Decision Support System for On-Farm Crop Water Irrigation Scheduling using Machine Learning approaches

A

Thesis submitted

for the award of the degree of

DOCTOR OF PHILOSOPHY

By

Mandeep Kaur Saggi
(901503038)

Under the guidance of

Dr. Sushma Jain
(Associate Professor, CSED)

Computer Science and Engineering Department
Thapar Institute of Engineering and Technology
Patiala - 147004, India

August 2021

CERTIFICATE

This is to certify that the Thesis entitled "**Decision Support System for On-Farm Crop Water Irrigation Scheduling using Machine Learning approaches**", submitted by **Mandeep Kaur Saggi** (901503038), a research scholar in the *Computer Science and Engineering Department, Thapar Institute of Engineering & Technology, Patiala*, for the award of the degree of **Doctor of Philosophy**, is a record of an original research work carried out under the supervision of Dr. Sushma Jain and refers work of other researchers which are duly listed in reference section. The Thesis has fulfilled all requirements as per the regulations of the Institute and in our opinion has reached the standard needed for submission.

The results embodied in this Thesis have not been submitted to any other University or Institute for the award of any degree.

Mandeep Kaur Saggi
Registration No. 901503038

This is to certify that the above statement made by the candidate is correct and true to the best of my knowledge and belief.

Date: 30-12-2021
Place: Patiala

Dr. Sushma Jain
Associate Professor,
Computer Science and Engineering Department,
Thapar Institute of Engineering and Technology, Patiala,
India - 147004

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to all those people who made this dissertation possible. At first and foremost, I would like to express my profound respect and gratitude to my supervisor, Dr. Sushma Jain, who has been my guiding force behind this work. I am greatly bounded for her constant encouragement, invaluable guidance, and her valuable comments on my work. More importantly, I would like to thank for the patience she has shown in carefully reading and commenting on the manuscripts, and countless revisions of this dissertation. Her commitments and dedication to research have been and will continue to be a constant source of inspiration for me. I am fortunate enough to have such an advisor who gave me the freedom to think independently and explore new ideas. I have no doubts that finishing my degree in proper and timely manner was impossible without her help. A special appreciation goes to Dr. Sushma Jain for giving critical and valuable suggestions. I am highly privileged to have got an opportunity to work with such a wonderful personality.

I would also like to extend my sincere thanks to the doctoral committee members Dr. Maninder Singh, Dr. Rinkle Rani, Dr. Sanmeet Bhatia and Dr. Alpana Agarwal for their invaluable suggestions, encouragements, and moral support that helped me to improve my research work. I am also thankful to the Head of the Department, and Dr. Rafat Siddique, Dean of Research and Sponsored Projects for their kind help carried out during my academic studies and entire research tenure.

My special thanks to my moral adviser Dr. Rakesh Sharda, Principal Scientist (Extension), Punjab Agricultural University (PAU), for his guidance, constant support and insightful comments during the entire journey of my Ph.D. Tenure. I wish to express my deep sense of gratitude to S. Mahinder Singh Dosanjh, Advisor to PAU, and Dr. Karanjeet Singh Thind, Additional Director of Research (Crop Improvement) PAU, for much needed support throughout the work.

It is with immense gratitude that I acknowledge Dr. Sanjay Jain, Professor, TIET, Patiala for his enriching suggestions, manuscript reading and editing whenever required.

On a personal note, I would like to thank my husband Amandeep Singh Bhatia for his constant support and being with me in every aspect of my life. I had a great time with my many friends at Thapar Institute of Engineering & Technology specially Jagroop Kaur, Prashu Bansal, Deepika Sharma, Natasha, and Divya. I would like to thank them for their support and encouragement. I would like to

thank and appreciate my Ph.D. fellows Gitika Sharma and Suman Saurabh Sarkar for correcting the Thesis.

I am grateful to my parents Father- *Harjit Singh Saggi*, Mother- *Parminder Kaur Saggi*, Brother-*Harpreet Singh Saggi*, Sister-in-Law- *Ramandeep Kaur* and my most heartfelt love to their child: *Karamveer Singh Saggi* whose love encouragement, and support made this research work possible. I would especially thank my Father in-law- *Dalbir Singh Bhatia* and Mother in-law- *Satinder Kaur*, son of brother in-law *Sehajveer Singh Bhatia*, and my dear cousins Ginni, Goldy, Bablu, Rinku, and Dolly, who are the joy of my life. I would also like to pay my sincere regards to my Sister-*Sandeep Kaur Saggi*, Brother-in-Law- *Simranjeet Singh Saini* and all my relatives for their constant motivation and support. They made this journey easier with words of encouragement which helped me in finishing my work.

I would like to express my sincere thanks to the Council of Scientific and Industrial Research, New Delhi under grant no 09/677(0040)/2019-EMR-I for awarding me as Senior Research Fellow to carry out my research work on project entitled "Decision Support System for On-Farm Crop Water Irrigation Scheduling using Machine Learning approaches".

I am thankful to Thapar Institute of Engineering & Technology for providing the research scholarship to undertake my Ph.D. research. Finally, I would like to thank the Almighty God for bestowing upon me this opportunity and showering his blessings on me to come out successful against all odds.

Mandeep Kaur Saggi

ABSTRACT

In India, irrigation is the largest consumer of fresh water and is drawing about 90% of groundwater. The requirement of irrigation system is much needed for the region such as Central Punjab, which occupies nearly 97.95% of the gross irrigated area for agricultural production. The water consumption is very high in commonly cultivated crops in Punjab that are Wheat, Maize, and Rice. This requires modern technologies in water management to meet the agricultural challenges. Hence, this system referred as Irrigation Water Management (IWM), is poised to be a key driver of smart farming to meet crop water requirement with a sufficient economic return without any damage to land and soil.

The major challenge in agriculture sustainability and dawdling is to utilize every drop of fresh water effectively and efficiently. The studies on water shortage suggest the development of innovative irrigation methods such as controlled deficit irrigation, partial root drying, and continuous deficit irrigation. In this context, the controlled irrigation, climate, soil fertility, crop quality, and time management are essential to the Decision Support System (DSS) to maximize the crop yield with minimum consumption of water.

Advanced Analytics and DSS can help farm managers in taking decision to solve complex irrigation problem. The Reference Evapotranspiration (ET_o) is one of the most valuable parameters for hydrological, climatologist investigation, and water resources management. An exact estimate of ET_o is necessary to analyze the water demand of irrigated agriculture, crop-water balance and improve the water quality. However, ET_o estimation is very difficult to achieve due to its dependency on many input parameters. Therefore, the primary objective of the research is to establish regression models for the estimation of ET_o with limited climate parameters. The estimation of daily ET_o can help in real-time prediction of crop evapotranspiration and crop irrigation demand.

The framework of ensemble-based modeling has been designed in this work using Machine Learning (ML) and Deep Learning (DL) models for estimating the Reference Evapotranspiration ET_o, Crop Evapotranspiration ET_c, and thus meet crop water requirement in Irrigation Scheduling (IS). The work is carried out with the following objectives:

- Estimation of Reference Evapotranspiration (ET_o) using H2O framework based on Deep Learning-Multilayer Perceptron's (DL), Generalized Linear Model

(GLM), Random Forest(RF), and Gradient-Boosting Machine (GBM) for classification and regression purposes.

- Estimation of single Crop Coefficient (K_c) and Crop Evapotranspiration (ET_c) using a novel multilevel ensemble model based on Fuzzy-Genetic (FG) and Regularization Random Forest (RRF) models for Wheat and Maize crops.

- Investigations on DSS based crop water requirement, Net Irrigation, Gross Irrigation, and Pumping Time in Irrigation Scheduling using Particle-Swarm optimization with Deep Neural Network (PSO-DNN), and Deep Learning models.

