall Maps

6.1 Introduction

This chapter presents the spatial analysis of rainfall in Sudan for the interval 2000-2012 in three levels (towns, states and regions). The chapter is organized as follows:

Section 6.2 shows the rainfall maps, which produced as results of the spatial analysis in different levels (towns, states and regions) in Sudan. In all maps polygon [275] has been used to represent regions, states and meteorological stations at town level. Graduation techniques [277], such as graduation by size and graduation by color have been utilized to analyze and compare the phenomenon of rain on different levels. Also analysis been made for rain at the same level for different months. Finally we analyzed rainfall with the most influenced variables such as temperature and relative humidity for different months and levels.

Section 6.3 discusses the rainfall maps in details by discussing each related set of maps separately to extract general patterns that control rain phenomenon and its distribution over different geographical areas in Sudan.

6.2 Rainfall Maps

The following maps show the spatial analysis for rain phenomenon in Sudan by linking the rainfall averages with their geographical places (stations, states and regions). Figures (6.1, 6.3, 6.5 and 6.7) used graduation by color to represent rainfall while figures (6.2, 6.4, 6.6 and 6.8) used graduation by size.

Figures (6.9, 6.11 and 6.13) compared between different levels (stations, states and regions) in monthly rainfall amount for some years.

Figures (6.10, 6.12 and 6.14) displayed comparison between averages monthly temperature, humidity and rainfall for some months to year 2012 on different levels.

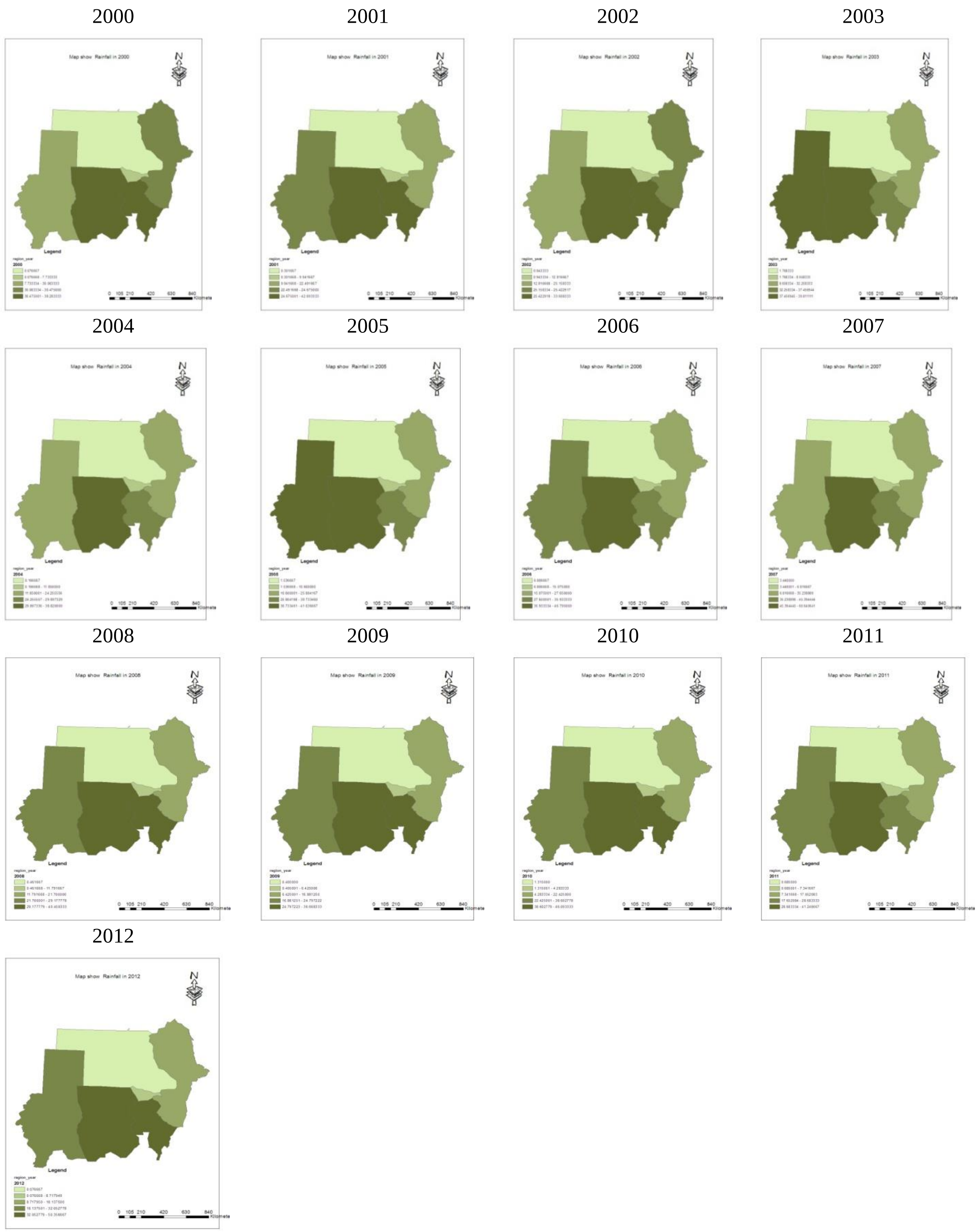


Figure 6.1. Rainfall on regions level for interval 2000 – 2012 by using graduation by color.

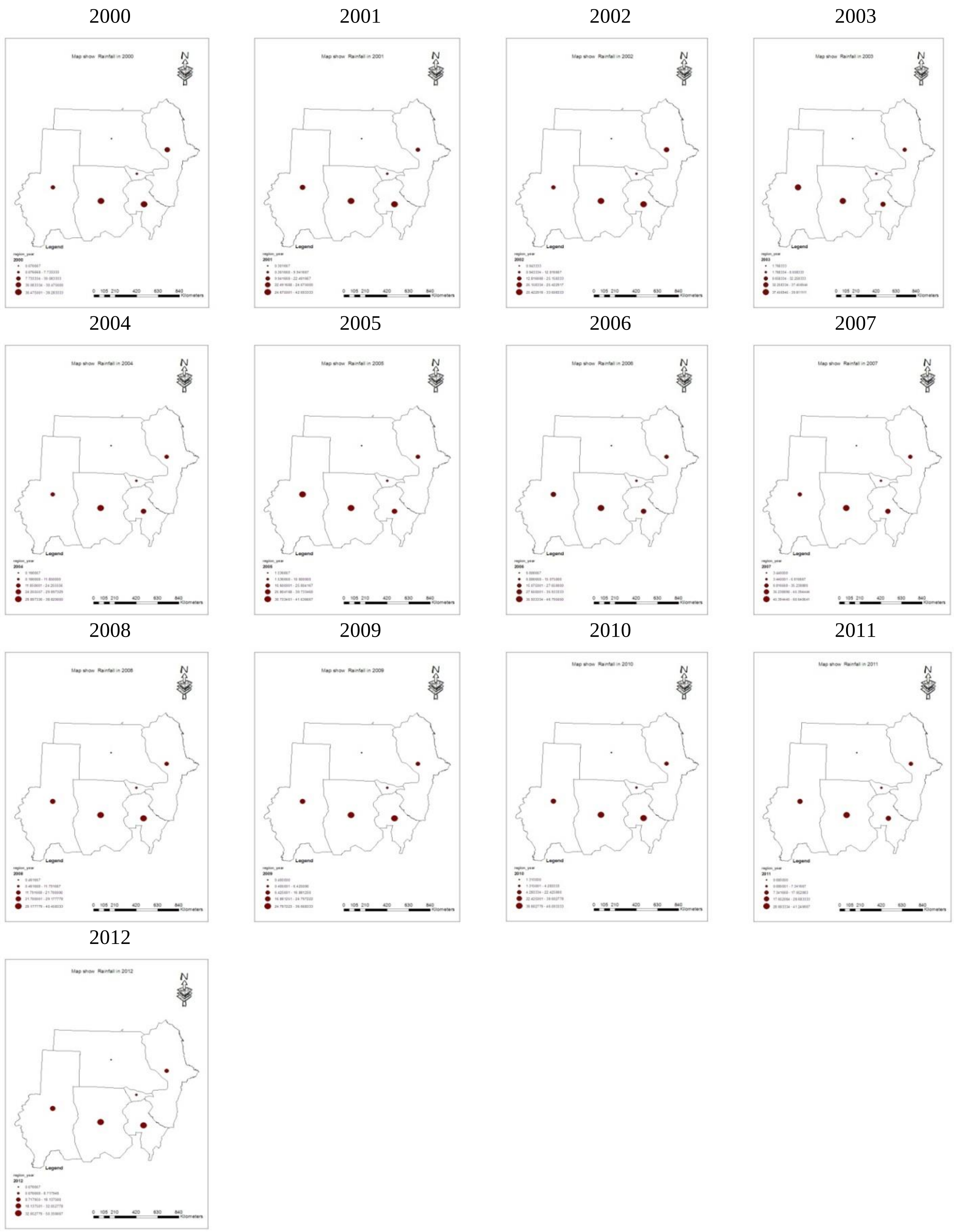


Figure 6.2. Rainfall on regions level for interval 2000 – 2012 using graduation by size.

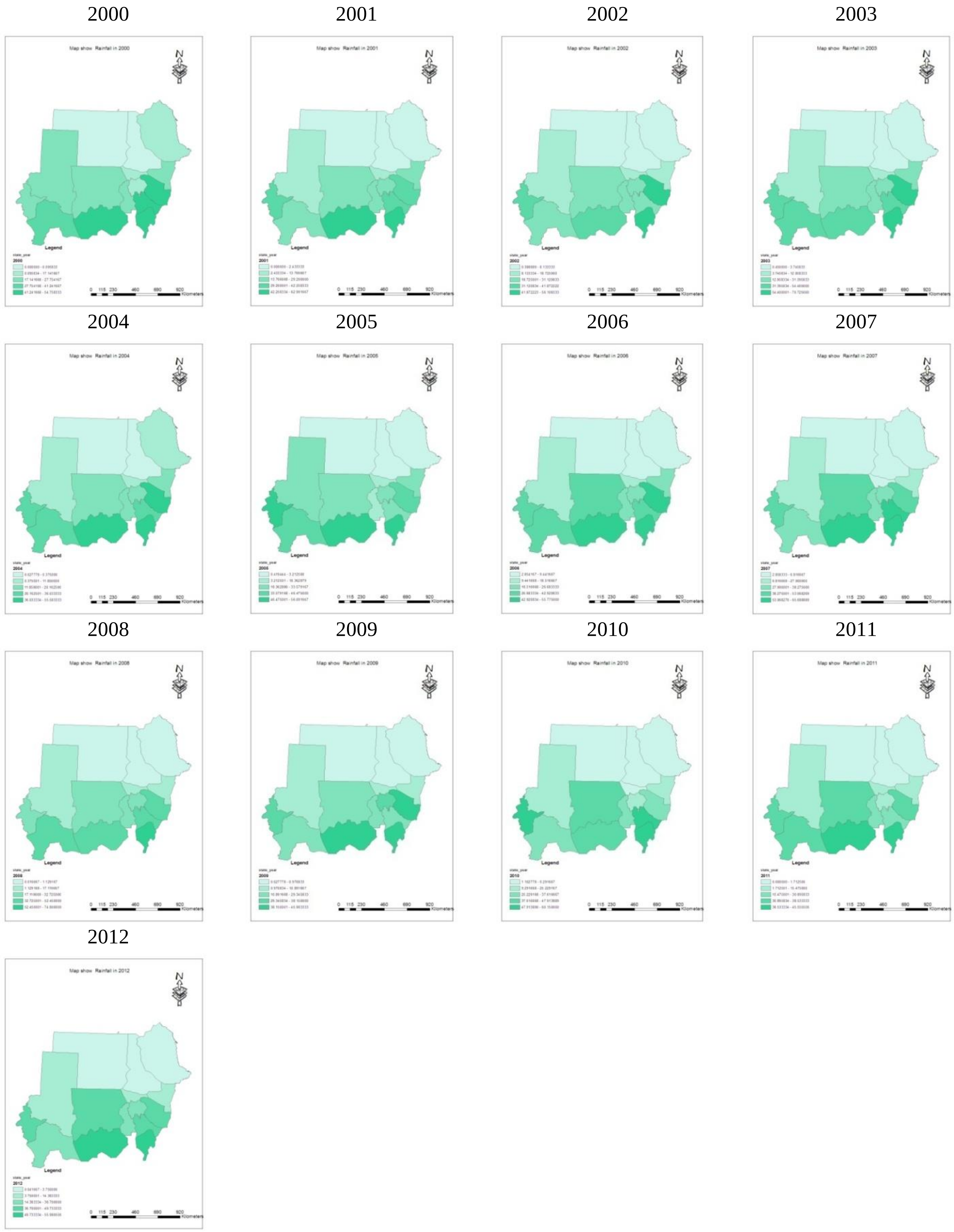


Figure 6.3. Rainfall on states level for interval 2000 – 2012 using graduation by color.

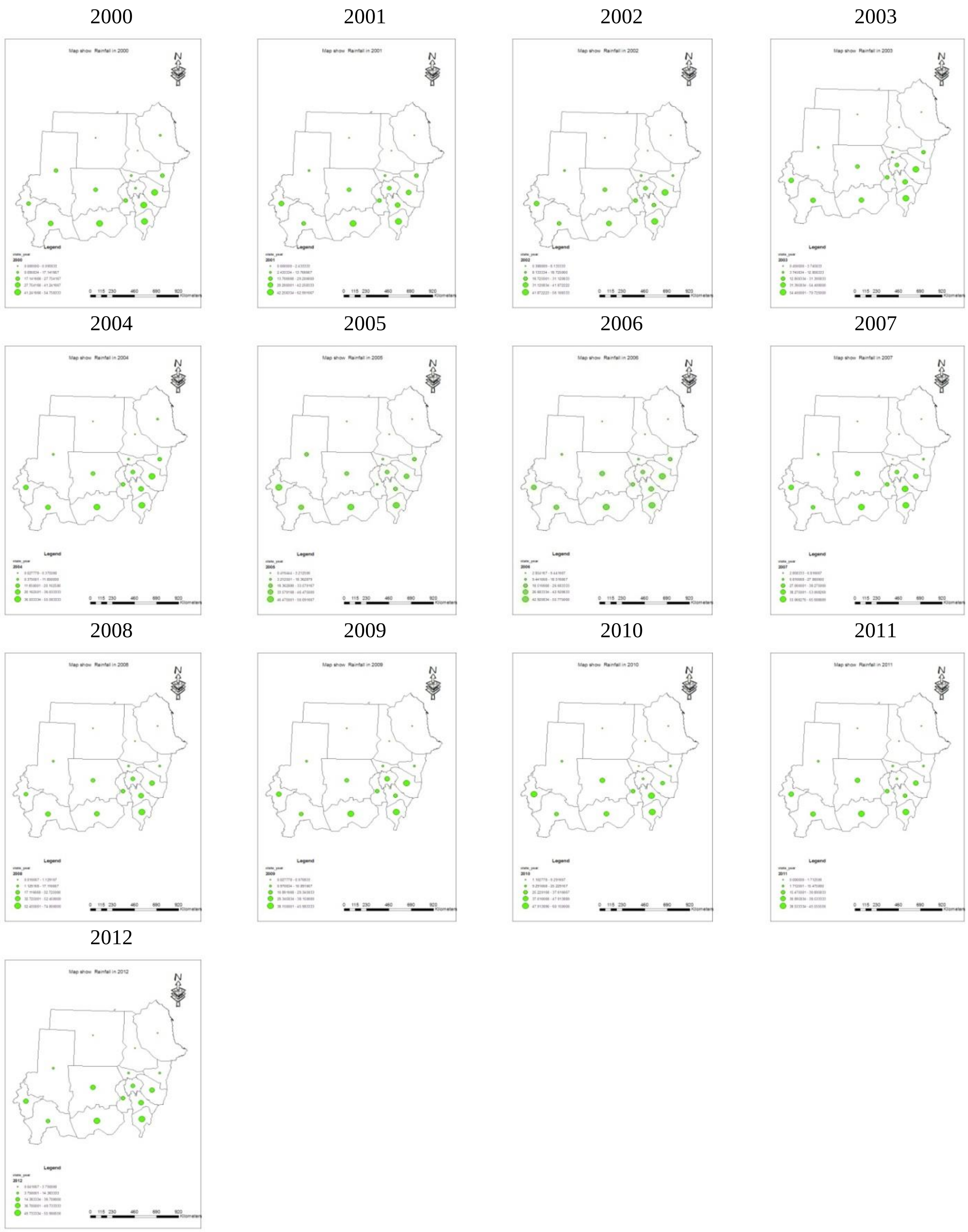


Figure 6.4. Rainfall maps on states level for interval 2000 – 2012 using graduation by size.

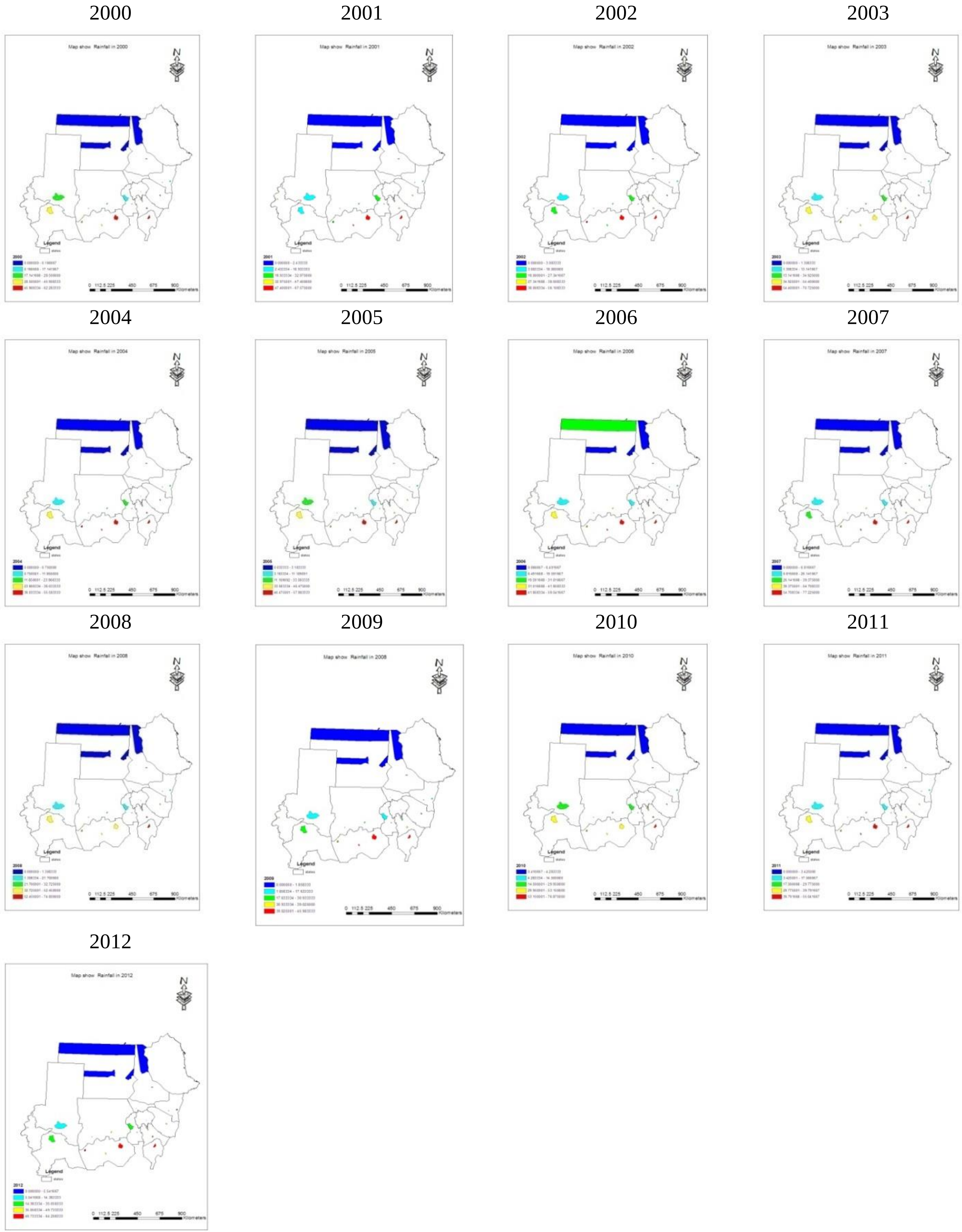


Figure 6.5. Rainfall on stations level for interval 2000 – 2012 using graduation by color.

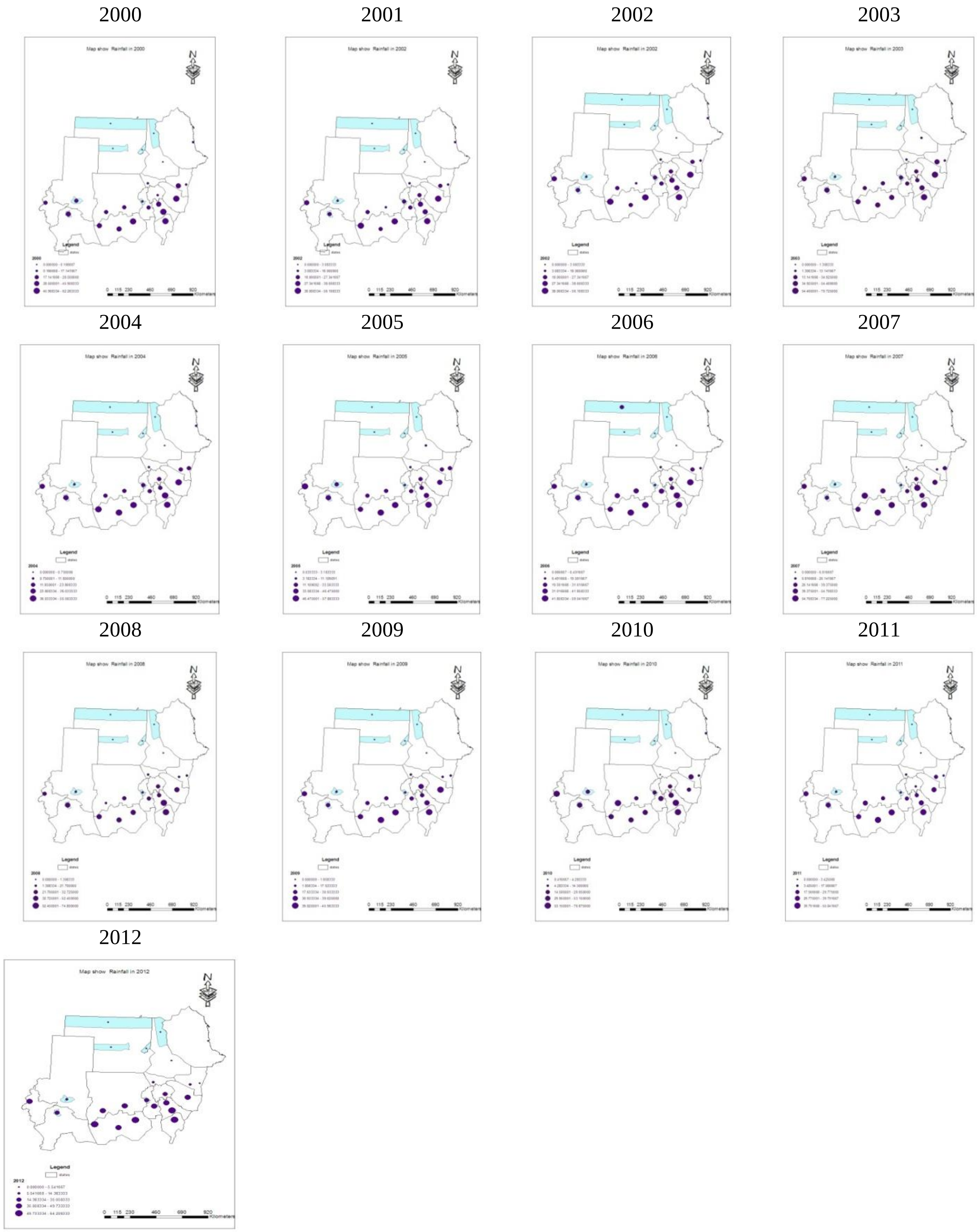


Figure 6.6. Rainfall on stations level for interval 2000 – 2012 using graduation by size.

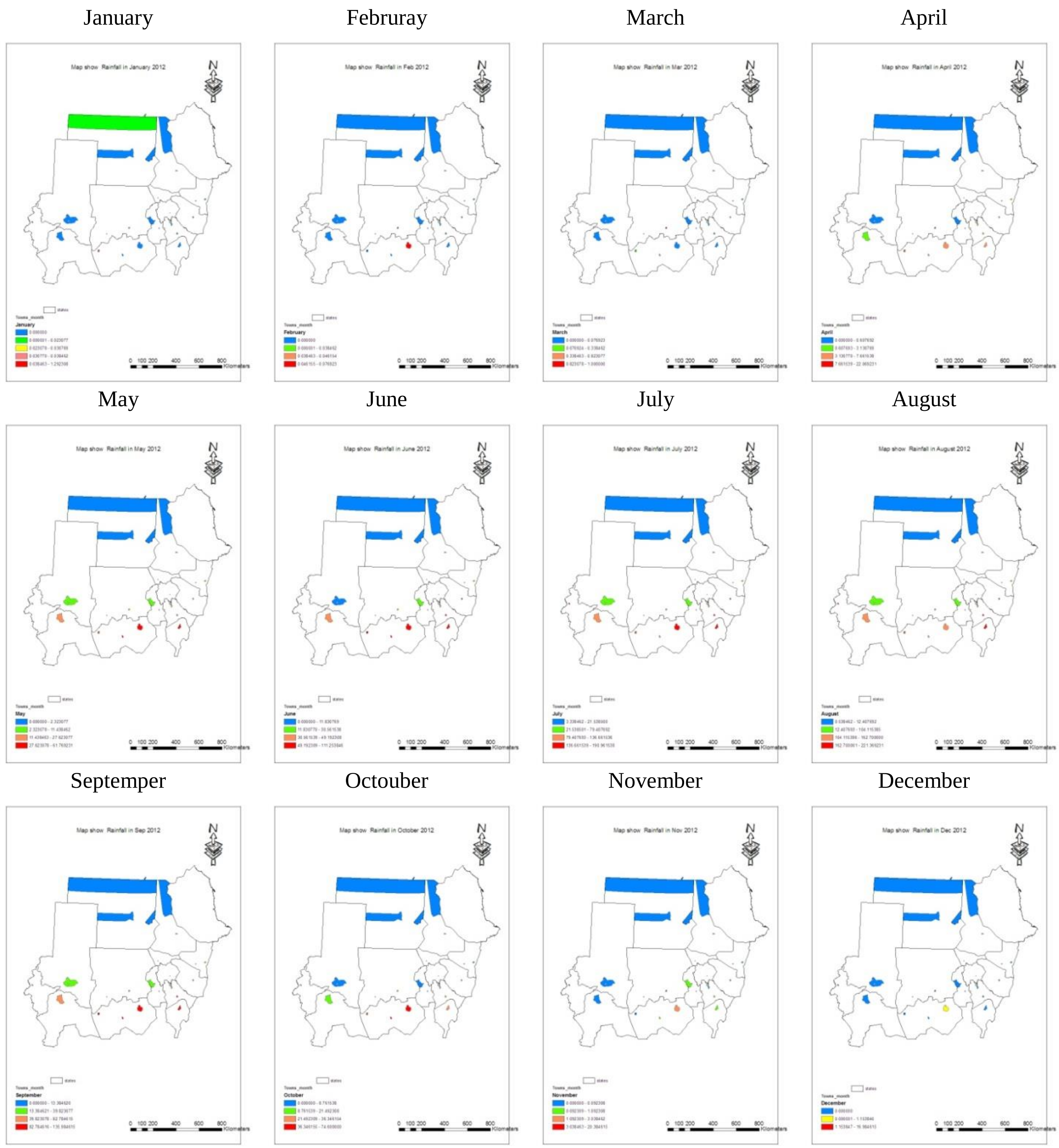
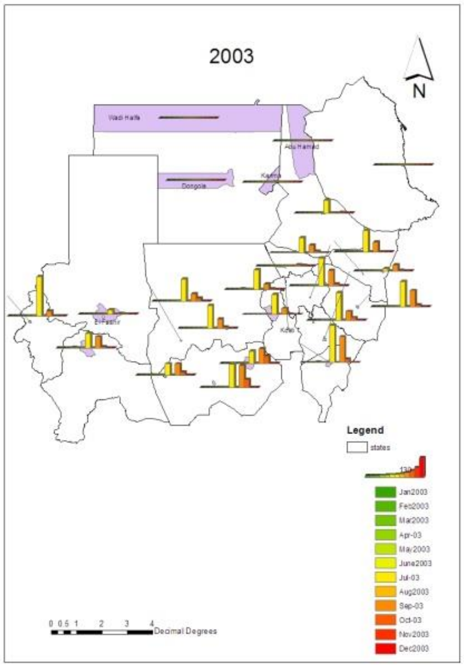


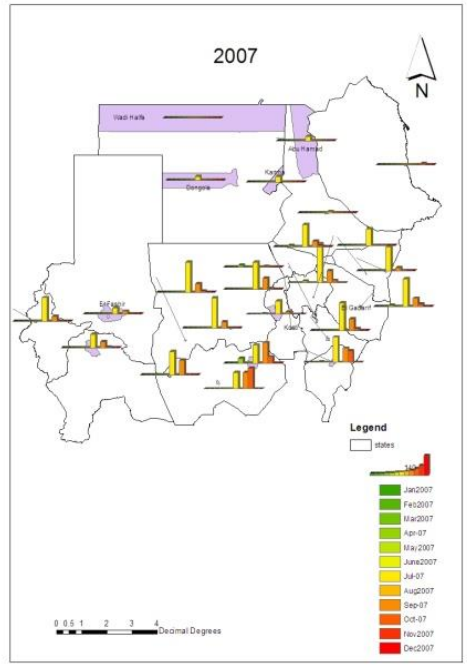
Figure 6.7. Average monthly rainfall on stations level for year 2012.

Figure 6.8. Average monthly rainfall on states level for year 2012 using gradation by size.

2003



2007



2012

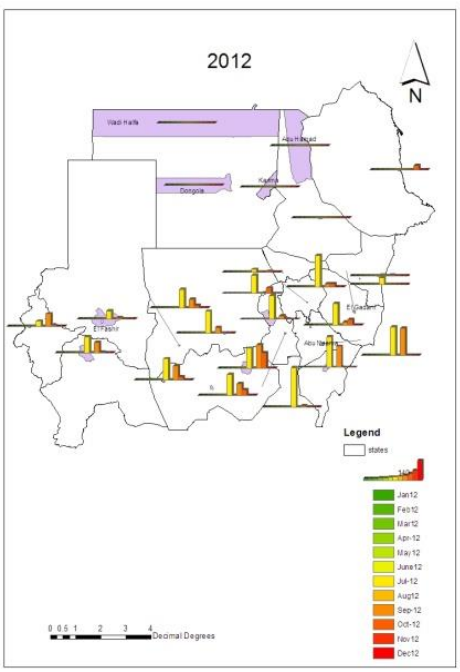


Figure 6.9. Comparison between average monthly rainfall on stations level for years 2003, 2007 and 2012.