The H2O model framework is investigated to determine the daily Reference Evapotranspiration (ET_o) for the Hoshiarpur and Patiala districts of Punjab. Daily Meteorological dataset is collected from Indian Meteorological Department (IMD), Pune having combination of six input variables i.e. T_{min}, T_{max}, R_H, u_2, I_s and R_s. The appropriate missing values have been filled using MissForest algorithms. Four supervised learning algorithms have been investigated to forecast ET_o. These are Deep Learning-Multi-Layer Perceptron (DL), Generalized Linear Model (GLM), Random Forest (RF), and Gradient-Boosting Machine (GBM). The FAO-56 Penman-Monteith method is used to calculate the Rs and ET_o. A three-layer Deep Learning model with Rectified Linear Unit (ReLU) function and Stochastic Gradient Descent (SGD) via Backpropagation have been developed to obtain the minimum prediction error. The multinomial classification as cross-entropy function has been applied for calculating the loss function. The effectiveness of the developed model is tested using performance metrics to obtain the accurate estimation of ET_o using classification and regression purposes. A three-layer multi- model ensemble machine learning approach has been investigated to predict ET_o for regression. The three supervised machine learning models: Extreme Machine Learning (ELM_1, ELM_2, ELM_3, ELM_4), Multi-Layer Perceptron's Neural Network (MLP_1, MLP_2, MLP_3, MLP_4), Support Vector Machine (SVM_1, SVM_2, SVM_3, and, SVM_4) models have been applied to investigate the abilities and applicability of the ensemble-based model.

The investigations on estimation of Crop Coefficient (K_c), and Crop Evapotranspiration (ET_c) are carried out using ensemble-based modeling. Two algorithms namely Fuzzy-Genetic (FG), and Regularized Random Forest (RRF) are applied to develop the FG-RRF based (ET_c) framework for three crops, namely (Maize, Wheat1, and Wheat2). Fuzzy-Genetic model is applied to simulate the K_c and ET_c values using a training dataset. The K_c and ET_c prediction probabilities are combined in dataset. Then, the ensembling dataset is used to train the RRF model for predicting the ET_c values of each sample of the crop. After getting the best accuracy

from a training model, the testing dataset is applied to validate the accuracy of the model. The proposed model is evaluated based on performance metrics to check the accuracy of results. The effectiveness of the developed model (FG-RRF based ET_c) is tested and compared with the SVM model including (Sigma and C) parameters.

The ensemble-based model is proposed to estimate the DSS-IS using Particle Swarm Optimization with Deep Neural Neural (PSO-DNN), and Deep Learning (DL). The fuzzy forest algorithm is applied for the selection of the best ten input features. The algorithm is based on random forest and designed to reduce and rank the important number of features in regression. These features are chosen based on a feature recursive exclusion function. Using these models, a DSS is developed with improved accuracy by reducing the number of features. The effectiveness of the developed model DSS-IS using PSO-DNN and DL is tested and compared with three samples of irrigation parameters. A decision support system can help the user not only in estimating the ET_o but also in selecting the best (ET_o) and (ET_c) (and intermediate parameters). The enhancement of water efficiency requires controlling the high demand for irrigated agriculture, which improves the capabilities to simulate the water cycle and its components accurately. The ensemble learning can enhance the effectiveness of classifiers by blending their decisions individually. A combination of Daily average temperature and solar radiation is the optimal combination for the ET_o and ET_c estimation. Ensemble models showed great applicability in modeling ET_o, and can be highly recommended for estimating (ET_c) in Punjab as well as other stations.

Contents

List of Figures

List of Tables

List of Acronyms

DSS	Decision Support System
ICTs	Information and Communication Technologies
ET	Evapotranspiration
ET_o	Reference Evapotranspiration
ET_c	Crop Evapotranspiration
K_c	Crop Coefficient
IS	Irrigation Scheduling
ICTs	Information and Communication Technologies
FAO	Food and Agriculture Organization
CWR	Crop Water Requirement
IWR	Irrigation Water Requirement
PM	Penman-Monteith
HS	Hargreaves-Samani
BC	Blaney-Criddle
GUI	Graphical User Interface
DL	Deep Learning
GBM	Gradient Boosting Machine
GLM	Generalized Linear Model
RF	Random Forest
FG	Fuzzy-Genetic
RRF	Regularization Random Forest
IRLSM	Iterative Re-weighted Least Squares
MPS	Matrix Product State
ELM	Extreme Machine Learning
ANN	Artificial Neural Network
WNN	Wavelet Neural Network
MLP	Multi-layer Perceptrons-Neural Network
SVM	Support Vector Machine
SLFNs	Single-hidden Layer Feed-forward Neural Networks
ANFIS	Adaptive Neuro-Fuzzy Inference System
BMAM	Bayesian Model Averaging Model

LS-SVM	Least Square Support Vector Machine
LS-SVR	Least Squares Support Vector Regression
MARS	Multivariate Adaptive Regression Splines
GEP	Gene Expression Programming
FG-RRF	Fuzzy-Genetic and Regularized Random Forest
FF	Fuzzy Forests
FA	Firefly Algorithm
RL	Relief
PCA	Principal Component Analysis
SVR	Support Vector Regression
WOA	Whale Optimization Algorithm
DWT	Discrete Wavelet Transform
WGCNA	Weighted Gene Co-expression Network Analysis
RFE-RFs	Recursive Feature Elimination Random Forests
VIM	Variable Importance Measure
DSS-IS	Decision Support System for Irrigation Scheduling
PSO	Particle Swarm Optimization
DNN	Deep Neural Networks
IP_t	Irrigation Period
IR_{gv}	Gross Irrigation Requirement
Net_{IR}	Net Irrigation Requirement
SAR	Synthetic Aperture Radar
MSE	Mean Square Error
RMSE	Root Mean Square Error
NRMSE	Normalized Root Mean Square Error
RMSLE	Root Mean Square Logarithmic Error
NSE	Nash–Sutcliffe Efficiency
ACC	Accuracy
MCE	Mean Per-Class Error
LL	Logloss

List of Symbols

Symbol	Description	Unit
T_{max}	air temperature maximum	(°C)
T_{min}	air temperature minimum	(°C)
T_{mean}	air temperature mean	(°C)
R_s	solar radiation	(MJ m^{-2} day^{-2})
I_s	sunshine hours	(h)
R_n	net radiation	(MJ $m^{-2}day^{-1}$)
P_e	effective rainfall	(mm)
P	rainfall	(mm)
R_H	relative humidity	(%)
RH_{max}	relative humidity maximum	(%)
RH_{min}	relative humidity minimum	(%)
RH_{mean}	relative humidity mean	(%)
u_2	wind speed	2 m height (ms^{-1})
E_f	field irrigation system efficiency	(%)
R_g/R_a	global radiation and solar radiation	(MJ $m^{-2}day^{-1}$)
Δ	slope of saturation vapor pressure function	(kPa $°C^{-1}$)
γ	psychometric constant	(kPa $°C^{-1}$)
G	soil heat flux density	(MJ $m^{-2}day^{-1}$)
e_a	actual vapour pressure	(kPa)
e_s	saturation vapour pressure	(kPa)
$(e_s - e_a)$	vapour pressure deficit	(kPa)
ρ	mean daily percentage of annual daytime hours	hours
$\bar{v}$	is the mean velocity of wind at some height z	
v	is the measured velocity of wind at some standard height z_1	
z_1	is the measure of surface roughness	
$M_i^{\sigma_i}$	complex square matrices	
G_e	groundwater contribution	

X_{true}	total data matrix
X_{imp}	imputed data matrix
N	total number of sample
R	pearson correlation
R_2	coefficient of determination
x_i	inputs
w_i	weights
b_i	bias
β	is determined by maximum likelihood
β_k	refers to the kth column (outcome)
β_j	the j^{th} row (vector of K coefficients for variable j)
β_c	is a vector of coefficients for class c
j	denotes the penalty factor for the $jthj^{th}$ variable
η	denotes the training output matrix
$Z^{\dagger}$	is the moore–penrose generalized inverse of the matrix Z

List of Publications

Journal Publications

1. **Mandeep Kaur Saggi** and Sushma Jain, Application of fuzzy-genetic and regularization random forest (FG-RRF): Estimation of crop evapotranspiration (ETc) for maize and wheat crops, **Agricultural Water Management**, https://doi.org/10.1016/j.agwat.2019.105907, 229, 105907, 2020. [SCIE Indexed, Impact Factor 4.51].

2. **Mandeep Kaur Saggi** and Sushma Jain, Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning, **Computers and electronics in agriculture**, https://doi.org/10.1016/j.compag.2018.11.031, 156, 387-398, 2019 [SCIE Indexed, Scopus, Impact Factor 5.56].