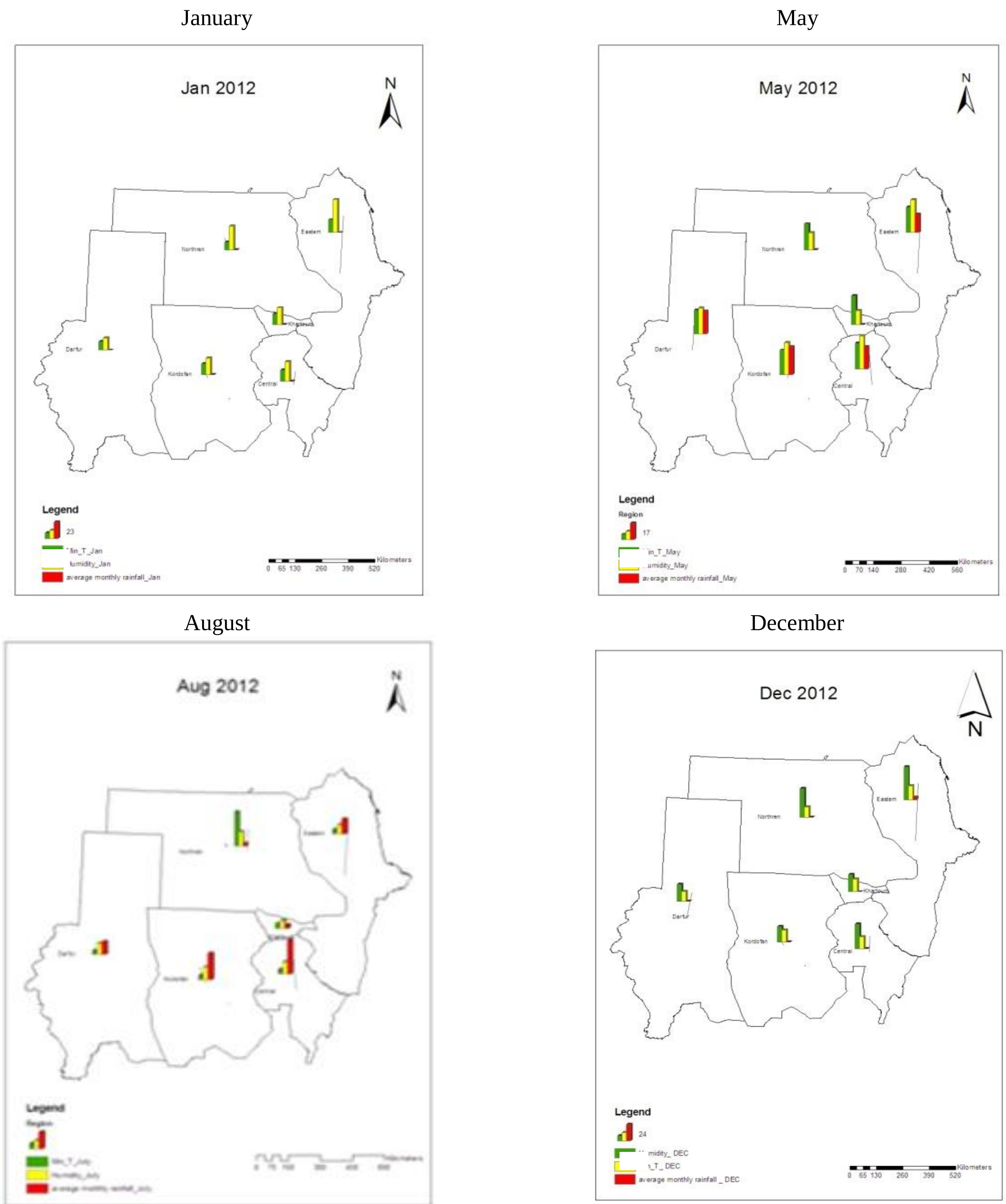
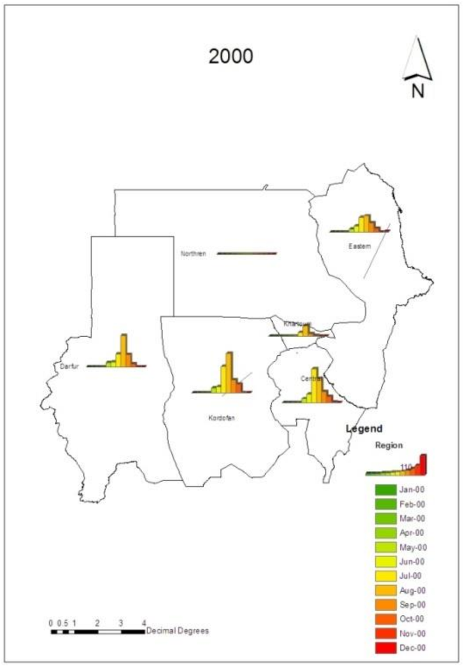
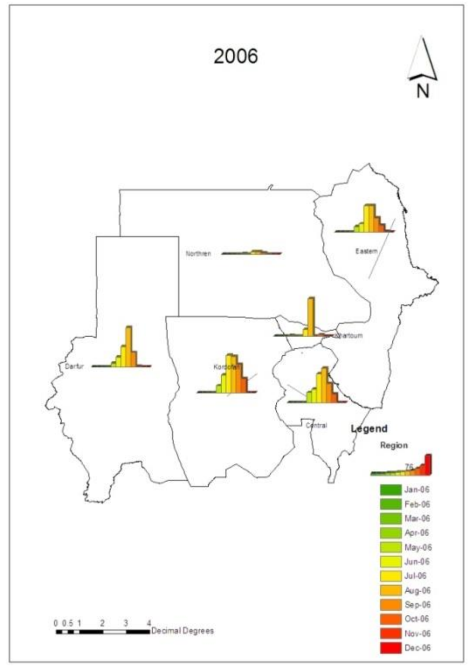


Figure 6.10. Comparison between average monthly temperature, humidity and rainfall on regions level for January, May, August and December 2012.

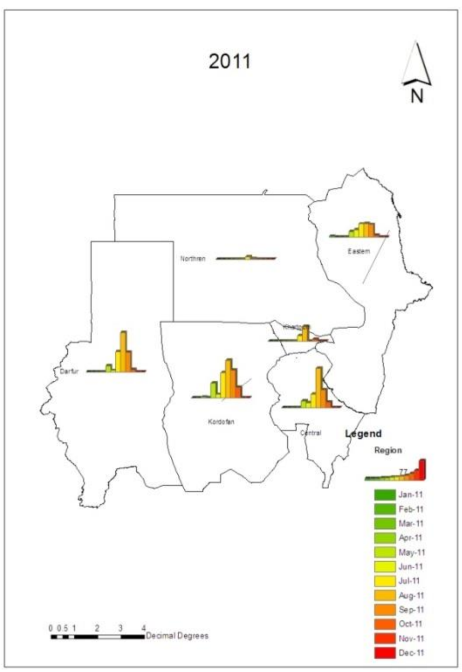
2000



2006



2011



2012

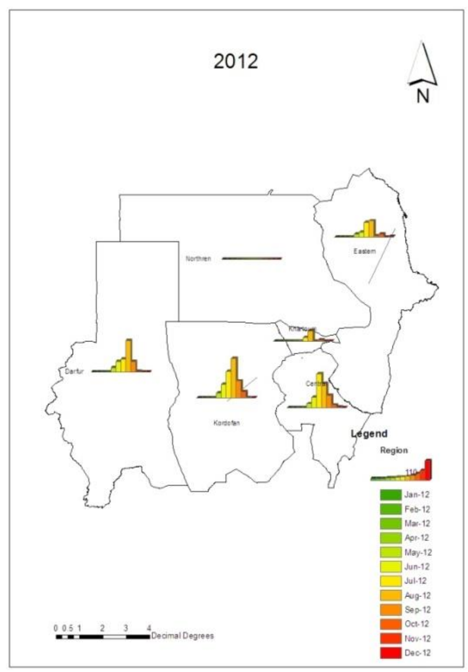


Figure 6.11. Comparison between average monthly rainfall on regions level for year 2000, 2006, 2011 and 2012.

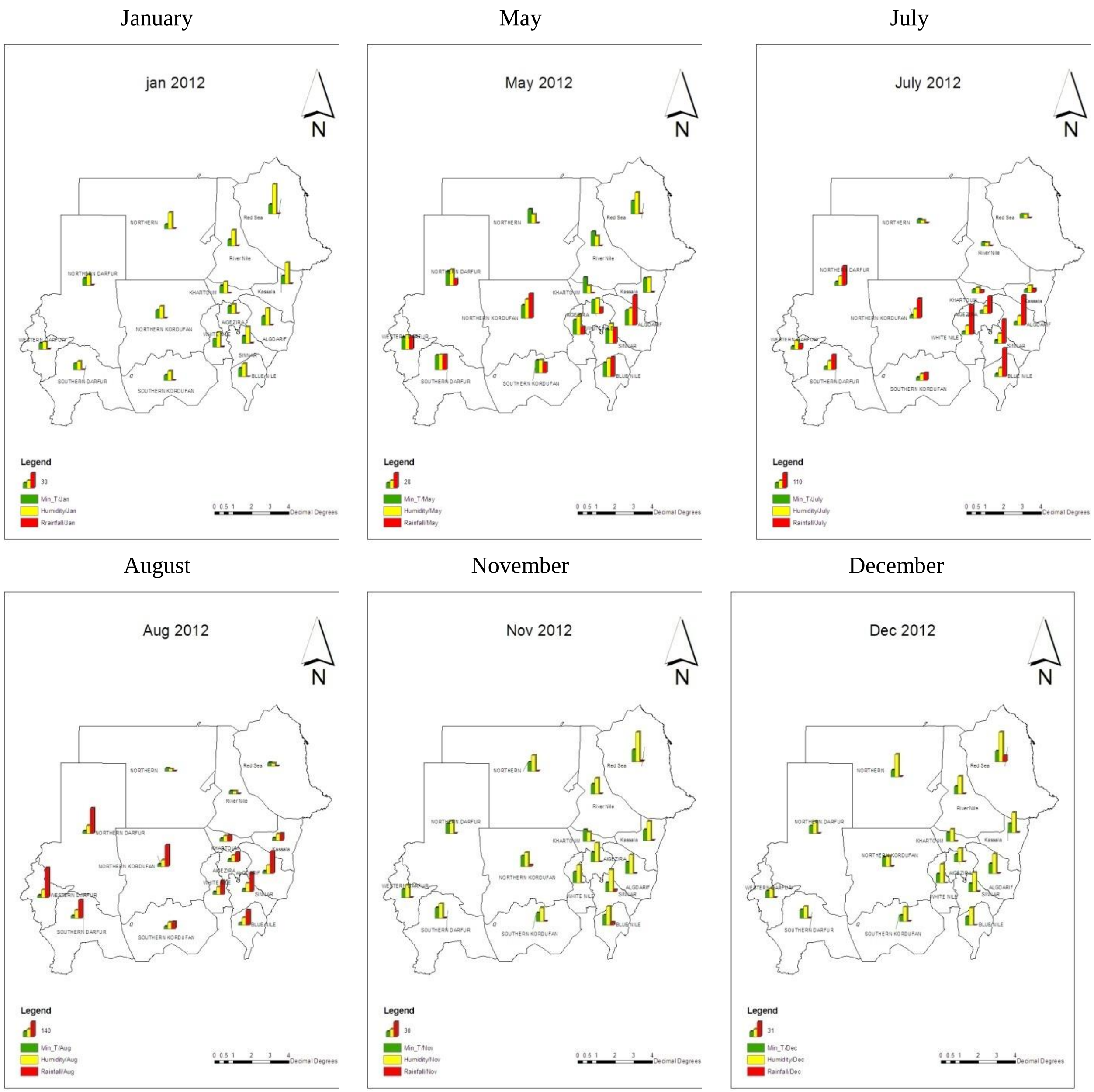
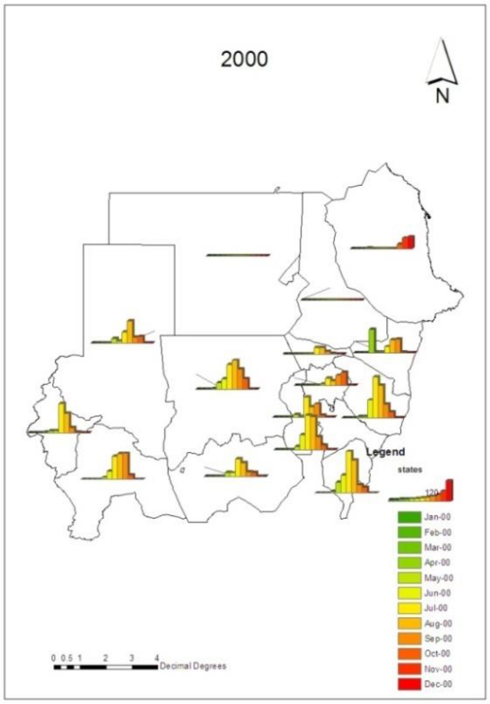
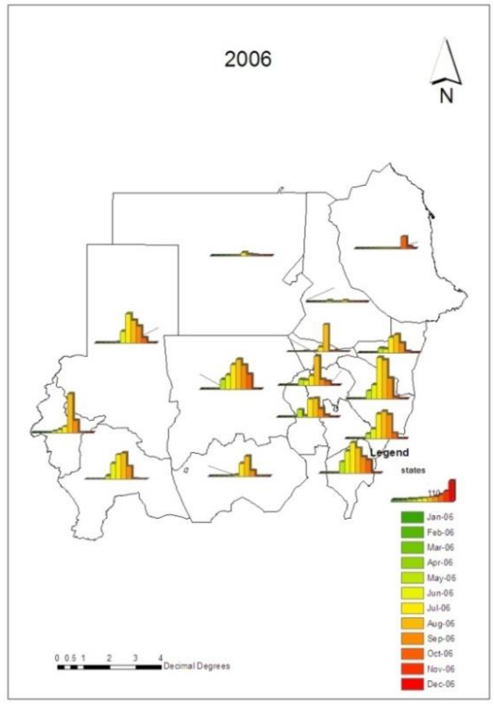


Figure 6.12. Comparison between average monthly temperature, humidity and rainfall on states level for January, May, July, August, November and December 2012.

2000



2006



2012

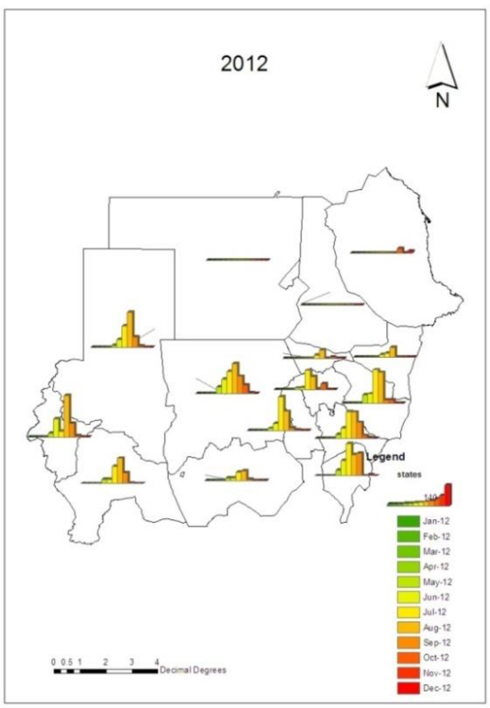


Figure 6.13. Comparison between average monthly rainfall on states level for year 2000, 2006 and 2012.

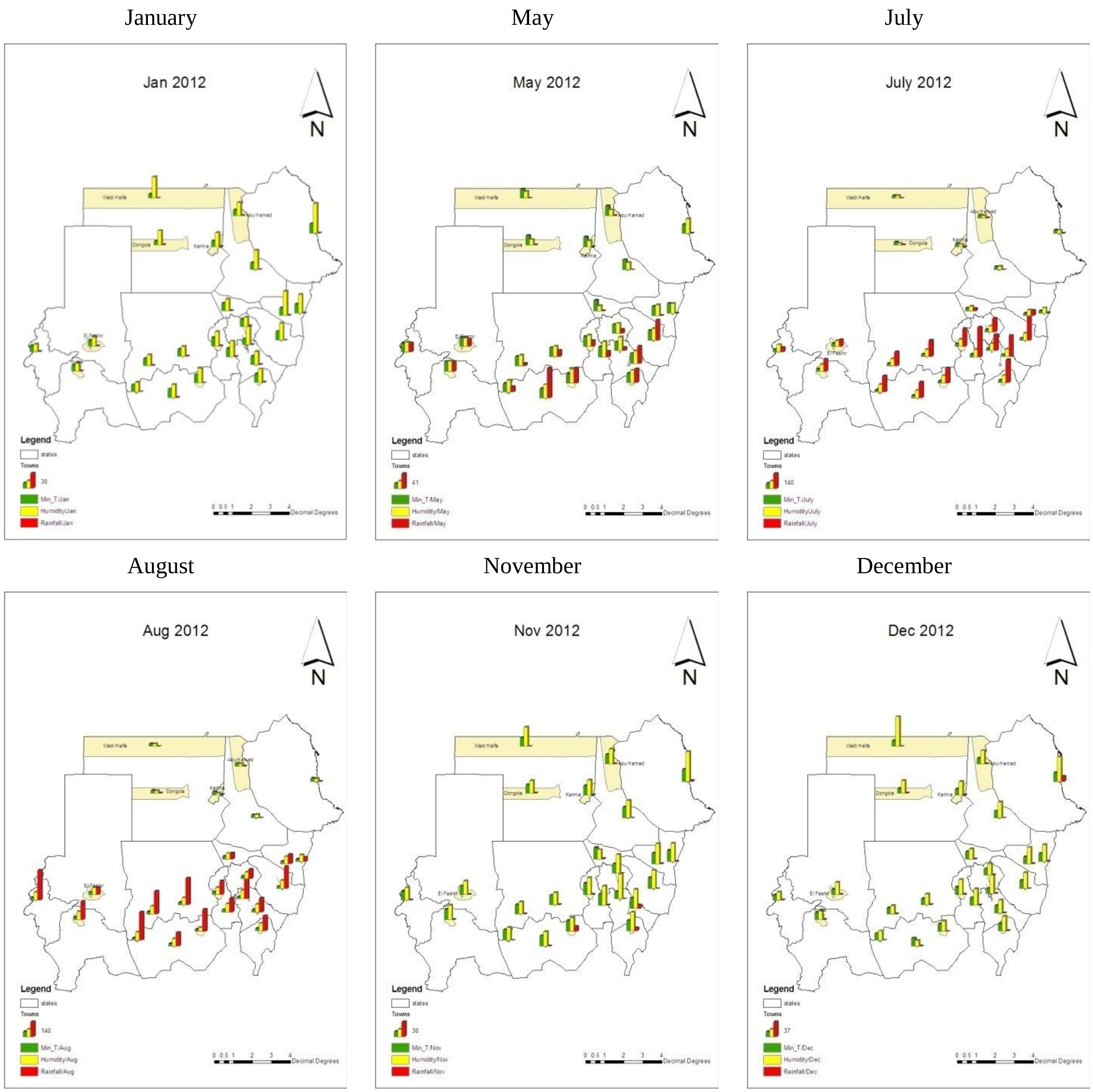


Figure 6.14. Comparison between average monthly temperature, humidity and rainfall on stations level for January, May, July, August, November and December 2012.

6.3 Discussions

In graduated color mapping, polygon feature is altered in color to reflect a particular value of the rainfall. Figure 6.1 shows rainfall on regions level for period 2000 – 2012 by using graduation by color, Kordofan region had the biggest amount of rainfall comparing with other regions but it had been decreased since 2003 there was a climatic change. In 2003 Darfur region recorded the highest rainfall amount sharing the Kordofan region however, that was exception and it generally recorded fewer amounts. For years 2008, 2009, 2010 and 2012 Central region shared Kordofan in the first order of rainfall amount. Northern region had the lowest rainfall amount in range 0.070667 - 3.44.

Figure 6.3 displays rainfall on states level for period 2000 – 2012 using graduation by color, generally Sinnar, Blue Nile, El Gadarif, South Kordofan and West Darfur had the highest rainfall amount, while Northern, River Nile, Red Sea had the lowest. A marked increase in the average amount of rain appeared in the years 2000 and 2005 in the North Darfur state.

Figure 6.5 illustrates rainfall on stations level for period 2000 – 2012 using graduation by color. Areas which appear in dark blue color (Wadi Halfa, Dongola, Karima, Atbara and Abu Hamad) acquired the lowest rainfall amount in range 0 - 5.541667. Towns such as El Fashir, Port Sudan, El Deweim and Kassala that represented by light blue in the most rainfall maps got amount of rain between 5.541668 and 25.141667. Green color which used to represent cities such as Kosti, El Deweim and Elobied covers the range 25.141668 – 39.3750. While the yellow color that represents the range 39.375001 – 54.40, appeared in some places, such as Elgeneina, Nyala, Kadugli and Babanusa. Areas that had the highest proportion of rain like El Gadarif, El damazen, Rashad, Abu Naama Babanusa and Kadugli been represented on the maps using the red color, which indicates the extent to 54.400001- 77.2250. Wadi Halfa is considered as desert areas of few rains but it showed a marked increase in the amount of rainfall in 2006. There is also a decrease in the amount of rain that fell in the town of Rashad in the years 2003, 2008 and 2010.

Figure 6.7 shows average monthly rainfall on stations level for year 2012 using graduation by color. In winter (November, December, January and February) there were little or no rain in most cities except for some stations such as Port Sudan, Rashad, Elobied, El Deweim, Babanusa El damazen and Abu Naama. In March and April a light rain (0.338463 – 7.661538) in the southern and western parts such as Nyala, Elgeneina, Elobied, Kadugli, Babanusa, El Gadarif, Kosti, El damazen, Abu Naama and Sinnar. In May and June moderate and heavy rain (27.623078 – 111.253846) in El damazen, Kadugli, El Gadarif, Rashad, Babanusa and Abu Naama. Beside few or no rain (0 – 11.830769) in the northern parts, Khartoum, El Fashir and Port Sudan. The months of July and August represent the peak of the fall, light rains (0.538462 – 21.6385) in each of Port Sudan, Khartoum, Wad Medani, and northern parts, while moderate rains (21.6386 – 104.115385) in El Fashir, El Deweim and Kassala, whilst heavy rains (104.115386 – 162.7) in Kosti, Sinnar, Elobied, El Nihood, Kadugli, Babanusa and Rashad and Nyala, however very heavy rains (162.700001 – 221.369231) in Elgeneina, El damazen, El Gadarif and Abu Naama. Only in January happened there was a marked increase in the rain pattern in Wadi Halfa station.

As shown in Figure 6.9 which appears comparison between average monthly rainfall on stations level for years 2003, 2007 and 2012 using graduation by color. There are constant patterns of rain, which often the possibility of their occurrence in the months July, August and September. In year 2003 there was very small quantity of rain in all months in Khartoum, Port Sudan and all towns of the northern region except Atbara, which it had more amount of rain in August. The highest rate of rain for the month of September was in Kadugli, Rashad, Wad Medani, El damazen and Sinnar. In 2007 it increased the amount of rain in Dongola, Karima and Khartoum, compared to 2003, as it rained in April in New Halfa, Khartoum, Wad Medani, Sinnar and Rashad. And also it rained in October in Wad Medani, El damazen, Abu Naama, Sinnar, Kadugli and Rashad. In year 2012 the amount of rain decreased significantly in Kassala, Khartoum and Northern region compared to 2007.

Figure 6.10 shows Comparison between average monthly temperature, humidity and rainfall on regions level for January, May, August and December 2012. In January temperature decreased, humidity increased and rainfall almost absent in all regions. While in May the temperature increased with rain in all regions except Northern and Khartoum. In August rains fell in all the regions and then returned to a decrease in December and absent except in the eastern region.

As shown in Figure 6.11, a comparison between average monthly rainfalls on regions level for year 2000, 2006, 2011 and 2012 been made. In the Northern region few or no rain in 2000 and 2012, little rain in July, August and September 2006 and a few rain in July 2011. The eastern region like other regions rainy season limited between May and October and up to a peak in July and August. Central, Kordofan and Darfur regions are considered the most rain compared to the rest, while we find that Khartoum and North are the least rain, but the eastern region is characterized medium rains. 2006 saw a clear change in the pattern of rainfall, increase of rain amount in all regions of Sudan.

Figure 6.12 compared between average monthly temperature, humidity and rainfall on states level for January, May, July, August, November and December 2012. In January rain absent from all States, while in May temperature increased, humidity decreased and rains began in most states except Northern, River Nile, Red Sea, Kassala and Khartoum. July and August period represent a fall season in most states except Northern, River Nile and the Red sea. In November and December temperature decreased and rain disappeared in most places except the Red Sea, Blue Nile, Sinnar, North Kordofan who witnessed rain in that period .

As shown in Figure 6.13 comparisons between average monthly rainfall on states level for year 2000, 2006 and 2012 been made. Rains were scarce or non-existent in the Nile River and the North, while rain was felt in the state of the Red Sea in October, November and December. It rained in other states in July, August and September, while it rained in April and May in El Gadarif, North Kordofan, South Kordofan, North Darfur, West Darfur, South Darfur, Sinnar and Blue Nile.

Figure 6.14 Compared between average monthly temperature, humidity and rainfall on stations level for years 2012. In January temperature decreased, humidly increased and no rain in all stations. In May temperature increased, humidity decreased and it rained in most stations except Dongola, Atbara, Abu Hamad, Karima, Wadi Halfa, Kassala, Port Sudan, New Halfa and Khartoum. While in July and August it rained in most stations except Dongola, Atbara, Abu Hamad, Karima and Port Sudan. Whilst in November and December humidity increased, temperature decreased and rain disappeared from all stations except Port Sudan, El damazen, Abu Naama and Rashad.

Also in graduated color mapping there is another technique known as graduation by size which used to produce the rainfall maps in Figures (6.2, 6.4, 6.6 and 6.8), polygon feature which used to represent stations, states and regions is altered in size to reflect the frequencies in the data. In this type of map, more than one incident at a given location is represented with a larger symbol.

Figure 6.2 shows rainfall on regions level for interval 2000 – 2012 using graduation by size. It can easily discriminate between the amounts of rain in the regions; we find that Northern region is least rain, followed by Khartoum, eastern and western regions respectively, finally Central and Kordofan comes as the two most rain regions.

Figure 6.4 displays rainfall maps on states level for interval 2000 – 2012 using graduation by size. It has been considered that states such as Northern, River Nile, Red Sea, Khartoum and North Darfur have the lowest rain in Sudan, while states such as El Gadarif, Sinnar, Blue Nile, South Kordofan, West Darfur and South Darfur have the highest amount of rain. In 2000 and 2010 Al Jazeera state recorded the lowest amount of rain compared with other years, while White Nile state obtained its minimum rain quantity in 2005.

In Figure 6.6 there is a comparison of rainfall on stations level for interval 2000 – 2012 using graduation by size. It appears that stations such as Dongola, Atbara, Abu Hamad, Karima and Wadi Halfa have few or no rain, while station such as Kassala, Khartoum, Port Sudan, and El Fashir recorded medium amount of rain, where stations like El Gadarif, El damazen, Abu Naama, Kadugli, Babanusa, Rashad, Nyala and Elgeneina recorded high amount of rain.

Figure 6.8 shows average monthly rainfall on states level for year 2012 using gradation by size. In January and February it rained few (0.000001 – 0.323077) in Khartoum, Eastern and Kordofan and no rain elsewhere. From March to June no or very few rain in Khartoum, Northern and Darfur, while a few and moderate rain in Central, Eastern region and Kordofan. Interval of July and August recorded few or medium (5.313846 – 88.656769) rain in Khartoum, Northern and Eastern and heavy and very heavy (88.656770 – 152.007652) rain in Central, Kordofan and Darfur. In September and October low rain (0.763077 – 13.384620) in Northern and Khartoum, while medium rain (13.384621 – 51.174350) in Darfur and Eastern, and high (51.174350 – 97.669231). In period of November and December no rain in Northern, Khartoum, Darfur and Central while very few (0.038453 – 0.736923) rain in Kordofan and few rain (0.736924 – 5.134615) in Eastern region.

6.4 Summary

Rainfall Maps proved that, they are very useful in analysis and comparisons based on rainfall data between different geographical areas, so as to its ability to display too much information in maps which link data with geographical location, that form of visualization is easy to read and understand. Also using rainfall maps enable us to work a large number of comparisons for further study and deeper understanding of the phenomenon.

From the spatial analysis of the phenomenon of rainfall and its distribution and based on the geographical location of stations, states and regions in Sudan we can come up with the following conclusions:

Rain is natural phenomenon do not adhere to a strict pattern in its occurrence and distribution, because it influenced by many variables in the atmosphere.

The interval From May to October represents the rainy season in most regions of Sudan and both July and August represent its peak, But In some stations such as Port Sudan, El damazen, Abu Naama and Rashad, few rains falls in the period from November to January.

Southern stations such as Elgeneina, Nyala, Kadugli, Babanusa, Rashad, El damazen, Abu Naama, Sinnar and El Gadarif are considered the areas which obtain the most amount of rainfall among the whole country, while stations such as Dongola, Atbara, Abu Hamad, Karima, Wadi Halfa, Port Sudan and Khartoum are considered the areas that obtain the least amount of rainfall. Other stations like Kassala, New Halfa, Wad Medani, El Deweim, Kosti, Elobied and El Nihood get a medium quantity of rainfall.

Some notable changes in the distribution pattern of rain appeared in some places for example, lack of rainfall in Kordofan since 2003. Also In 2003 Darfur region recorded the highest rainfall amount sharing Kordofan region, however, that was exception and it generally recorded fewer amounts.

7. Conclusions and Future Work

7.1 Introduction

Weather forecasting is a complex, inexact process and difficult to predict, which lends itself to computational intelligence techniques. This research showed that computational intelligence methods could successfully be used in addition to other methods for rainfall prediction.

In this thesis, we studied the rain fall phenomenon, which has been widely studied by several researchers working in different disciplines. We explored machine leaning techniques, ensemble methods, adaptive neuro-fuzzy system and geographic information system. We have applied all the above techniques to deal with rainfall prediction in Sudan.

This chapter summarizes the research work and the finding for this study. Section 7.2 provides the summary on the highlighted issues in this research. The essential contributions of weather forecasting particularly long-term rainfall prediction using computational intelligence and machine learning approaches are presented in Section 7.3. Section 7.4 gives the limitation and recommendation for future works of this research.

7.2 Thesis Summary

The meteorological data that used in this research has been brought from Central Bureau of Statistics, Sudan for 13 years from 2000 to 2012 for 24 meteorological stations over the country with 3732 total number of examples. These stations were: (Khartoum, Dongola, Atbara, Abu Hamad, Karima, Wadi Halfa, Wad Medani, El Deweim, Kassala, Port Sudan, El Gadarif, Elobied, El Nihood, Kadugli, Nyala, Elgeneina, El Fashir, Kosti, El damazen, New Halfa, Babanusa, Rashad, Abu Naama and Sinnar). The dataset had eight (8) attributes containing monthly averages data.