3. **Mandeep Kaur Saggi**, Amandeep Singh Bhatia, Sushma Jain, and Ajay Kumar, Matrix Product State–Based Quantum Classifier, **Neural computation**, 31(7), https://doi.org/10.1162/necoa01202, 1499-1517, 2019 [SCIE Indexed, Scopus, Impact Factor 2.505].

4. **Mandeep Kaur Saggi**, Sushma Jain, A survey towards an integration of big data analytics to big insights for value-creation. Information Processing & Management, 54(5), https://doi.org/10.1016/j.ipm.2018.01.010, 758-790, 2018 [SCIE Indexed, Scopus, Impact Factor 6.22].

Under Communicated

5. **Mandeep Kaur Saggi**, Sushma Jain, Amandeep Singh Bhatia, Rakesh Sharda. Proposition of New Ensemble Data-Intelligence Model for Evapotranspiration Process Simulation. [Communicated in Journal of Ambient Intelligence & Humanized Computing].

6. **Mandeep Kaur Saggi**, Sushma Jain, A Survey Towards Decision Support System on Smart Irrigation Scheduling using Machine Learning. [Under review in Archives of Computational Methods in Engineering].

7. **Mandeep Kaur Saggi**, Sushma Jain, Decision support system for smart irrigation system using deep learning. [Submitted in Computers and electronics in agriculture]

Chapter 1

Introduction

"*WATER- Every drop is precious, save it*". Water is the main limiting factor of agricultural development in semi-arid and arid climates. "Arthur Keith said that the discovery of agriculture is the first major step for civilized life". For the last seven decades, Indian agriculture has played a major part in the country's economic development.

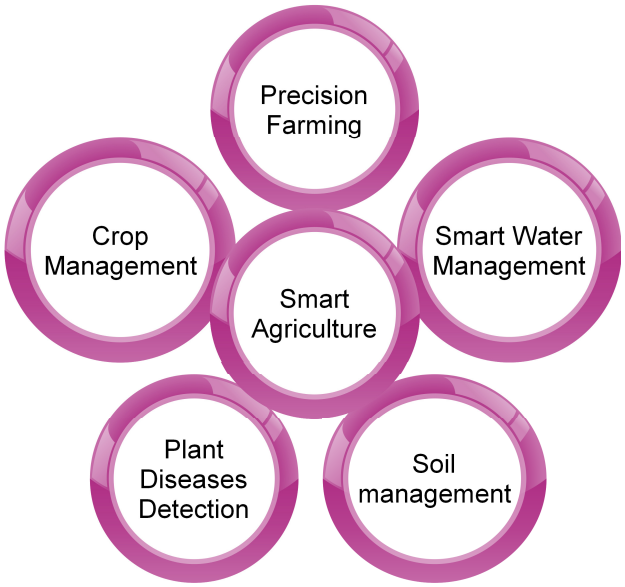

Figure 1.1: Smart agriculture management

However, the agriculture sector of India has been transformed from traditional to modern practices by the deployment of Information and Communication Technologies (ICTs) which provide various services (such as- Smart water management, Soil management, Plant diseases, Precision Farming, and Crop management as depicted in Fig. 1.1). The demand of water for the India's agriculture and industry sectors is constantly increasing to fulfill the requirements of 1.366 billion people. Central Punjab has occupied a high percentage of the area for agricultural production which requires proper irrigation system.

Punjab has 97.95% of irrigation area [1]. Therefore, achievement of the Green Revolution is endangered by a significant decline in water resources. Consequently, water conservation and precision agriculture are becoming vital issues in tropical climate areas. Wheat, maize and rice are the most commonly cultivated crops that have high water consumption in Punjab. The major challenge in agriculture sustainability and dawdling is due to climate change; therefore, every drop of freshwater needs to be utilized effectively and efficiently. Kumar et al. [2] investigated the changes in the frequency, area, and population exposure affected by both dry and wet extremes in India under the warming climate.

To overcome these challenges, the multivariate, complex, and unpredictable agricultural ecosystems must be well understood by continuously analyzing, measuring, and monitoring several physical aspects and phenomena [3]. Advanced Analytic and DSS can help to solve the complex problems of farm managers to take decision regarding irrigation.

1.1 Overview

"*Irrigation is defined as the process of artificially supplying water to the soil for the growth of crop*" [4]. Therefore, it is becoming a very important managerial activity to achieve effective and efficient utilization of water. The primary objective of good irrigation scheduling is to apply the right amount of water at right time. There is need to predict the water requirement for the crops development.

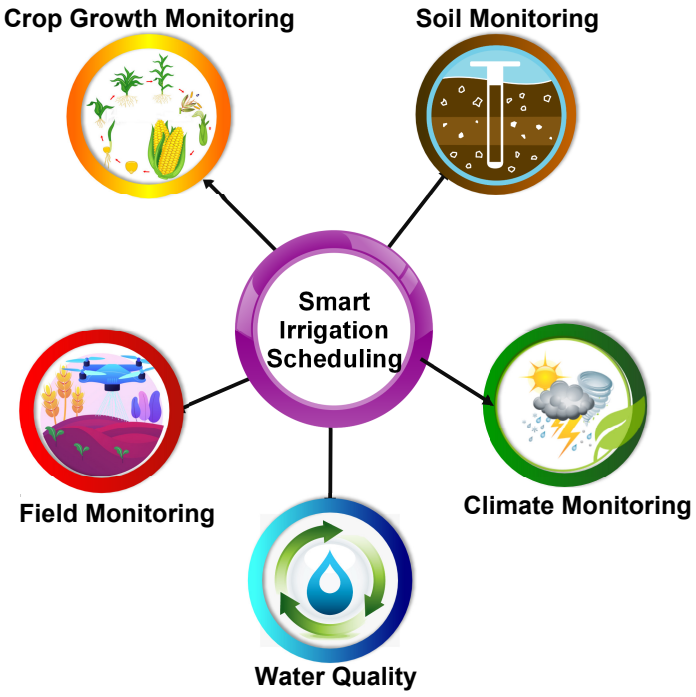

Figure 1.2: Smart irrigation scheduling

Irrigation scheduling identifies the crop-water balance, control sensor technologies, and improve the water quality by estimation of evapotranspiration (ET_o) [5]. Fig. 1.2, presents the approaches for Smart Irrigation Scheduling such as Crop growth monitoring, Soil monitoring, Climate monitoring, water quality, and Field monitoring using drones.

The ultimate sustainable irrigation potential of India has been estimated in 1991 United Nations' Food and Agriculture Organization (FAO) report to be 139.5 million hectares, comprising 58.5 mha from major and medium river-fed irrigation canal schemes, 15 mha from minor irrigation canal schemes, and 66 mha from groundwater well fed irrigation [6]. It is estimated that even after achieving the full irrigation potential, nearly 50 percent of the total cultivated area will remain rain-fed.

The expected problem in irrigation scheduling is "when and how much to irrigate?". Currently, irrigation decision-making systems applied to the field of agriculture mostly aim at a given area and for specific crops [7] But it is difficult to be applied to different areas and crops. Under the growing environment, the amount of irrigation is defined as the quantity of water required to meet the crop water loss due to evapotranspiration.

It can be obtained via prediction using indirect channels or field measurement techniques. However, the amount of water and timing of irrigation have major impact on quality of crop and its yield. Several methods are applied for the irrigation scheduling such as pan evaporation, soil moisture, leaf water potential, and crop growth-stages. The demand of water can be fulfilled by complete or partial irrigation in all methods.

1.2 Reference Evapotranspiration (ET_o)

"*Evapotranspiration is the sum of evaporation from the soil surface, and transpiration from plant surfaces to the atmosphere* " [8].

Irrigation scheduling is applied to meet the needs of daily crop evapotranspiration and its accurate estimation. Evapotranspiration of crops differs significantly over the growing season mainly due to alterations in climatic conditions and crop cover. It also varies among the crops. Therefore, estimation of water required for a crop is a crucial parameter in the planning, and development of irrigation management systems. Fig. 1.3 presented the concept of ET_o.

The reference evapotranspiration is approximated from meteorological data with the following parameters (relative humidity, air temperature, wind duration/wind speed, sunshine hours and solar radiation) by using the Penman-Monteith equation [8].