Statistical analysis for the data set has been made followed by data transformation then data preprocessing and data normalization.

To determine the most influencing and important variables that affect on the long term rainfall prediction out of the existing ones, many attributes evaluator algorithms such as (correlation based feature selection subset evaluator, Classifier subset evaluator, relief attribute evaluator and Wrapper subset evaluator) have been implemented with appropriate different search methods such as (best-first, evolutionary search, exhaustive search, genetic search, greedy stepwise, linear forward selection, PSO search, random search, scatter searchV1, subset size forward selection, Tabu Search and Ranker).

In the experiments we built 10 base algorithm models (Gaussian Processes, Linear Regression, Multilayer Perceptron, IBk, KStar, Decision Table, M5Rules, M5P, REP Tree and User Classifier.) and 7 Meta algorithms (Additive Regression, Bagging, Multi Scheme, Random Subset, Regression by Discretization, Stacking, and Vote) to predict rainfall. Then we compared between their results according to specific performance measure criteria and determined the best ones.

After that we proposed a new ensemble model (ensemble Vote +3 algorithms) for long term rainfall prediction. The new novel ensemble model has been constructed based of Meta classifier Vote combining with the best three base algorithms IBK, K-star and M5P. The models have been evaluated by using correlation coefficient; mean absolute error and root mean-squared error as performance metrics. Also we used both time taken to build the model and time taken to test model on supplied test set to compare and differentiate among the models. The results showed that the new novel ensemble method (ensemble Vote +3 algorithms) has the best performance comparing to basic, Meta algorithms and other ensemble models.

As an attempt to more improvement for outcomes we proposed Adaptive Neuro-fuzzy Inference Systems (ANFIS) to predict rainfall, because their learning and reasoning capabilities. We built several ANFIS models using different types of membership functions, different optimization methods and different dataset ratios for training and testing. The proposed models have been evaluated and compared by using correlation coefficient, mean absolute error and root mean squared error as performance metrics. The results show that ANFIS neuro-fuzzy produced satisfactory results, so it may be useful in long term rainfall prediction.

The empirical results indicate that ANFIS neuro-fuzzy and ensemble Vote+3 models are able to capture the dynamic behavior of the rainfall data and they may be useful in long term rainfall prediction. Their results converged dramatically, but ANFIS model outperformed the ensemble Vote+3model and produces better results. Also ANFIS model has ability to interpret and explain its results by using rules, and this feature is not available for other models

Finally we generated the rainfall maps, which produced results of the spatial analysis in different levels (towns, states and regions) in Sudan. In all maps polygon [275] has been used to represent regions, states and meteorological stations at town level. Graduation techniques [277], such as graduation by size and graduation by color have been utilized to analyze and compare the phenomenon of rain on different levels. Also analysis been made for rain at the same level for different months. Finally we analyzed rainfall with the most influenced variables such as temperature and relative humidity for different months and levels.

7.3 Research Contributions

In this research, the proposed objectives have been achieved:

1. The most influencing variables (Date, Minimum Temperature, Humidity, and Wind Direction) that affect on the long-term rainfall prediction out of the 7 variables have been specified.
2. We have presented a formulation of the adaptive neuro-fuzzy inference system (ANFIS) model for long-term rainfall prediction and we concluded that:
   - Different types of membership functions have been used to decide which one is the most appropriate for the target, the results showed that the model which used generalized bell shaped membership outperforms the others and achieve the highest performance accuracy.
   - Dividing dataset into 70-30 for training and testing respectively considered the best choice for dividing our dataset, since it provided the best results comparing with the other choices for both training and testing phases.
   - Our experiments ensured that the hybrid learning method is much better than back propagation method as learning algorithm for ANFIS model and it gave better and more accurate results.

3. We developed a long-term rainfall prediction model by using computational intelligence techniques and machine learning approaches for different stations in Sudan and compared them to determine which one is the most accurate and has highest performance. Moreover, this research proposed both an Adaptive Neuro-fuzzy System (ANFIS) and a novel ensemble method for long-term rainfall prediction. The experimental results indicate that ANFIS neuro-fuzzy and ensemble Vote+3 algorithms are able to capture the dynamic behavior of the rainfall data and they may be useful in long-term rainfall prediction. The results of two proposed models are converged, but ANFIS model outperformed the ensemble Vote+3 algorithms and produces better results. Also ANFIS model has the ability to interpret and explain its results by using rules such as:

   ***If* (Date is medium) and (Temperature is high) and (Humidity is high) and (Wind Direction is low) *Then* (Rainfall is very low)**
4. We performed the spatial analysis of rainfall phenomenon in Sudan for period 2000-2012 and produced rainfall maps. Monthly and annual analysis has been made for different levels (stations, states and regions). Analysis was made for rain at the same level for different months. Finally we analyzed rainfall with the most influenced variables such as temperature and relative humidity for different months and levels. Rainfall Maps proved that, they are very useful in analysis and comparisons based on rainfall data between different geographical areas, so as to its ability to display too much information in maps which link data with geographical location, that form of visualization is easy to read and understand. Also using rainfall maps enable us to work a large number of comparisons for further study and deeper understanding of the phenomenon.

7.4 Limitations and Future Work

The data set which has been used in this research that brought from central bureau of statistics Sudan were monthly averages, and we failed to obtain the daily rainfall data from the meteorological body or anywhere else. If it possible, to train the proposed models in this study on the daily rainfall data, that will produce daily rainfall prediction models, thus short-term prediction of rain for each station in Sudan has been available.

From other side, in this study we considered only seven predictors for rainfall prediction. The predictors were date, minimum temperature, maximum temperature, relative humidity, wind speed and wind direction. If we use some more climate factors such as atmosphere pressure, sea surface temperature, etc, so we may obtain more accurate predictions.

References

[1] J. R. Holton and G. J. Hakim, *An introduction to dynamic meteorology* vol. 88: Academic press, 2012.

[2] P. Jagannathan and B. Parthasarathy, "Trends and periodicities of rainfall over India," *Monthly Weather Review*, vol. 101, pp. 371-375, 1973.

[3] P. Lynch, "The origins of computer weather prediction and climate modeling," *Journal of Computational Physics*, vol. 227, pp. 3431-3444, 2008.

[4] A. Ajith, P. Sajith, and J. Babu, "Will We Have a Wet Summer? Long-term Rain Forecasting Using Soft Computing Models," *Modeling and Simulation*, pp. 1044-1048, 2001.

[5] R. Kimura, "Numerical weather prediction," *Journal of Wind Engineering and Industrial Aerodynamics*, vol. 90, pp. 1403-1414, 2002.

[6] G. Shrivastava, S. Karmakar, M. K. Kowar, and P. Guhathakurta, "Application of artificial neural networks in weather forecasting: a comprehensive literature review," *International Journal of Computer Applications (0975–8887) Volume*, 2012.

[7] A. P. Engelbrecht, *Computational intelligence: an introduction*: John Wiley & Sons, 2007.

[8] A. Konar, *Computational intelligence: principles, techniques and applications*: Springer Science & Business Media, 2006.

[9] M. Negnevitsky, *Artificial intelligence: a guide to intelligent systems*, second ed.: Pearson Education, 2005.

[10] I. H. Witten and E. Frank, *Data Mining: Practical machine learning tools and techniques*, second ed.: Morgan Kaufmann, 2005.

[11] A. K. Pujari, *Data mining techniques*: Universities press, 2001.

[12] N. Bushara and A. Abraham, "Weather Forecasting in Sudan Using Machine Learning Schemes," *Journal of Network and Innovative Computing*, vol. 2, pp. 309-317, 2014.

[13] T. Mitchell, "Generalization as search," *Artificial Intelligence*, vol. 18, pp. 203–226, 1982.

[14] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, pp. 1-39, 2010.

[15] G. Seni and J. F. Elder, "Ensemble methods in data mining: improving accuracy through combining predictions," *Synthesis Lectures on Data Mining and Knowledge Discovery*, vol. 2, pp. 1-126, 2010.

[16] K. Tumer and J. Ghosh, "Error correlation and error reduction in ensemble classifiers," *Connection science*, vol. 8, pp. 385-404, 1996.

[17] A. Krogh and J. Vedelsby, "Neural network ensembles, cross validation, and active learning," *Advances in neural information processing systems*, vol. 7, pp. 231-238, 1995.

[18] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*: John Wiley & Sons, 2004.

[19] L. I. Kuncheva and C. J. Whitaker, "Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy," *Machine learning*, vol. 51, pp. 181-207, 2003.

[20] X. Hu, "Using rough sets theory and database operations to construct a good ensemble of classifiers for data mining applications," in *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, 2001, pp. 233-240.
[21] G. Brown, J. Wyatt, R. Harris, and X. Yao, "Diversity creation methods: a survey and categorisation," *Information Fusion*, vol. 6, pp. 5-20, 2005.
[22] R. E. Banfield, L. O. Hall, K. W. Bowyer, and W. P. Kegelmeyer, "A comparison of decision tree ensemble creation techniques," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 173-180, 2007.
[23] A. K. Palit and D. Popovic, *Computational intelligence in time series forecasting: theory and engineering applications*: Springer Science & Business Media, 2006.
[24] T. Takagi and M. Sugeno, "Fuzzy identification of systems and its applications to modeling and control," *Systems, Man and Cybernetics, IEEE Transactions on*, pp. 116-132, 1985.
[25] J.-S. R. Jang, "ANFIS: adaptive-network-based fuzzy inference system," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 23, pp. 665-685, 1993.
[26] J.-S. R. Jang and C.-T. Sun, "Neuro-fuzzy modeling and control," *Proceedings of the IEEE*, vol. 83, pp. 378-406, 1995.
[27] J.-S. R. Jang, C.-T. Sun, and E. Mizutani, "Neuro-fuzzy and soft computing; a computational approach to learning and machine intelligence," 1997.
[28] C.-T. Lin and C. G. Lee, "Neural-network-based fuzzy logic control and decision system," *Computers, IEEE Transactions on*, vol. 40, pp. 1320-1336, 1991.
[29] C.-T. Lin, *Neural fuzzy systems: a neuro-fuzzy synergism to intelligent systems*: Prentice hall PTR, 1996.
[30] T. E. Alhanafy, F. Zaghlool, and A. S. E. D. Moustafa, "Neuro fuzzy modeling scheme for the prediction of air pollution," *Journal of American Science*, vol. 6, pp. 605-616, 2010.
[31] T. Nazmy, H. El-Messiry, and B. Al-Bokhity, "Adaptive neuro-fuzzy inference system for classification of ECG signals," in *Informatics and Systems (INFOS), 2010 The 7th International Conference on*, 2010, pp. 1-6.
[32] A. Sengur, "An expert system based on linear discriminant analysis and adaptive neuro-fuzzy inference system to diagnosis heart valve diseases," *Expert Systems with Applications*, vol. 35, pp. 214-222, 2008.
[33] O. Špakov and D. Miniotas, "Visualization of eye gaze data using heat maps," *Elektronika ir Elektrotechnika*, vol. 74, pp. 55-58, 2015.
[34] J.-H. Shin, S. Blay, B. McNeney, and J. Graham, "LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms," *Journal of Statistical Software*, vol. 16, pp. 1-10, 2006.
[35] A. Sopan, A. S.-I. Noh, S. Karol, P. Rosenfeld, G. Lee, and B. Shneiderman, "Community Health Map: A geospatial and multivariate data visualization tool for public health datasets," *Government Information Quarterly*, vol. 29, pp. 223-234, 2012.
[36] W. S. Cleveland, C. S. Harris, and R. McGill, "Human Factors and Behavioral Science: Experiments on Quantitative Judgments of Graphs and Maps," *Bell System Technical Journal*, vol. 62, pp. 1659-1674, 1983.
[37] S. Silva, B. S. Santos, and J. Madeira, "Using color in visualization: A survey," *Computers & Graphics*, vol. 35, pp. 320-333, 2011.

[38] M. Harrower and C. A. Brewer, "ColorBrewer. org: an online tool for selecting colour schemes for maps," *The Cartographic Journal,* vol. 40, pp. 27-37, 2003.
[39] S. H. Alvi, "Climatic changes, desertification and the Republic of Sudan," *GeoJournal,* vol. 33, pp. 393-399, 1994.
[40] R. Pankhurst, *The Ethiopian borderlands: Essays in regional history from ancient times to the end of the 18th century*: The Red Sea Press, 1997.
[41] C. B. o. Statistics, "Statistical monitoring 2004-2013," December 2014.
[42] C. B. o. Statistics, "Statistical monitoring 2003-2012," December 2013.
[43] C. B. o. Statistics, "2008 Census Final Tabulation Sudan labour force, table-2.," 2008.
[44] C. B. o. Statistics, "2008 Census Final Tabulation, northern Sudan labour force, table-2," 2008.
[45] C. B. o. Statistics, " Census Final Tabulation Northern States, Economic and Social characteristics, Table E3," 1993.
[46] N. Anbarci, J. Boyd, E. Floehr, J. Lee, and J. J. Song, "Population and income sensitivity of private and public weather forecasting," *Regional science and urban economics,* vol. 41, pp. 124-133, 2011.
[47] M. Tektaş, "Weather forecasting using ANFIS and ARIMA models," *Environmental Research, Engineering and Management,* vol. 51, pp. 5-10, 2010.
[48] S. Asmita and K. Shukla, "Review on the Architecture, Algorithm and Fusion Strategies in Ensemble Learning," *International Journal of Computer Applications,* vol. 108, 2014.
[49] S. Lee, S. Cho, and P. M. Wong, "Rainfall prediction using artificial neural networks," *Journal of Geographic Information and Decision Analysis,* vol. 2, pp. 233-242, 1998.
[50] Y. Tang, W. W. Hsieh, B. Tang, and K. Haines, "A neural network atmospheric model for hybrid coupled modelling," *Climate dynamics,* vol. 17, pp. 445-455, 2001.
[51] C. W. Dawson and R. Wilby, "An artificial neural network approach to rainfall-runoff modelling," *Hydrological Sciences Journal,* vol. 43, pp. 47-66, 1998.
[52] R. M. Trigo and J. P. Palutikof, "Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach," *Climate Research,* vol. 13, pp. 45-59, 1999.
[53] E. Toth, A. Brath, and A. Montanari, "Comparison of short-term rainfall prediction models for real-time flood forecasting," *Journal of Hydrology,* vol. 239, pp. 132-147, 2000.
[54] K. C. Luk, J. E. Ball, and A. Sharma, "An application of artificial neural networks for rainfall forecasting," *Mathematical and Computer modelling,* vol. 33, pp. 683-693, 2001.
[55] F.-J. Chang, J.-M. Liang, and Y.-C. Chen, "Flood forecasting using radial basis function neural networks," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 31, pp. 530-535, 2001.
[56] S. C. Michaelides, C. S. Pattichis, and G. Kleovoulou, "Classification of rainfall variability by using artificial neural networks," *International Journal of Climatology,* vol. 21, pp. 1401-1414, 2001.
[57] I. Maqsood, M. R. Khan, and A. Abraham, *Weather forecasting models using ensembles of neural networks*: Springer, 2003.
[58] A. J. Cannon and P. H. Whitfield, "Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models," *Journal of Hydrology,* vol. 259, pp. 136-151, 2002.

[59] D.-I. Jeong and Y.-O. Kim, "Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction," *Hydrological processes,* vol. 19, pp. 3819-3835, 2005.
[60] V. Somvanshi, O. Pandey, P. Agrawal, N. Kalanker, M. R. Prakash, and R. Chand, "Modelling and prediction of rainfall using artificial neural network and ARIMA techniques," *J. Ind. Geophys. Union,* vol. 10, pp. 141-151, 2006.
[61] M. Janga Reddy and R. Maity, "Regional rainfall forecasting using large scale climate teleconnections and artificial intelligence techniques," *Journal of Intelligent Systems,* vol. 16, pp. 307-322, 2007.
[62] R. Bustami, N. Bessaih, C. Bong, and S. Suhaili, "Artificial neural network for precipitation and water level predictions of Bedup River," *IAENG International journal of computer science,* vol. 34, pp. 228-233, 2007.
[63] M. Hayati and Z. Mohebi, "Application of artificial neural networks for temperature forecasting," *World Academy of Science, Engineering and Technology,* vol. 28, pp. 275-279, 2007.
[64] A. Aytek and M. Alp, "An application of artificial intelligence for rainfall-runoff modeling," *Journal of Earth System Science,* vol. 117, pp. 145-155, 2008.
[65] F. O. Hocaoglu, Y. Oysal, and M. Kurban, "Missing wind data forecasting with adaptive neuro-fuzzy inference system," *Neural Computing and Applications,* vol. 18, pp. 207-212, 2009.
[66] K. Solaimani, "Rainfall-runoff prediction based on artificial neural network (A Case Study: Jarahi Watershed)," *American-Eurasian J. Agric. & Environ. Sci,* vol. 5, pp. 856-865, 2009.
[67] A. Y. Shamseldin, "Application of a neural network technique to rainfall-runoff modelling," *Journal of Hydrology,* vol. 199, pp. 272-294, 1997.
[68] P. C. Nayak, K. Sudheer, D. Rangan, and K. Ramasastri, "A neuro-fuzzy computing technique for modeling hydrological time series," *Journal of Hydrology,* vol. 291, pp. 52-66, 2004.
[69] A. Abraham, N. S. Philip, and P. Mahanti, "Soft computing models for weather forecasting," *International Journal of Applied Science and Computations,* vol. 11, pp. 106-117, 2004.
[70] 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, pp. 115-125, 2005.
[71] P. Luenam, S. Ingsriswang, L. Ingsrisawang, P. Aungsuratana, and W. Khantiyanan, "A Neuro-Fuzzy Approach for Daily Rainfall Prediction over the Central Region of Thailand," in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, 2010.
[72] E. Vamsidhar, K. Varma, P. S. Rao, and R. Satapati, "Prediction of rainfall using backpropagation neural network model," *International Journal on Computer Science and Engineering,* vol. 2, pp. 1119-1121, 2010.
[73] C. Patil and A. Ghatol, "Rainfall forecasting using local parameters over a meteorological station: an artificial neural network approach," *International Journal of Engineering Research & Industrial Applications,* vol. 3, pp. 341-356, 2010.
[74] J. Joshi and V. M. Patel, "Rainfall-Runoff Modeling Using Artificial Neural Network (A Literature Review)," in *National Conference on Recent Trends in Engineering & Technology, BVM Engineering College, VV Nagar, Gujarat, India*, 2011, pp. 13-14.

[75] A. El-Shafie, A. Noureldin, M. Taha, and A. Hussain, "Dynamic versus static neural network model for rainfall forecasting at Klang River Basin, Malaysia," *Hydrology and Earth System Sciences,* vol. 8, pp. 6489-6532, 2011.

[76] D. Sawaitul, K. Wagh, and P. Chatur, "Classification and prediction of future weather by using back propagation algorithm-an approach," *International Journal of Emerging Technology and Advanced Engineering,* vol. 2, pp. 110-113, 2012.

[77] M. Nelson, T. Hill, B. Remus, and M. O’Connor, "Can neural networks be applied to time series forecasting and learn seasonal patterns: An empirical investigation," in *Twenty seventh Annual Hawaii International Conference on System Sciences,* Hawaii 1994, pp. 649–655.

[78] R. Sharda and R. B. Patil, "Connectionist approach to time series prediction: an empirical test," *Journal of Intelligent Manufacturing,* vol. 3, pp. 317-323, 1992.

[79] S. Y. Kang, "An investigation of the use of feedforward neural networks for forecasting," 1992.

[80] T. Hill, L. Marquez, M. O'Connor, and W. Remus, "Artificial neural network models for forecasting and decision making," *International Journal of Forecasting,* vol. 10, pp. 5-15, 1994.

[81] T. Hill, M. O’Connor, and W. Remus, "Neural network models for time series forecasts," *Management Sciences,* vol. 42, pp. 1082– 1092, 1996.

[82] K. Abhishek, A. Kumar, R. Ranjan, and S. Kumar, "A rainfall prediction model using artificial neural network," in *Control and System Graduate Research Colloquium (ICSGRC), 2012 IEEE*, 2012, pp. 82-87.

[83] A. Mohammad, "Weather Temperature Forecasting Using Artificial Neural Network," *Journal of Engineering and Development,* vol. 15, 2011.

[84] S. De and A. Debnath, "Artificial neural network based prediction of maximum and minimum temperature in the summer monsoon months over India," *Applied Physics Research,* vol. 1, p. p37, 2009.

[85] S. S. Baboo and I. K. Shereef, "An efficient weather forecasting system using artificial neural network," *International journal of environmental science and development,* vol. 1, pp. 2010-0264, 2010.

[86] C. J. Devi, B. S. P. Reddy, K. V. Kumar, B. M. Reddy, and N. R. Nayak, "ANN approach for weather prediction using back propagation," *International Journal of Engineering Trends and Technology,* vol. 3, 2012.

[87] M. K. Mayilvaganan and K. Naidu, "ANN and Fuzzy Logic Models for the Prediction of groundwater level of a watershed," *International Journal on Computer Science and Engineering,* vol. 3, pp. 2523-2530, 2011.

[88] A. El-Shafie, M. Mukhlisin, A. A. Najah, and M. Taha, "Performance of artificial neural network and regression techniques for rainfall-runoff prediction," *International Journal of the Physical Sciences*, vol. 6, pp. 1997-2003, 2011.

[89] D. Lekkas, C. Onof, M. Lee, and E. Baltas, "Application of artificial neural networks for flood forecasting," *Global Nest Journal,* vol. 6, pp. 205-211, 2004.

[90] G. Geetha and R. S. Selvaraj, "Prediction of monthly rainfall in Chennai using back propagation neural network model," *International Journal of Engineering Science and Technology,* vol. 3, 2011.

[91] A. R. Naik and S. Pathan, "Weather Classification and forecasting using Back Propagation Feed-forward Neural Network," *International Journal of Scientific and Research Publications,* vol. 2, 2012.

[92] G. Tiron and S. Gosav, "The July 2008 rainfall estimation from Barnova WSR-98 D radar using artificial neural network," *Romanian Reports in Physics,* vol. 62, pp. 405-413, 2010.

[93] Ö. A. Dombaycı and M. Gölcü, "Daily means ambient temperature prediction using artificial neural network method: A case study of Turkey," *Renewable Energy,* vol. 34, pp. 1158-1161, 2009.

[94] M. Afzali, A. Afzali, and G. Zahedi, "Ambient Air Temperature Forecasting Using Artificial Neural Network Approach," in *International Conference on Environmental and Computer Science IPCBEE,* 2011, pp. 176-180.

[95] M. Trejo-Perea, G. Herrera-Ruiz, J. Rios-Moreno, R. C. Miranda, and E. Rivasaraiza, "Greenhouse energy consumption prediction using neural networks models," *training,* vol. 1, p. 2, 2009.

[96] S. M. Paras, A. Kumar, and M. Chandra, "A feature based neural network model for weather forecasting," *International Journal of Computational Intelligence,* vol. 4, 2009.

[97] T. Santhanam and A. Subhajini, "An efficient weather forecasting system using radial basis function neural network," *Journal of Computer Science,* vol. 7, p. 962, 2011.

[98] H. Abdul-Kader, "Neural Networks Training Based on Differential Evolution Algorithm Compared with Other Architectures for Weather Forecasting34," *IJCSNS,* vol. 9, p. 92, 2009.

[99] I. Maqsood, M. R. Khan, and A. Abraham, "Intelligent weather monitoring systems using connectionist models," *NEURAL PARALLEL AND SCIENTIFIC COMPUTATIIONS,* vol. 10, pp. 157-178, 2002.

[100] T. A. Awchi, "Application of radial basis function neural networks for reference evapotranspiration prediction," *Al-Rafidain Engineering,* vol. 16, pp. 117-130, 2008.

[101] G.-F. Lin and L.-H. Chen, "A non-linear rainfall-runoff model using radial basis function network," *Journal of Hydrology,* vol. 289, pp. 1-8, 2004.

[102] T. Chow and S. Cho, "Development of a recurrent sigma-pi neural network rainfall forecasting system in Hong Kong," *Neural Computing & Applications,* vol. 5, pp. 66-75, 1997.

[103] R. P. Deshmukh and A. Ghatol, "Modular neural network approach for short term flood forecasting a comparative study," *International Journal,* 2010.

[104] X. L. Gong, Z. J. Hu, M. L. Zhang, and H. Wang, "Wind Power Forecasting Using Wavelet Decomposition and Elman Neural Network," in *Advanced Materials Research*, 2013, pp. 628-632.

[105] C. Meng and J. Wu, "A novel nonlinear neural network ensemble model using K-PLSR for rainfall forecasting," in *Bio-Inspired Computing and Applications*, ed: Springer, 2012, pp. 41-48.

[106] E. C. Özelkan, F. Ni, and L. Duckstein, "Relationship between monthly atmospheric circulation patterns and precipitation: Fuzzy logic and regression approaches," *Water resources research,* vol. 32, pp. 2097-2103, 1996.

[107] H. Liu and V. Chandrasekar, "Classification of hydrometeor type based on multiparameter radar measurements- Development of a neuro-fuzzy system and in-situ

verification," in *International Conference on Radar Meteorology, 29 th, Montreal, Canada*, 1999, pp. 172-175.
[108] C. Mahabir, F. Hicks, and A. R. Fayek, "Application of fuzzy logic to forecast seasonal runoff," *Hydrological processes,* vol. 17, pp. 3749-3762, 2003.
[109] D.-H. Bae, D. M. Jeong, and G. Kim, "Monthly dam inflow forecasts using weather forecasting information and neuro-fuzzy technique," *Hydrological Sciences Journal,* vol. 52, pp. 99-113, 2007.
[110] R. A. Aliev, B. Fazlollahi, R. R. Aliev, and B. Guirimov, "Linguistic time series forecasting using fuzzy recurrent neural network," *Soft Computing,* vol. 12, pp. 183-190, 2008.
[111] L. Jiang and J. Wu, "Hybrid PSO and GA for Neural Network Evolutionary in Monthly Rainfall Forecasting," in *ACIIDS (1)*, 2013, pp. 79-88.
[112] D. M. Riley and G. K. Venayagamoorthy, "Comparison of a recurrent neural network PV system model with a traditional component-based PV system model," in *Photovoltaic Specialists Conference (PVSC), 2011 37th IEEE*, 2011, pp. 002426-002431.
[113] S. Zhao and L. Wang, "The model of rainfall forecasting by support vector regression based on particle swarm optimization algorithms," in *Life System Modeling and Intelligent Computing*, ed: Springer, 2010, pp. 110-119.
[114] I. S. Isa, S. Omar, Z. Saad, N. M. Noor, and M. K. Osman, "Weather forecasting using photovoltaic system and neural network," in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2010 Second International Conference on*, 2010, pp. 96-100.
[115] M. Lunagariya, S. Mishra, and V. Pandey, "Verification and usability of medium range weather forecast for Anand region," *Journal of Agrometeorology*, vol. 11, pp. 228-233, 2009.
[116] A. P. Dawid, "Present position and potential developments: Some personal views: Statistical theory: The prequential approach," *Journal of the Royal Statistical Society. Series A (General),* pp. 278-292, 1984.
[117] H. R. Glahn and D. A. Lowry, "The use of model output statistics (MOS) in objective weather forecasting," *Journal of applied meteorology,* vol. 11, pp. 1203-1211, 1972.
[118] S.-F. Su, C.-B. Lin, and Y.-T. Hsu, "A high precision global prediction approach based on local prediction approaches," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 32, pp. 416-425, 2002.
[119] M. Kannan, S. Prabhakaran, and P. Ramachandran, "Rainfall forecasting using data mining technique," 2010.
[120] E. Rupa and S. Jain, "Neuro-Fuzzy Based Region Rainfall Forecasting."
[121] A. Kusiak, X. Wei, A. P. Verma, and E. Roz, "Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach," *Geoscience and Remote Sensing, IEEE Transactions on,* vol. 51, pp. 2337-2342, 2013.
[122] T. Sivaramakrishnan and S. Meganathan, "Point Rainfall Prediction using Data Mining Technique," *Res. Journal of Applied Sciences, Engineering and Technology,* vol. 4, pp. 1899-1902, 2012.
[123] E. G. Petre, "A decision tree for weather prediction," *PP,* vol. 77, p. 82, 2009.
[124] L. Ingsrisawang, S. Ingsriswang, S. Somchit, P. Aungsuratana, and W. Khantiyanan, "Machine learning techniques for short-term rain forecasting system in the northeastern

part of Thailand," in *Proceedings of World Academy of Science, Engineering and Technology*, 2008.

[125] N. KHANDELWAL and R. DAVEY, "Climatic Assessment Of Rajasthan's Region For Drought With Concern Of Data Mining Techniques," *Concern*, vol. 2, 2012.

[126] D. Chauhan and J. Thakur, "Data Mining Techniques for Weather Prediction: A Review."

[127] G. Shoba and G. Shobha, "Rainfall Prediction Using Data Mining techniques: A Survey," *Journal of Engineering and Computer Science*, vol. 3, pp. 6206-6211, 2014.

[128] F. Olaiya and A. B. Adeyemo, "Application of data mining techniques in weather prediction and climate change studies," *International Journal of Information Engineering and Electronic Business (IJIEEB)*, vol. 4, p. 51, 2012.

[129] N. Sethi and D. K. Garg, "Exploiting Data Mining Technique for Rainfall prediction," *International Journal of Computer Science and Information Technologies,* vol. 5, pp. 3982-3984, 2014.

[130] N. Singhrattna, B. Rajagopalan, M. Clark, and K. Kumar, "Seasonal forecasting of Thailand summer monsoon rainfall," *International Journal of Climatology*, vol. 25, pp. 649-664, 2005.

[131] W. T. Zaw and T. T. Naing, "Empirical statistical modeling of rainfall prediction over Myanmar," *World Academy of Science, Engineering and Technology,* vol. 2, pp. 500-504, 2008.

[132] J. Kajornrit, K. W. Wong, and C. C. Fung, "Rainfall prediction in the northeast region of Thailand using modular fuzzy inference system," in *Fuzzy Systems (FUZZ-IEEE), 2012 IEEE International Conference on*, 2012, pp. 1-6.

[133] V. B. Nikam and B. Meshram, "Modeling Rainfall Prediction Using Data Mining Method," in *Fifth International Conference on Computational Intelligence, Modeling and Simulation*, 2013.

[134] S.-Y. Ji, S. Sharma, B. Yu, and D. H. Jeong, "Designing a rule-based hourly rainfall prediction model," in *Information Reuse and Integration (IRI), 2012 IEEE 13th International Conference on*, 2012, pp. 303-308.

[135] J. W. Tukey, "Exploratory data analysis," 1977.

[136] L. K. Hansen and P. Salamon, "Neural network ensembles," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 993-1001, 1990.

[137] R. E. Schapire, "The strength of weak learnability," *Machine learning,* vol. 5, pp. 197-227, 1990.

[138] Y. Freund and R. E. Schapire, "Experiments with a new boosting algorithm," in *ICML*, 1996, pp. 148-156.