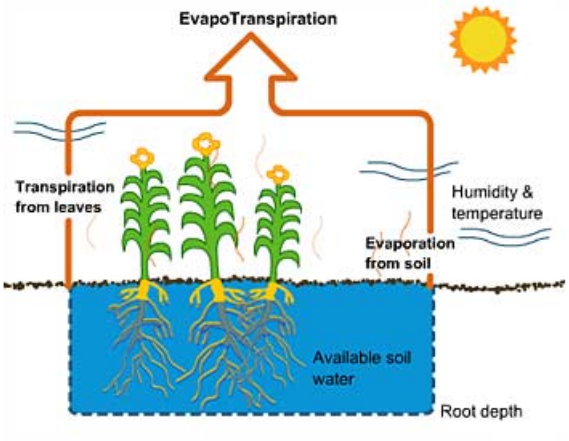

Figure 1.3: Evapotranspiration ET_o process

Factors Affecting ET_o

Climate and weather play significant role in determining long-term and day-to-day activities in the agriculture. The demand of crop-water is determined largely by weather variables. Rainfall is the foremost weather variable that affects the water resources for irrigation planning. The main climatic factors for agriculture are rainfall, maximum/minimum temperature, solar radiation, humidity, photo-period or sunshine hour, wind speed, and night temperature [9] as depicted in Fig. 1.4.

Weather elements control the crop water demand and crop evapotranspiration which depends upon the different weather elements such as humidity, temperature, sunshine hours, solar radiation, wind speed, etc. It is also affected by rainfall. The weather elements affecting ET_o is presented by Table 1.1:

FAO ET_o Estimation Models

There are various mathematical models used to calculate the reference evapotranspiration (ET_o). The procedures for calculating the ET_o is introduced by the Food and Agriculture Organization of the United Nations (FAO) and it is presented in Table 1.2.

So, the FAO-56 Penman Monteith method is the most dominant as compared to others methods [8] [10] [11]. Moreover, it requires several climatic data factors to estimate ET_o Maroufpoor et al. [12].

The FAO-56 Penman Monteith method [8] has been broadly used to analyze ET_o from meteorological factors and it is suggested as the standard technique by the FAO Ladlani et al. [13] and calculated by Eq. (1.1).

$$ET_0 = \frac{0.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot \frac{900}{T_{mean}+273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{1.1}$$

where, ET_o is the reference evapotranspiration (mm/day^{-1}); Δ= slope of saturation vapor pressure function (kPa $°C^{-1}$); R_n= net radiation (MJ $m^{-2}day^{-1}$); γ = psychometric constant (kPa $°C^{-1}$); G = soil heat flux density (MJ $m^{-2}day^{-1}$); u_2 = average 24-h wind speed at 2 m height (ms^{-1}); T_{mean} = mean air temperature (°C); e_a = actual vapour pressure (kPa); e_s = saturation vapour pressure (kPa); and $(e_s - e_a)$ = vapour pressure deficit (kPa).

Table 1.1: Weather affecting parameters

Parameters	Models	References
Solar Radiation	Richardson Model $R_g/R_a = a(T_{max} - T_{min})^b$	Richardson [14]
	Angstrom Model $R_g = R_a(a+b)(n/N)$	Angstrom [15]
	Regression Model $R_g = a_0 + a_1x_1 + a_2x_2 + ... + a_nx_n$	Ali et al. [16]
Air Temperature	$T_{mean} = \frac{T_{max} + T_{min}}{2}$	Allen et al. [17]
Air Humidity	Relative Humidity Instruments for measurements Psychrometer, Hair hygrometer Dew-point hygrometer	ALi [4]
Wind	$\frac{\bar{v}}{v} = \frac{z}{z_1}^k$, $\frac{\bar{v}}{v_1} = \frac{ln(z/z_0+1)}{ln(z1/z_0+1)}$	Linsley [18] Sutton [19]
Sunshine hour	$N = (2/15)cos^{-1}(-tan\delta tan\phi)$	Duffie and Beckman [20]
Rainfall	Arithmetic average, Theissen weight, Isohyetal weight $P = (P_1 + P_2 + ... + P_n)/n$ $P = (P_A.W_A) + (P_B.W_B) + (P_C.W_C) + ..$ $P = \frac{\sum P_iA_i}{\sum A_i}$ Effective rainfall=Total rainfall-Surface runoff	Ali [4]

To conclude, FAO-56 Penman Monteith method can estimate ET_o most accurately, if attributes (vapor pressure deficiency, wind speed, minimum and maximum air temperatures, and solar radiation) are being made available from meteorological weather stations [21].

Thus, in case of insufficient data or nonavailability of data other models could be required for estimating ET_o. So, this creates the requirement of exploring more flexible models or developing new models to calculate ET_o using fewer weather properties and a reasonable precision.

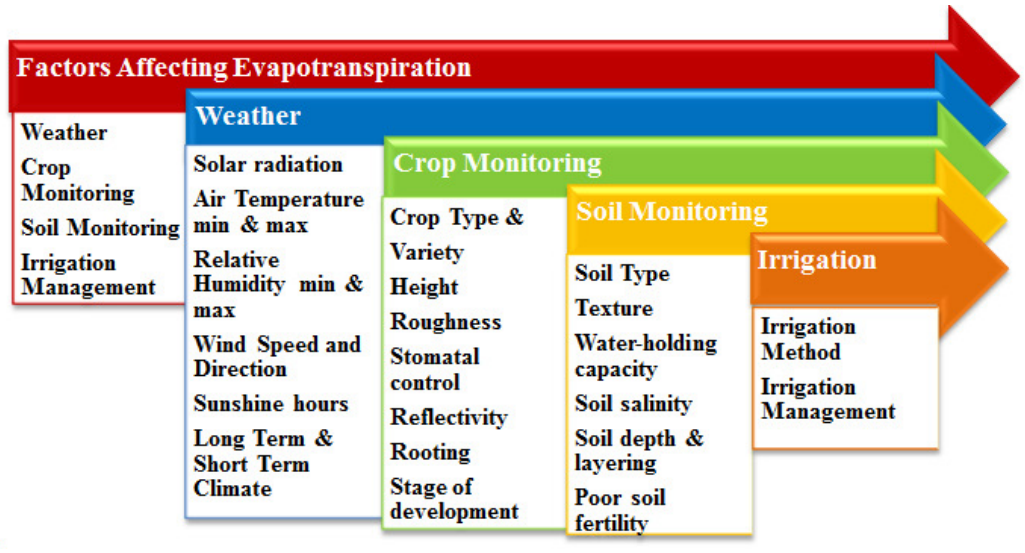

Figure 1.4: Factors affecting evapotranspiration

Table 1.2: ET_o estimation empirical methods

Method	Formula	Ref.
A standard scientific empirical model		
FAO Penman Monteith	$ET_o = \frac{0.0864}{\lambda} . \frac{\triangle(R_N - G) + c_p \rho_a DPV / r_a}{\triangle + \gamma(1 + r_c / r_a)}$	[17]
Temperature based estimation models		
FAO Blaney-Criddle method	$ET_o = [\rho(0.46T + 8)]$	[22]
Rational use of the FAO Blaney-Criddle method	$ET_o = [b\rho(0.46T + 8.13)(1 + 0.0001E)]$	[23]
Hargreaves and Samani method	$ET_o = (0.00023R_a)(T_{mean} + 17.8)TD^{0.5}$	[24]
Hargreaves and Samani method 1	$ET_o = (0.0030 * 0.408R_a)(T_{mean} + 20)TD^{0.4}$	[25]
Thornthwaite	$ET_(o) = 16\frac{T_{mean}}{i}^i$	[26]
Pan evaporation method	$ET_o = K_p E_{pan}$	[27]
Radiation based estimation models		
FAO radiation method	$ET_o = a + b[\frac{\triangle}{\triangle + gamma} * R_s]$	[28]
Priestley-Taylor	$ET_o = 1.26\frac{\triangle}{\triangle + \gamma} * \frac{R_n - G}{\lambda}]$	[29]
Jesen-Haise	$ET_o = 0.408 * C_{T_{mean}} * (T - T_k) * K_{T_{mean}} * R_a * TD^{0.5}$	[30]

1.3 Crop Evapotranspiration (ET_c)

"*The crop evapotranspiration denoted as ET_c, is the water requirement of crop.*" Crop Evapotranspiration is the most significant factor to understand the requirement of crop water, the utilization of water resources for irrigation scheduling, and the monitoring of crop yield production. The crop evapotranspiration can be computed in two steps, where in first step a reference evapotranspiration ET_o is determined and then, in second step, the crop evapotranspiration ET_c is estimated. Various climate conditions and crop-growth models are developed to predict the accurate crop requirement, and crop production [31]. There are three factors that affect the estimation of ET_c: The Climate, Growth of crop, and Type of crop [32]:

Climate based ET_c

It is determined by climatic factors including Solar Radiation, Temperature, Wind, and Humidity as well as environmental factors. The evaporation account for approximately 10% of the overall evapotranspiration and the transpiration of crops constitutes the remaining 90% [33]. Hence, there is a need to estimate ET_c in area's of hot, dry, windy, and sunny seasons. Value of ET_c is lowest during the season of cold, humid and cloudy season without wind. Climate impact on crop water requirements is determined by reference crop evapotranspiration ET_o and generally expressed per unit of time in millimeters i.e. mm/day, mm/month, or mm/season.

Crop Type and Crop Growth Stages

This section presents the influence of the type of crop and stages of growth in crop water requirement. The daily and seasonal crop water needs are influenced by the type of crop and that affects the accuracy of ET_c estimation. For example, rice crop requires more water than maize crop [32].

The crop evapotranspiration ET_c is calculated by multiplying a crop coefficient (K_c) with the reference evapotranspiration ET_o.

The (K_c) value is sensitive and depends on several aspects such as type of crop, weather variables, canopy cover density, stage of growth, soil-moisture and agricultural operations.

However, K_c method has the capability to determine the actual crop evapotranspiration ET_c precisely. The two-step crop coefficient K_c reference evapotranspiration has been a successful method to predict the evapotranspiration ET_o and crop water requirements. The total growing period is divided into four stages of growth shown in Fig. 1.5 [34] [35]. According to the FAO methodology, the four growing stages of a crop are the initial, development, mid-season and end-season stages Allen et al. [8].

- The initial stage: The period ranges from transplanting or sowing till crop covers the 10% of ground.
- The crop development stage: The period begins at the completion of the initial stage and remains until the full ground has been covered (ground cover 70-80%).
- The mid-season stage: The period begins at the completion of the crop development stage and remains till maturity; it consists of grain-setting and flowering time.
- The late-season stage: The period starts at the completion of mid-season stage and remains until the last day of the harvest; it consists ripening of the crop.

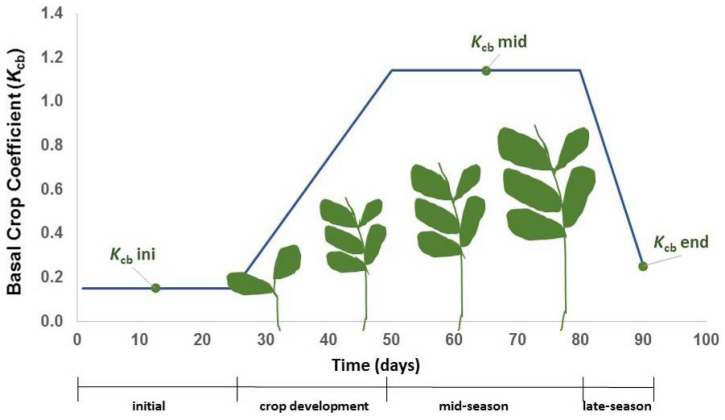

Figure 1.5: Various crop-growth development stages

The idea of crop coefficient K_c is proposed by Jensen [36] and has been improved by many researchers like Doorenbos [37]; Burman et al. [38]. The crop coefficient method can be expressed as follows in Eq. (1.2):

$$ET_c = K_c \times ET_o \tag{1.2}$$

where ET_c represents the requirement of crop water (mm d^{-1}), ET_o the reference crop water requirement (mm d^{-1}), and K_c the crop coefficient. Besides, Kingra et al. [39] computed crop water requirement for Wheat and transplanted Rice at Ludhiana, reported that the Wheat crop used about 315 mm water whereas the Rice crop used about 780 mm water during its growing season.

The K_c values of $Wheat_1$ crop as 0.4, 1.15 & 0.4 and $Wheat_2$ as 0.5, 1.36, 1.42, and 0.42 are considered in the initial, mid and last stage of growth respectively. The length of time in (days) for four stages of $Wheat_1$ as 29, 55, 14 & 32 days while for $Wheat_2$ as 24, 46, 35, 42 days are considered.

The K_c values of 0.7, 0.85, 1.15 & 1.05 for Maize crop are considered at different stages. The length of time (days) for four stages of Maize are 35, 18, 17, and 15 days used.

1.4 Irrigation Scheduling

Howell explored the irrigation schedule for effective and efficient use of water [40]. This efficiency can be enhanced by using advanced methods of irrigation. For the accurate estimation of crop-water requirement, we deduct the amount of effective rainfall from the total amount required for a crop. It is calculated in mm/month or mm/day. The main goal of good irrigation scheduling is to apply the correct amount of water at the right time, and make sure that water becomes available when the crop requires it.

According to predetermined schedules, the irrigation water is supplied to the cultivation by keeping in mind the status of the soil and the need of crop [41]. Judicious and accurate usage of water for crop production needs a knowledge of the quality of water, soil, weather, crop, and drainage situation. The soil type and climate condition plays a vital role for determining correct amount of water to be supplied to any specific crop.

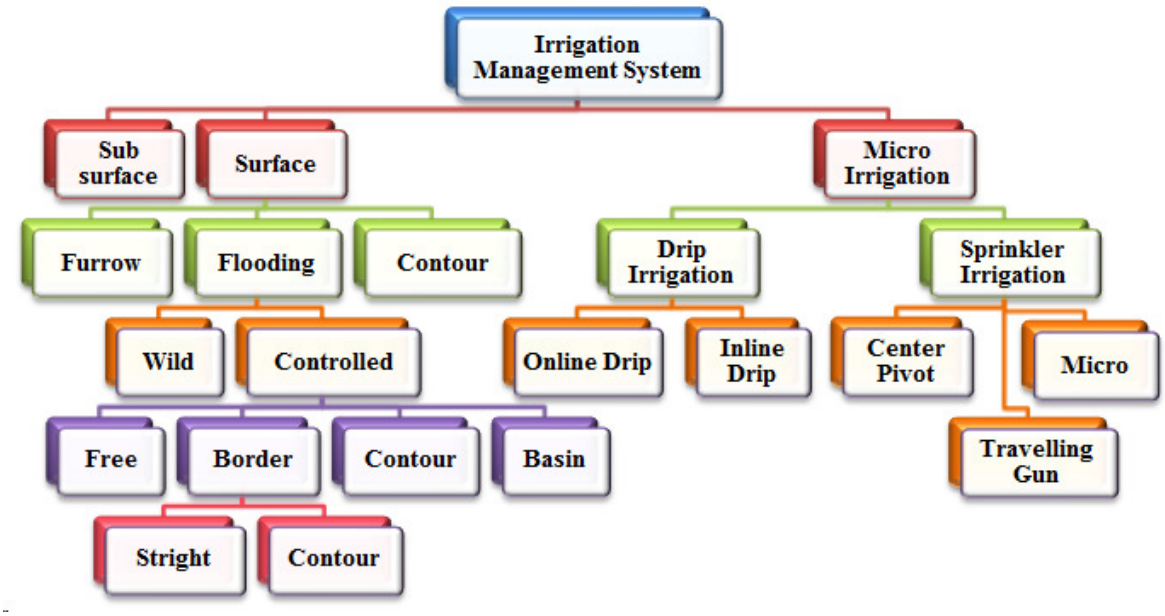

Figure 1.6: Types of irrigation scheduling

Irrigation scheduling is a very time consuming and complex process [32]. However, the advent of advanced technology has made it simpler and the water supply can be scheduled precisely to meet the water requirements of crop. Application is required to fix date and time of irrigation for a specific crop and to fix the quantity

of water required for irrigation. Otherwise, excess water would reduce the efficiency of its usage and hinder the development and growth of crop.