[139] S. Merler, B. Caprile, and C. Furlanello, "Parallelizing AdaBoost by weights dynamics," *Computational statistics & data analysis,* vol. 51, pp. 2487-2498, 2007.

[140] C.-X. Zhang and J.-S. Zhang, "A local boosting algorithm for solving classification problems," *Computational Statistics & Data Analysis*, vol. 52, pp. 1928-1941, 2008.

[141] M. Alhamdoosh and D. Wang, "Fast decorrelated neural network ensembles with random weights," *Information Sciences,* vol. 264, pp. 104-117, 2014.

[142] J. T. DeWeber and T. Wagner, "A regional neural network ensemble for predicting mean daily river water temperature," *Journal of Hydrology,* vol. 517, pp. 187-200, 2014.

[143] L. Li, Q. Hu, X. Wu, and D. Yu, "Exploration of classification confidence in ensemble learning," *Pattern Recognition,* vol. 47, pp. 3120-3131, 2014.

[144] L. Zhang and P. N. Suganthan, "Random forests with ensemble of feature spaces," *Pattern Recognition,* vol. 47, pp. 3429-3437, 2014.
[145] A. S. M. Salih and A. Abraham, "Novel Ensemble Decision Support and Health Care Monitoring System," *Journal of Network and Innovative Computing,* vol. 2, pp. 041-051.
[146] L. Li, R. Stolkin, L. Jiao, F. Liu, and S. Wang, "A compressed sensing approach for efficient ensemble learning," *Pattern Recognition,* vol. 47, pp. 3451-3465, 2014.
[147] W.-C. Chen, L.-Y. Tseng, and C.-S. Wu, "A unified evolutionary training scheme for single and ensemble of feedforward neural network," *Neurocomputing,* vol. 143, pp. 347-361, 2014.
[148] L. Rokach, A. Schclar, and E. Itach, "Ensemble methods for multi-label classification," *Expert Systems with Applications,* vol. 41, pp. 7507-7523, 2014.
[149] A. M. Canuto, M. C. Abreu, L. de Melo Oliveira, J. C. Xavier, and A. d. M. Santos, "Investigating the influence of the choice of the ensemble members in accuracy and diversity of selection-based and fusion-based methods for ensembles," *Pattern recognition letters,* vol. 28, pp. 472-486, 2007.
[150] V. Bolón-Canedo, N. Sánchez-Maroño, and A. Alonso-Betanzos, "Data classification using an ensemble of filters," *Neurocomputing,* vol. 135, pp. 13-20, 2014.
[151] M. Jin, R. Li, Z. Xu, and X. Zhao, "Reliable fault diagnosis method using ensemble fuzzy ARTMAP based on improved Bayesian belief method," *Neurocomputing,* vol. 133, pp. 309-316, 2014.
[152] D. S. Nascimento, A. L. Coelho, and A. M. Canuto, "Integrating complementary techniques for promoting diversity in classifier ensembles: A systematic study," *Neurocomputing,* vol. 138, pp. 347-357, 2014.
[153] Q. Hu, L. Li, X. Wu, G. Schaefer, and D. Yu, "Exploiting diversity for optimizing margin distribution in ensemble learning," *Knowledge-Based Systems,* vol. 67, pp. 90-104, 2014.
[154] G. Timms, A. Turnbull, and A. Rahman, "Ensemble aggregation methods for relocating models of rare events," *Engineering Applications of Artificial Intelligence,* vol. 34, pp. 58-65, 2014.
[155] N. Kourentzes, D. K. Barrow, and S. F. Crone, "Neural network ensemble operators for time series forecasting," *Expert Systems with Applications,* vol. 41, pp. 4235-4244, 2014.
[156] X.-C. Yin, K. Huang, C. Yang, and H.-W. Hao, "Convex ensemble learning with sparsity and diversity," *Information Fusion,* vol. 20, pp. 49-59, 2014.
[157] J.-F. Díez-Pastor, C. García-Osorio, and J. J. Rodríguez, "Tree ensemble construction using a GRASP-based heuristic and annealed randomness," *Information Fusion,* vol. 20, pp. 189-202, 2014.
[158] H. S. Own and A. Abraham, "A Novel-weighted Rough Set-based Meta Learning for Ozone Day Prediction," *Acta Polytechnica Hungarica,* vol. 11, 2014.
[159] W. Pedrycz, *Fuzzy sets engineering*: CRC press, 1995.
[160] M. M. Gupta, "Fuzzy neural networks: theory and applications," in *Photonics for Industrial Applications*, 1994, pp. 303-325.
[161] A. D. Kulkarni and C. D. Cavanaugh, "Fuzzy neural network models for classification," *Applied Intelligence,* vol. 12, pp. 207-215, 2000.
[162] M. Brown and C. J. Harris, "Neurofuzzy adaptive modelling and control," 1994.
[163] E. H. Mamdani and S. Assilian, "An experiment in linguistic synthesis with a fuzzy logic controller," *International journal of man-machine studies,* vol. 7, pp. 1-13, 1975.

[164] F. Castellanos and N. James, "Average hourly wind speed forecasting with ANFIS," in *11th American Conference on Wind Engineering, San Juan, Puerto Rico (June 2009),* 2009.
[165] M. Sharma, "Seasonal Weather Parameters Forecasting Using Soft Computing Techniques," *EXCEL INDIA PUBLISHERS NEW DELHI,* p. 328, 2014.
[166] N. R. Babu, C. B. A. Babu, D. P. Reddy, and M. Gowtham, "Comparison of ANFIS and ARIMA Model for Weather forecasting," *Indian Journal of Science and Technology,* vol. 8, pp. 70-73, 2015.
[167] K. Ramesh, A. Kesarkar, J. Bhate, M. Venkat Ratnam, and A. Jayaraman, "Adaptive neuro-fuzzy inference system for temperature and humidity profile retrieval from microwave radiometer observations," *Atmospheric Measurement Techniques,* vol. 8, pp. 369-384, 2015.
[168] M. K. Mayilvaganan and K. Naidu, "Comparison of membership functions in adaptive-network-based fuzzy inference system (ANFIS) for the prediction of groundwater level of a watershed," *J Computer Appl Res Dev,* vol. 1, pp. 35-42, 2011.
[169] D. C. Bisht and A. Jangid, "Discharge modelling using adaptive neuro-fuzzy inference system," *International Journal of Advanced Science and Technology,* vol. 31, pp. 99-114, 2011.
[170] Z. Alipour, A. M. A. Ali, F. Radmanesh, and M. Joorabyan, "Comparison of three methods of ANN, ANFIS and Time Series Models to predict ground water level:(Case study: North Mahyar plain)," 2014.
[171] S. A. Akrami, A. El-Shafie, and O. Jaafar, "Improving rainfall forecasting efficiency using modified adaptive Neuro-Fuzzy Inference System (MANFIS)," *Water resources management,* vol. 27, pp. 3507-3523, 2013.
[172] Y. Ytoui, "Rainfall-runoff modeling at monthly and daily scales using conceptual models and Neuro-fuzzy inference system," in *2nd International Conference - Water resources and wetlands*, Tulcea (Romania), 2014, pp. 263-270.
[173] E. Aldrian and Y. S. Djamil, "Application of multivariate ANFIS for daily rainfall prediction: influences of training data size," *Makara, Sains,* vol. 12, pp. 7-14, 2008.
[174] S. Mayya, "Runoff Modelling using different member ship functions in Adaptive Neuro fuzzy inference system," *International Journal of Advances in Engineering Sciences,* vol. 4, pp. 48-51, 2014.
[175] M. Vafakhah, F. Sedighi, and M. Javadi, "Modelling the Rainfall-Runoff data in snow affected watershed," *International Journal of Computer and Electrical engineering,* vol. 6, pp. 40-43, 2014.
[176] A. Solgi, V. Nourani, and A. Pourhaghi, "Forecasting Daily Precipitation Using Hybrid Model of Wavelet-Artificial Neural Network and Comparison with Adaptive Neurofuzzy Inference System (Case Study: Verayneh Station, Nahavand)," *Advances in Civil Engineering*, vol. 2014, 2014.
[177] M. T. Dastorani, "Application of ANN and ANFIS Models on Dryland Precipitation Prediction (Case study: Yazd in central Iran," *Journal of Applied Sciences,* vol. 10, 2010.
[178] N. Charaniya and S. Dudul, "Adaptive and regressive model for rainfall prediction," *International journal of advanced research in engineering and technology,* vol. 4, pp. 183-191, 2013.

[179] E. K. Lafdani, A. M. Nia, A. Pahlavanravi, and A. Ahmadi, "Rainfall-Runoff Simulation using MIKE11/NAM and ANFIS Models (A Case Study: Qleh Shahrokh Basin in Iran)," *World of Sciences Journal,* 2013.

[180] R. Panchal, T. Suryanarayana, and F. Parekh, "Adaptive Neuro-Fuzzy Inference System For Rainfall-Runoff Modeling."

[181] A. El-Shafie, O. Jaafer, and A. Seyed, "Adaptive neuro-fuzzy inference system based model for rainfall forecasting in Klang River, Malaysia," *Int J Phys Sci,* vol. 6, pp. 2875-2888, 2011.

[182] M. Vafakhah, S. Janizadeh, and S. Khosrobeigi Bozchaloei, "Application of Several Data-Driven Techniques for Rainfall-Runoff Modeling," *ECOPERSIA,* vol. 2, pp. 455-469, 2014.

[183] G. Fallah-Ghalhary, M. Mousavi-Baygi, and M. Nokhandan, "Soft computing models for annual rainfall prediction," in *7th ECAC Annual Meeting of the EMS*, 2008.

[184] E. Olyaie, H. Banejad, K.-W. Chau, and A. M. Melesse, "A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States," *Environmental monitoring and assessment,* vol. 187, pp. 1-22, 2015.

[185] B. Bekuretsion and T. Beshah, "Prediction of Rainfall for Ethiopian Upper Blue Nile Basin Using Soft Computing Models," *HiLCoE Journal of Computer Science and Technology,* p. 26, 2013.

[186] S. S. Sharifi, R. Delirhasannia, V. Nourani, A. A. Sadraddini, and A. Ghorbani, "Using Artificial Neural Networks (ANNs) and Adaptive Neuro-Fuzzy Inference System (ANFIS) for Modeling and Sensitivity Analysis of Effective Rainfall," *Recent Advances in Continuum Mechanics, Hydrology and Ecology, Mladenov V (eds),* pp. 133-139, 2013.

[187] A. Abraham, D. Steinberg, and N. S. Philip, "Rainfall forecasting using soft computing models and multivariate adaptive regression splines," *IEEE SMC Transactions, Special Issue on Fusion of Soft Computing and Hard Computing in Industrial Applications,* vol. 1, pp. 1-6, 2001.

[188] J. Patel and F. Parekh, "Forecasting Rainfall Using Adaptive Neuro-Fuzzy Inference System (ANFIS)."

[189] P. Niksaz and A. mohammad Latif, "Rainfall Events Evaluation Using Adaptive Neural-Fuzzy Inference System," *International Journal of Information Technology and Computer Science (IJITCS),* vol. 6, p. 46, 2014.

[190] R. Faulina and Suhartono, "Hybrid ARIMA-ANFIS for Rainfall Prediction in Indonesia," *nternational Journal of Science and Research,* vol. 2, pp. 159-162, 2013.

[191] B. Raup, A. Kääb, J. S. Kargel, M. P. Bishop, G. Hamilton, E. Lee, F. Paul, F. Rau, D. Soltesz, and S. J. S. Khalsa, "Remote sensing and GIS technology in the Global Land Ice Measurements from Space (GLIMS) project," *Computers & Geosciences,* vol. 33, pp. 104-125, 2007.

[192] M. P. Bishop, J. A. Olsenholler, J. F. Shroder, R. G. Barry, B. H. Raup, A. B. Bush, L. Copland, J. L. Dwyer, A. G. Fountain, and W. Haeberli, "Global Land Ice Measurements from Space (GLIMS): remote sensing and GIS investigations of the Earth's cryosphere," *Geocarto International,* vol. 19, pp. 57-84, 2004.

[193] J. Santillan, M. Makinano, and E. Paringit, "Integrating Remote sensing and GIS and Hydrologic Models for predicting land cover change impact on surface runoff and

sediment yield in a critical watershed in Mindano, Philippines," *Remote Sensing and Spatial Information Science,* vol. 38, pp. 436-441, 2010.
[194] M. Anderson and W. Kustas, "Thermal remote sensing of drought and evapotranspiration," *Eos,* vol. 89, pp. 233-234, 2008.
[195] J. Sun, G. D. Salvucci, and D. Entekhabi, "Estimates of evapotranspiration from MODIS and AMSR-E land surface temperature and moisture over the Southern Great Plains," *Remote Sensing of Environment,* vol. 127, pp. 44-59, 2012.
[196] A. V. Bradley, J. E. Thornes, L. Chapman, D. Unwin, and M. Roy, "Modelling spatial and temporal road thermal climatology in rural and urban areas using a GIS," *Climate Research,* vol. 22, pp. 41-55, 2002.
[197] R. Lal, "Soils and sustainable agriculture: A review," in *Sustainable agriculture*, ed: Springer, 2009, pp. 15-23.
[198] A. V. Ines, K. Honda, A. D. Gupta, P. Droogers, and R. S. Clemente, "Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture," *agricultural water management,* vol. 83, pp. 221-232, 2006.
[199] S. Frolking, J. Qiu, S. Boles, X. Xiao, J. Liu, Y. Zhuang, C. Li, and X. Qin, "Combining remote sensing and ground census data to develop new maps of the distribution of rice agriculture in China," *Global Biogeochemical Cycles,* vol. 16, pp. 38-1-38-10, 2002.
[200] L. Chapman and J. E. Thornes, "The use of geographical information systems in climatology and meteorology," *Progress in physical geography,* vol. 27, pp. 313-330, 2003.
[201] K. Solaimani, S. Modallaldoust, and S. Lotfi, "Investigation of land use changes on soil erosion process using geographical information system," *International Journal of Environmental Science & Technology,* vol. 6, pp. 415-424, 2009.
[202] M. Mohamed and B. Abdo, "Spatial variability mapping of some soil properties in El-Multagha agricultural project (Sudan) using geographic information systems (GIS) techniques," *Journal of Soil Science and Environmental Management,* vol. 2, pp. 58-65, 2011.
[203] M. Kavak and S. Karadogan, "The relationship between sea surface temperature and chlorophyll concentration of phytoplanktons in the Black Sea using remote sensing technique

" *Journal of Environmental Biology,* vol. 32, pp. 493-498, 211.
[204] E. Politi, M. E. Cutler, and J. S. Rowan, "Using the NOAA Advanced Very High Resolution Radiometer to characterise temporal and spatial trends in water temperature of large European lakes," *Remote Sensing of Environment,* vol. 126, pp. 1-11, 2012.
[205] S. B. Seng, A. K. Chong, and A. Moore, "Geostatistical modelling, analysis and mapping of epidemiology of dengue fever in Johor State, Malaysia," in *The 17th Annual Colloquium of the Spatial Information Research Centre, University of Otago, Dunedin, New Zealand*, 2005, pp. 24-25.
[206] B. C. B. Haarman, R. F. Riemersma-Van der Lek, W. A. Nolen, R. Mendes, H. A. Drexhage, and H. Burger, "Feature-expression heat maps–A new visual method to explore complex associations between two variable sets," *Journal of biomedical informatics,* vol. 53, pp. 156-161, 2015.