Methods of Irrigation Scheduling

Different methods of irrigation can be applied which have their own benefits and limitations. Earlier, watering is being done by the use of bucket. But it is very time-consuming and difficult method. More advanced water application are available such as i.e. sprinkler, or drip irrigation, depending upon following aspects:

- Natural conditions
- Type of crop
- Type of technology
- Previous experience with irrigation
- Required labor inputs
- Costs and benefits

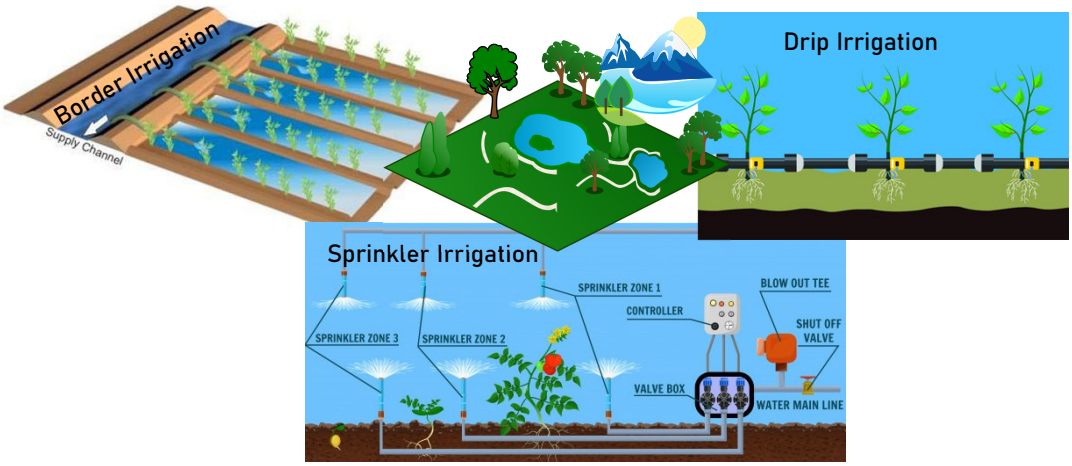

Figure 1.7: Irrigation methods

Surface irrigation or irrigation methods may be categorized based on mode of water application as depicted in Fig. 1.6. [42]: and Fig. 1.7 shown the different types of irrigation methods.

Border irrigation: Border irrigation is a traditional method of surface irrigation. Surface irrigation is where the water is applied to the surface of field by gravity flow. The water is applied into small channels (furrows) or complete surface is flooded with water (basin irrigation) or strips of land (borders). Borders are uniformly long, graded strips of land and, separate by earth bunds. They are also

known as border strips. There are several ways to fed the irrigation water to the border such as by using small gates, launching the channel bank, and using spiles or siphons. A sheet of water is guided by the earth bunds and flows down the slope of the border. It is mostly suited to cover long uninterrupted field lengths to facilitates machine operations in the huge mechanized farms.

Sprinkler Irrigation: It is equivalent to natural rainfall, where water is pumped using a pipe system and then rotating sprinkler heads used to spray onto the crops. Further, the spray is applied into the air via sprinklers to breaks it into small drops of water that fall on the grounds uniformly.

Drip Irrigation: It involves application of dripping water to the soil from a small diameter plastic pipe systems attached with outlets called emitters or drippers at very low rates (2-20 liters/hour). It is also called as trickle irrigation. Water is applied very close to plants so that only part of the soil in which the roots grow is wetted, unlike sprinkler and surface irrigation [32]. It is well suited for trees, vine crops and row crops (vegetables, soft fruit).

Figure 1.8: The need for irrigation scheduling

Need of Irrigation Scheduling

Hydrological, Climatologist, and Agronomical processes are required for the development of irrigation system. These studies are mainly developed to estimate daily, weekly, or monthly evapotranspiration. The precise approximation of evapotranspiration is an important process that plays a key role in crop planning, deployment and production of irrigation systems.

In recent years, several approaches have been developed to overcome the problems and obstacles that occur with smart farming, such as species recognition, yield prediction, disease detection, drought, crop productivity problems and irrigation management as depicted in Fig.1.8.

So, there is a great need to explore research to enhance the scalability of irrigation scheduling based on advanced data analytic and machine learning. Some research has been done in the decision support system to improve the right decision on agricultural data.

Sustainable irrigation scheduling, enhances the crop yield, monitor the quality of water, protect soil quality, limit use of water, increase irrigation frequency, and minimizes water logging problems.

1.5 Decision Support System

Decision support system (DSS) was introduced in 1970's to assist users in complex decision-making processes and for making efficient use of irrigation water at the farm level [43].

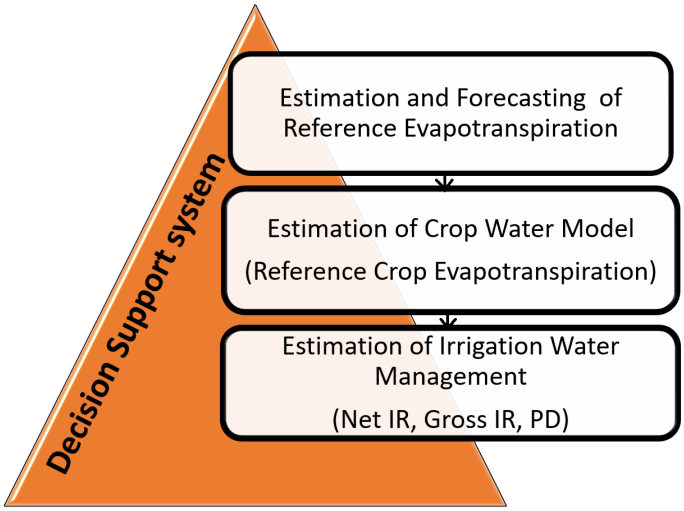

Figure 1.9: Process of DSS for crop water irrigation scheduling

The decision support system is an integrated approach to solve complex problems, combining the computer calculation and data storage capacities with human language and perception, support of mathematical model statistics, providing decision-maker. It is known as a primary tool in management for better decision making and environmental resources. Fig 1.9 presented the process of DSS for crop water irrigation scheduling and Fig. 1.10 shows the DSS for agriculture application.

Guariso et al. [44] introduced the concept of DSS. Various researchers surveyed the advanced use of the DSS management for utilizing water resources. DSS system decides about the requirement of water on the basis of knowledge of condition of soil, climate and type of crop.

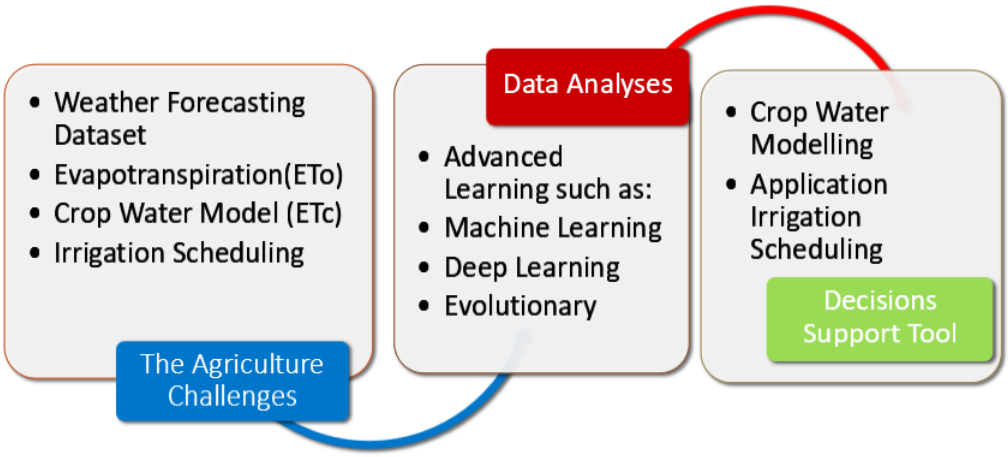

Figure 1.10: DSS for agriculture application

Statistics

"Statistic is the scientific method that works upon collected data to do analysis and representing the facts for making decisions in every domain.". Statistical learning is known as applied statistics, and is a powerful mathematical tool that is similar to predictive modeling in machine learning. Statistics can be used to prepare, summarize, visualize data for hypothesis testing.

It is very useful for us to explore the description and understand the variability. Bayesian method is used to estimate the hydrologic properties and irrigation needs for an under-constrained mass balance model. They presented an approach Markov chain Monte Carlo algorithm to solve for spreading of values for each unknown parameter in a conceptual mass balance model [45].

Machine Learning

"Machine learning (ML) is described as the scientific method which allows the machines to learn without being strictly programmed" [46]. From the last decade, machine learning and data analytics is becoming a wide research area in the domain of agriculture.

Machine learning models have demonstrated excellent results for crop-based modeling in recent days. There is a variety of machine learning models based on prediction for reference crop evapotranspiration. Various researchers applied the machine-learning and data analytics modeling to predict the crop evapotranspiration. Yamaç and Todorovic presented the satisfactory results of machine learning modeling using climate data [47]. A comparative analysis has been performed by Shiri, Nazemi, et al. [48] to estimate the ET_o using various intelligent models, namely ANN, ANFIS, Support Vector Machine (SVM), and GEP.

In the domain of agriculture, Big Data Analytics Technology have offered newly predictive models for ET_o estimation, e.g. Generalized Neuro-fuzzy models (Kisi et al. [49]), Artificial Neural Network (ANN), (Kumar et al. [50]), Adaptive Neuro-fuzzy Inference System (Tabari et al. [51]), Multi-layer Perceptrons Neural Network (MLPNN), (Zaji and Bonakdari [52]), Extreme Learning Machine (ELM) (Abdullah et al. [53]), Multivariate Adaptive Regression Splines (MARS) and Least Square Support Vector Regression (LSSVM) (Kisi [54]), GRNN (2016), ELM, WNN and GANN by Feng et al. [55]. Moreover, the Auto-ML technique is found to show excellent results while applying a border and sprinkler irrigation methods.

Deep Learning

"Deep learning (DL), sub-field of Artificial Intelligence (AI), is the most fostering domain considered for research and industry". In order to deal with the complex input-output data mapping, DL is significantly correlated with ML approach. DL technique handles the large amount of data applying on Artificial Neural Network (ANN).

Deep feed-forward neural networks are based on Multi-Layer Perceptrons (MLPs) published by Alexey Ivakhnenko and Lapa [56]. It can be used to model the complicated relationship between input and output due to its high hierarchical structure model training, construction and feature learning [57].

DL is applied in the hydrological and agricultural domain to overcome the difficulty of software such as data availability, cost factor, and its complexity for e.g. in application of modeling the crop evapotranspiration [40]. Brahma et al. [58] forecasted the solar-irradiance from NASA's POWER project based on DL model in two location of India. Rumar detection system using Deep learning (convolutional neural network) is proposed by [59].

ML and DL models are equivalent in artificial intelligence method to address the complex input-output relationships. [57]. However, DL has a benefit over traditional ML, due to its high hierarchical structure model.

Big Data Analytics

"Big Data Analytics (BDA) is a process of analyzing the big data to provide past, present and future statistics". The developments in Big Data Analytics provide a new paradigm and solutions for big data sources, storage, and advanced analytics. The BDA helps in acquiring a deep understanding and useful insights of various sectors such as: Smart Agriculture.

Big Data is a fascinating new field which include advanced analytics, data science, statistics, and machine learning. Big data analytics have tremendous development benefits in the agriculture economy. Advanced big data analytics have improved

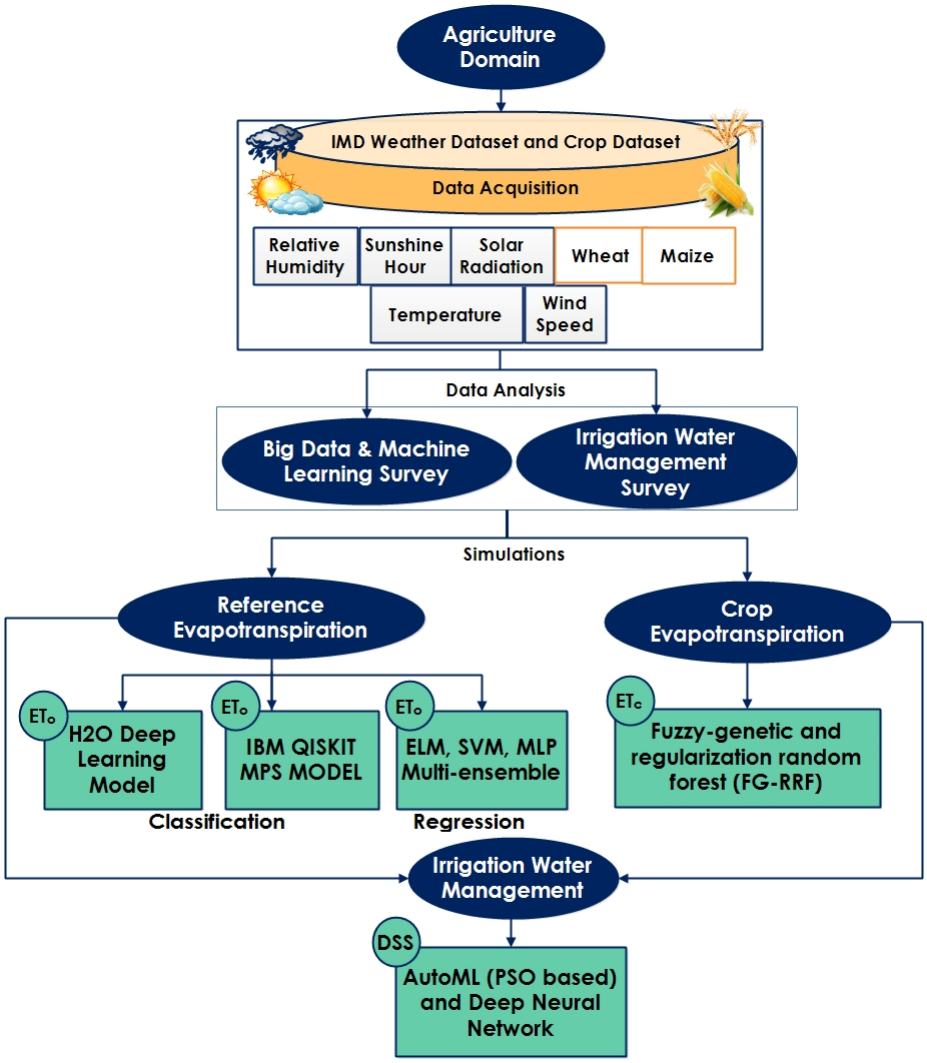

Figure 1.11: Schematic diagram of research work

the tools and technologies that changed the way of real-time applications to make better decision provide high-performing platform for efficient analysis, capturing, storing and managing large scale of big data. In addition, agriculture practices are becoming increasingly data-derived and data-enabled with the recent development of 5G in Internet of Things, Machine Learning Analytics, Big data and Artificial Intelligence approaches [60].

The goal of the research is to develop irrigation management protocols that will lead to improved decision system (allocation, application, and optimization). This study is expected to provide a decision tool that will assist irrigators and water managers in determining the reference evapotranspiration (ET_o), Crop Water Requirement (CWR), Irrigation Water Requirement (IWR), and irrigation scheduling for more effective water allocation and its application. Fig. 1.11 presents the schematic diagram and structure of research work.

1.6 Objectives

Based on extensive Literature Survey, following objectives were outlined:-

- To analyze existing empirical models for estimating evapotranspiration ET_o, and ET_c for irrigation scheduling.
- To develop a crop water model for estimating changes in evapotranspiration for varying meteorological data of different locations.
- To develop an on-farm smart irrigation decision support system of different irrigation scheduling.
- Validation of the proposed decision support system and comparison with existing models.

1.7 Contributions

The Thesis contributions in the following ways:

- Theoretical Contribution 1

  From the theoretical point of view, a comprehensive investigation has been conducted by using Big Data Analytics for gaining useful insights and attaining a deep understanding of several sectors such as healthcare, agriculture, smart cities, cyber-physical system, and social media analytics, etc. A systematic and extensive methodological review has been given on several technologies and tools of BDA and also described the research gaps for further investigation.

- Theoretical Contribution 2

  Beginning with the introduction to smart agriculture, this study provides an overview of irrigation scheduling which includes reference evapotranspiration, crop evapotranspiration, and crop water need, using various decision support system technologies. This study introduces the theoretical framework that is necessary to understand this process and the following chapters. Machine learning, Deep Learning, and Big Data Analytics have been investigated for the irrigation scheduling application.