[207] D. Toddenroth, T. Ganslandt, I. Castellanos, H.-U. Prokosch, and T. Bürkle, "Employing heat maps to mine associations in structured routine care data," *Artificial intelligence in medicine*, vol. 60, pp. 79-88, 2014.
[208] O. A. Koita, L. Sangaré, H. A. Sango, S. Dao, N. Keita, M. Maiga, M. Mounkoro, Z. Fané, A. S. Maiga, and K. Traoré, "Effect of seasonality and ecological factors on the prevalence of the four malaria parasite species in northern mali," *Journal of tropical medicine*, vol. 2012, 2012.
[209] S. Sudhakar, T. Srinivas, A. Palit, S. Kar, and S. Battacharya, "Mapping of risk prone areas of kala-azar (Visceral leishmaniasis) in parts of Bihar state, India: an RS and GIS approach," *Journal of vector borne diseases*, vol. 43, p. 115, 2006.
[210] A. Kawasaki, R. Fujimaki, N. Kaneko, and S. SADOHARA, "Using GIS for assessing stream water chemistry in a forested watershed," *研究*, vol. 17, pp. 53-62, 2009.
[211] P. Pramojanee, C. Tanavud, C. Yongchalermchai, and C. Navanugraha, "An application of GIS for mapping of flood hazard and risk area in Nakorn Sri Thammarat Province, South of Thailand," in *Proceedings of International Conference on Geo-Information for Sustainable Management*, 1997, pp. 17-21.
[212] A. S. Dhakal, T. Amada, and M. Aniya, "Landslide hazard mapping and its evaluation using GIS: an investigation of sampling schemes for a grid-cell based quantitative method," *Photogrammetric Engineering and Remote Sensing*, vol. 66, pp. 981-989, 2000.
[213] C. Chien-Yuan, C. Tien-Chien, Y. Fan-Chieh, and L. Sheng-Chi, "Analysis of time-varying rainfall infiltration induced landslide," *Environmental Geology*, vol. 48, pp. 466-479, 2005.
[214] J. M. da Silva, C. V. Damásio, A. M. Sousa, L. Bugalho, L. Pessanha, and P. Quaresma, "Agriculture pest and disease risk maps considering MSG satellite data and land surface temperature," *International Journal of Applied Earth Observation and Geoinformation*, vol. 38, pp. 40-50, 2015.
[215] C. Daly, E. H. Helmer, and M. Quinones, "Mapping the climate of Puerto Rico, Vieques and Culebra," *International Journal of Climatology*, vol. 23, pp. 1359-1381, 2003.
[216] F. Lindberg, C. Grimmond, N. Yogeswaran, S. Kotthaus, and L. Allen, "Impact of city changes and weather on anthropogenic heat flux in Europe 1995–2015," *Urban Climate*, vol. 4, pp. 1-15, 2013.
[217] S. y. Obara, J. Morel, D. Konno, and Y. Utsugi, "Output prediction of large-scale photovoltaics by wind-condition analysis using 3D topographic maps," *Solar Energy*, vol. 105, pp. 157-169, 2014.
[218] J. Wilk and L. Andersson, "GIS-supported modelling of areal rainfall in a mountainous river basin with monsoon climate in southern India," *Hydrological sciences journal*, vol. 45, pp. 185-202, 2000.
[219] Q. Ye, S. Kang, F. Chen, and J. Wang, "Monitoring glacier variations on Geladandong mountain, central Tibetan Plateau, from 1969 to 2002 using remote-sensing and GIS technologies," *Journal of Glaciology*, vol. 52, pp. 537-545, 2006.
[220] C. Omuto, "A new approach for using time-series remote-sensing images to detect changes in vegetation cover and composition in drylands: A case study of eastern Kenya," *International journal of remote sensing*, vol. 32, pp. 6025-6045, 2011.
[221] J. D. Pleil, M. A. Stiegel, M. C. Madden, and J. R. Sobus, "Heat map visualization of complex environmental and biomarker measurements," *Chemosphere*, vol. 84, pp. 716-723, 2011.

[222] C. Santhi, R. Muttiah, J. Arnold, and R. Srinivasan, "A GIS-based regional planning tool for irrigation demand assessment and savings using SWAT," *Transactions of the ASAE*, vol. 48, pp. 137-147, 2005.
[223] S. Yelwa and N. Eniolorunda, "Simulating the Movement of Desertification in Sokoto and its Environs, Nigeria using 1km SPOT-NDVI Data," *Environmental Research Journal*, vol. 6, pp. 175-181, 2012.
[224] J. Choo, H. Lee, J. Kihm, and H. Park, "iVisClassifier: An interactive visual analytics system for classification based on supervised dimension reduction," in *Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on*, 2010, pp. 27-34.
[225] S. d. B. Ferraz, C. A. Vettorazzi, and D. M. Theobald, "Using indicators of deforestation and land-use dynamics to support conservation strategies: a case study of central Rondônia, Brazil," *Forest Ecology and Management*, vol. 257, pp. 1586-1595, 2009.
[226] W. L. Gorr, D. Nagin, and J. Szczypula, "Comparative study of artificial neural network and statistical models for predicting student grade point averages," *International Journal of Forecasting*, vol. 10, pp. 17-34, 1994.
[227] S. He and S. Liong, "Artificial Neural Network Technique for Raingauge Based Rainfall Nowcasing," in *Advances in Water Resources and Hydraulic Engineering*, ed: Springer, 2009, pp. 40-44.
[228] Y. Freund and R. E. Schapire, "A desicion-theoretic generalization of on-line learning and an application to boosting," in *Computational learning theory*, 1995, pp. 23-37.
[229] Y. Freund and R. E. Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting," *Journal of computer and system sciences*, vol. 55, pp. 119-139, 1997.
[230] A. Nachev and B. Stoyanov, "PRODUCT QUALITY ANALYSIS USING SUPPORT VECTOR MACHINES," *INFORMATION MODELS & ANALYSES*, p. 179, 2012.
[231] K. Kira and L. A. Rendell, "A practical approach to feature selection," in *Proceedings of the ninth international workshop on Machine learning*, 1992, pp. 249-256.
[232] I. Kononenko, "Estimating attributes: analysis and extensions of RELIEF," in *Machine Learning: ECML-94*, 1994, pp. 171-182.
[233] M. Robnik-Šikonja and I. Kononenko, "An adaptation of Relief for attribute estimation in regression," in *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, 1997, pp. 296-304.
[234] A. Shahi, R. Atan, and M. Sulaiman, "An effective fuzzy C-mean and type-2 fuzzy logic for weather forecasting," *Journal of Theoretical and Applied Information Technology*, vol. 5, pp. 556-567, 2009.
[235] C. Rodriguez Rivero, J. Pucheta, H. Patino, J. Baumgartner, S. Laboret, and V. Sauchelli, "Analysis of a Gaussian process and feed-forward neural networks based filter for forecasting short rainfall time series," in *Neural Networks (IJCNN), The 2013 International Joint Conference on*, 2013, pp. 1-6.
[236] M. M. Rahman, M. Rafiuddin, and M. M. Alam, "Seasonal forecasting of Bangladesh summer monsoon rainfall using simple multiple regression model," *Journal of Earth System Science*, vol. 122, pp. 551-558, 2013.
[237] C. Udomboso and G. Amahia, "Comparative analysis of rainfall prediction using statistical neural network and classical linear regression model," *Journal of Modern Mathematics and Statistics*, vol. 5, pp. 66-70, 2011.

[238] R. R. Deshpande, "On the Rainfall Time Series Prediction Using Multilayer Perceptron Artificial Neural Network," *International Journal of emerging technology and advanced engineering,* vol. 2, pp. 2250-2459, 2012.
[239] S. Vijayarani and M. Muthulakshmi, "Comparative Analysis of Bayes and Lazy Classification Algorithms," *International Journal of Advanced Research in Computer and Communication Engineering,* vol. 2, pp. 3118-3124, 2013.
[240] S. Kotsiantis, A. Kostoulas, S. Lykoudis, A. Argiriou, and K. Menagias, "Using data mining techniques for estimating minimum, maximum and average daily temperature values," *International Journal of Mathematical, Physical and Engineering Sciences,* vol. 1, pp. 16-20, 2008.
[241] B. Bhattacharya and D. P. Solomatine, "Neural networks and M5 model trees in modelling water level–discharge relationship," *Neurocomputing,* vol. 63, pp. 381-396, 2005.
[242] K. Wisaeng, "A Comparison of Decision Tree Algorithms For UCI Repository Classification."
[243] E. Alpaydin, "Introduction to machine learning (adaptive computation and machine learning series)," ed: The MIT Press Cambridge, 2004.
[244] J. R. Quinlan, "Simplifying decision trees," *International journal of man-machine studies,* vol. 27, pp. 221-234, 1987.
[245] M. Zontul, F. Aydın, G. Doğan, S. Şener, and O. Kaynar, "Wind Speed Forecasting Using Reptree And Bagging Methods In Kirklareli-Turkey," 2013.
[246] S. Peyvandi, H. M. Shirazi, and A. Faraahi, "Proposing a Classification Algorithm for User Identification According To User Web Log Analysis," *Australian Journal of Basic and Applied Sciences,* vol. 5, pp. 645-652, 2011.
[247] G. Wang, J. Hao, J. Ma, and H. Jiang, "A comparative assessment of ensemble learning for credit scoring," *Expert systems with applications,* vol. 38, pp. 223-230, 2011.
[248] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics & Data Analysis,* vol. 38, pp. 367-378, 2002.
[249] L. Breiman, "Bagging predictors," *Machine learning,* vol. 24, pp. 123-140, 1996.
[250] H. Lan, E. Frank, and M. Hall, "Data mining: Practical machine learning tools and techniques," ed: Morgan Kaufman, Boston, 2011.
[251] T. K. Ho, "The random subspace method for constructing decision forests," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 20, pp. 832-844, 1998.
[252] M. Skurichina and R. P. Duin, "Bagging, boosting and the random subspace method for linear classifiers," *Pattern Analysis & Applications,* vol. 5, pp. 121-135, 2002.
[253] C. Lai, M. J. Reinders, and L. Wessels, "Random subspace method for multivariate feature selection," *Pattern recognition letters,* vol. 27, pp. 1067-1076, 2006.
[254] S. M. Mwagha, M. Muthoni, and P. Ochieg, "Comparison of Nearest Neighbor (ibk), Regression by Discretization and Isotonic Regression Classification Algorithms for Precipitation Classes Prediction," *International Journal of Computer Applications,* 2013.
[255] D. H. Wolpert, "Stacked generalization," *Neural networks,* vol. 5, pp. 241-259, 1992.
[256] A. K. Seewald and J. Fürnkranz, "An evaluation of grading classifiers," in *Advances in Intelligent Data Analysis,* ed: Springer, 2001, pp. 115-124.
[257] L. Breiman, "Stacked regressions," *Machine learning,* vol. 24, pp. 49-64, 1996.

[258] M. Graczyk, T. Lasota, B. Trawiński, and K. Trawiński, "Comparison of bagging, boosting and stacking ensembles applied to real estate appraisal," in *Intelligent Information and Database Systems*, ed: Springer, 2010, pp. 340-350.
[259] L. Felföldi, "Classifier Combination Systems and their Application in Human Language Technology," *Research Group on Artificial Intelligence, Szeged, Hungary,* 2008.
[260] L. Kuncheva, "Fuzzy vs Non-fuzzy in combining classifiers designed by boosting," in *IEEE Transaction on Fuzzy Systems*, 2003, pp. 729-741.
[261] J. Kittler, M. Hatef, R. P. Duin, and J. Matas, "On combining classifiers," *Pattern Analysis and Machine Intelligence, IEEE Transactions on,* vol. 20, pp. 226-239, 1998.
[262] N. Ullah and P. Choudhury, "Flood Flow Modeling in a River System Using Adaptive Neuro-Fuzzy Inference System," *Environmental Management and Sustainable Development,* vol. 2, pp. 54-68, 2013.
[263] R. Rojas, "Neural Networks-A Systematic Introduction Springer-Verlag," *New York,* 1996.
[264] R. Kirkby, E. Frank, and P. Reutemann, " WEKA Explorer User Guide for Version 3-5-8," University of Waikato2008.
[265] O. Folorunsho, "Comparative study of different data mining techniques performance in knowledge discovery from medical database," *International Journal of Advanced Research in Computer Science and Software Engineering,* vol. 3, pp. 11-15, 2013.
[266] E. L. Lehmann and G. Casella, *Theory of point estimation* vol. 31: Springer Science & Business Media, 1998.
[267] D. J. Cowen, "GIS versus CAD versus DBMS: What are the differences?," *Photogramm. Eng. Remote Sens.,* vol. 54, pp. 1551-1555, 1988.
[268] R. J. H. Robles, "Development of Univ. of San Agustin Geographic Information System (USAGIS)," 2014.
[269] W. C. Hession and V. Shanholtz, "A geographic information system for targeting nonpoint-source agricultural pollution," *Journal of Soil and Water conservation,* vol. 43, pp. 264-266, 1988.
[270] M. Batty, "Urban modeling in computer-graphic and geographic information system environments," *Environment and Planning B,* vol. 19, pp. 663-663, 1993.
[271] D. J. Peuquet and D. F. Marble, *Introductory readings in geographic information systems*: CRC Press, 2003.
[272] H. J. Miller and E. A. Wentz, "Representation and spatial analysis in geographic information systems," *Annals of the Association of American Geographers,* vol. 93, pp. 574-594, 2003.
[273] T. C. Bailey and A. C. Gatrell, *Interactive spatial data analysis* vol. 413: Longman Scientific & Technical Essex, 1995.
[274] G. Upton and B. Fingleton, *Spatial data analysis by example. Volume 1: Point pattern and quantitative data*: John Wiley & Sons Ltd., 1985.
[275] H. Samet, "Applications of spatial data structures," 1990.
[276] A. Hackeloeer, K. Klasing, J. M. Krisp, and L. Meng, "Georeferencing: a review of methods and applications," *Annals of GIS,* vol. 20, pp. 61-69, 2014.
[277] A. T. Murray, I. McGuffog, J. S. Western, and P. Mullins, "Exploratory spatial data analysis techniques for examining urban crime implications for evaluating treatment," *British Journal of criminology,* vol. 41, pp. 309-329, 2001.

[278] T. Erden and H. Karaman, "Analysis of earthquake parameters to generate hazard maps by integrating AHP and GIS for Küçükçekmece region," *Natural Hazards and Earth System Science*, vol. 12, pp. 475-483, 2012.
[279] M. Negnevitsky, C. Potter, and M. Ringrose, "Short Term Wind Forecasting Techniques for Power Generation," in *Australasian Universities Power Engineering Conference*, 2004.
[280] J.-S. R. Jang, "Input selection for ANFIS learning," in *Proceedings of the fifth IEEE international conference on fuzzy systems*, 1996, pp. 1493-1499.
[281] C. M. Bishop, *Neural networks for pattern recognition*: Oxford university press, 1995.

List of Author’s Publications

1. Computational Intelligence in Weather Forecasting: A Review, Journal of Network and Innovative Computing (JNIC), Volume 1, pp. 320-331, 2013.
2. Weather forecasting in Sudan using Machine Learning Schemes, Journal of Network and Innovative Computing (JNIC), Volume 2, pp. 309-317, 2014.
3. Novel Ensemble Method for Long Term Rainfall Prediction, International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Volume 7, pp. 116-130, 2015.
4. Using Adaptive Neuro-Fuzzy Inference System (ANFIS) to Improve the Long-term Rainfall Forecasting, Journal of Network and Innovative Computing (JNIC), Volume 3, pp. 146-158, 2015.
5. Spatial Diversity of Monthly and Annual Rainfall in Sudan: Using Rainfall Predictions and Geographical Information System (GIS) to Produce Rainfall Maps, International Journal of Computer Information Systems and Industrial Management Applications (IJCISIM), Volume 7, pp. 151-172, 2015.