- To calculate Reference Evapotranspiration (ET_o), H2O model framework is applied to determine the ET_o for Patiala station, and Hoshiarpur station of Punjab for classification. In this work, the Missforest imputation algorithm has been applied to impute missing values in the original dataset. Four data-driven models (such as Random Forest, Deep Learning, Generalized Linear Model,

and Gradient Boosting Machine) have been analyzed under H2O framework. The estimation of daily ET_o using DL, RF, GLM, and GBM models is done based on performance metrics such as Pearson Correlation (r), Root Mean Square Error (RMSE), Normalized Root Mean Square Error (NRMSE), Root Mean Square Logarithmic Error (RMSLE), Coefficient of Determination (r2) or (R), Mean Square Error (MSE), Nash–sutcliffe Efficiency (NSE), Accuracy (ACC) and Mean Per-class Error (MCE) using the training, validation and testing datasets respectively. The main results of this contributions are as follows:

  - DL model provided the most accurate results among the considered models with average (RMSLE=0.0693, r=0.98, r2=0.99, NSE=0.98, NRMSE= 13.90%, RAE=0.0346) for Hoshiarpur station.
  - GBM model is a slightly better than GLM model with average (MSE=0.0522, RMSE=0.2285, LL=0.218, ACC=0.932, MCE=0.0842 and RMSLE=0.0894, r=0.98, r2= 0.97, NSE=0.97, NRMSE=16.80%, RAE=0.0508), but GBM gives over-fitting in case of training dataset.
  - GLM had slightly better accuracy than RF with (MSE=0.0641, RMSE=0.253, LL=0.235, ACC=0.93, MCE=0.0848) and (r=0.96, r2= 0.97, NSE=0.97), NRMSE =17.80% and RAE=0.0567).

- To calculate Reference Evapotranspiration (ET_o), a Matrix Product State (MPS) classifier has been proposed to classify classical and quantum data. To illustrate the validation of the developed model, a ibmqx4 quantum computer simulation has been applied in order to determine the accuracy of model, and the experimental outcomes shows the superiority of the proposed model through MPS circuits.The main results of this contribution are as follows:

  - The accuracy of the testing data set of $Agri_1$ is just slightly greater. It has been investigated that the training accuracy of the $Agri_2$ and $Agri_3$ samples is marginally higher than the testing samples, respectively.
  - In the case of the training data set of the $Agri_1$ sample, the specificity is approximately 0.98, that is, the MPS classifier identifies more negative results compared to the testing set's 0.76. Therefore, the true positive value of the training set is less than that of the testing set for $Agri_1$.

- A novel multi-level ensemble model is developed using various ML approaches to ensure the generalization, and robustness of the model. Three different models Extreme Machine Learning (ELM_1, ELM_2, ELM_3, ELM_4), Multi-layer Perceptrons Neural Network (MLP_1, MLP_2, MLP_3, MLP_4), Support Vector

Machine (SVM_1, SVM_2, SVM_3, and, SVM_4) based on various parameters (such as activation function, kernel functions, and hidden layers) used to estimate the reference evapotranspiration (ET_o). We presented a study to find accurate quantification for ET_o, using CROPWAT8.0 software with Ludhiana station. The ensemble-SVM model achieved good accuracy (99.46% to 99.72%) to predict the daily ET_o and the correlation coefficient is closed to 1 during training, validation, and testing datasets. The RMSE of (0.0085 to 0.0935) and MAE of (0.0614 to 0.0639) results are found minimum as compared to other ensemble ML models.

- An innovative multilevel ensemble model based on Regularization Random Forest (RRF) and Fuzzy-Genetic (FG) model is proposed for accurate estimation of crop coefficient (K_c) and reference evapotranspiration (ET_c). The fuzzy approach ensures that the rules determining the (K_c) and (ET_c) are always comprehensible, accessible, and adjustable to achieve the target of saving water. This study presents the water requirement of three crops namely (Maize), ($Wheat_1$), and ($Wheat_2$). The main results of this contribution are as follows:

  - The ET_c is ranged from 1.22 to 7.05 mm day-1 for maize crop, 0.65 to 3.70 mm day-1 for $wheat_1$ crop, and 1.22 to 7.05 mm day-1 for $wheat_2$ crop. In this respect, our analysis depicted that the models have high performance for modeling daily K_c and ET_c (e.g. MSE= 0.0134-0.156, RMSE= 0.1160-0.396, r2= 0.830-0.99, ACC= 94-99) in testing set.

  - During the growth stage of Maize crop, the Actual total water requirement is recorded as 263 mm and Predicted as 255 mm. For $Wheat_1$ crop Actual requirement recorded as 245 mm and Predicted as 255.1 mm, for $Wheat_2$, Actual and Predicted requirements are recorded as 191 and 190 mm, respectively on using testing dataset.

  - Overall results of model simulation performance of K_c and ET_c methods using the proposed model showed that Maize crop performed better than the other crops ($Wheat_1$ and $Wheat_2$) in training and testing scenario, respectively.

- To calculate irrigation need, the Auto-ML based on Particle Swarm optimization (PSO) and Deep Learning (DL), have been applied to estimate the three parameters of irrigation scheduling. The net irrigation, gross irrigation, and pumping discharge are used to estimate the water requirement for Border irrigation and Sprinkler irrigation methods.

1.8 Thesis Organization

The research work discussed in this thesis is intended to establish a few analytic techniques and DSS for irrigation water management. The proposed work encapsulates the approaches which are weaved into Eight chapters. This thesis is devoted to design a crop water model for estimating changes in evapotranspiration by using data of different meteorological stations.

Chapter 1: Introduction

An approach towards better smart agriculture, this chapter presents an overview of irrigation scheduling using advanced technologies. This chapter provides coverage on some aspects of smart agriculture like an introduction to the concept of smart water management, reference evapotranspiration, crop evapotranspiration and irrigation scheduling. The goal of the research is to develop irrigation management protocols that will lead to improved irrigation decision systems (analytic and modeling application). This chapter elaborates the related concepts of the new paradigm of agriculture application based on decision support systems using machine learning, deep learning, and big data analytics.

Chapter 2: Literature Review

This chapter provides a literature survey of the existing reference evapotranspiration (ET_o), and crop evapotranspiration (ET_c) to estimate the water requirement for different crops in irrigation scheduling (IS). In this chapter, we explain methodological surveys on various machine learning methods, statistics, analytics, and deep learning for estimation of (ET_o), (ET_c) and (IS). We examined how such developments can be leveraged to develop and execute the next generation of data, models, analytics, and decision support tools for agricultural irrigation water systems.

Chapter 3: Reference Evapotranspiration ET_o using H2O Framework

In this chapter, we have investigated H2O model for classification of data. We presented a H2O model framework to determine the daily ET_o for the Hoshiarpur and Patiala districts of Punjab. The four supervised learning models: Deep Learning-Multilayer Perceptrons (DL), Generalized Linear Model (GLM), Gradient-Boosting Machine (GBM), and Random Forest (RF) are applied to predict the overall ability to predict the future ET_o.

Chapter 4: Reference Evapotranspiration (ET_o) using MPS

The concept of matrix product state classifier is proposed and applied to classify the quantum encoded data. The binary classification of the classical machine learn-

ing data set is applied on the agriculture dataset encoded into a quantum state. The main advantage of MPS quantum classifier is that training can be implemented with high efficiency by considering the various parameters on the ibmqx4 quantum computer. Furthermore, it is implemented using agriculture dataset to classify the ET_o for the Patiala station.

Chapter 5: Reference Evapotranspiration ET_o using Multi-Ensembling

In this chapter, we have investigated some regression based models for modeling the ET_o for Ludhiana station. We have applied three machine learning models, and a Multi-ensembling technique to build the model accurately. Recently machine learning (ML) techniques like Extreme Machine Learning (ELM), Artificial Neural Network (ANN), Support Vector Machines (SVM) are being widely used for modeling the process of evapotranspiration. The proposed framework of the Multi-level Ensembling ET_o Forecasting Model (MLE-ET_o) consists of five phases: (i) Input dataset, (ii) Partitioning of dataset, (iii) Model development, (iii) Multi-level ensembling for ET_o, and (v) Quantified prediction output.

Chapter 6: Crop Evapotranspiration ET_c using FG-RRF

We developed an innovative multilevel ensemble model for accurate estimation of crop coefficient (K_c) and reference evapotranspiration (ET_c) using Regularization Random Forest (RRF) and Fuzzy-Genetic (FG) models. This chapter presented the water requirement of three crops namely (Maize, $Wheat_1$, and $Wheat_2$). The developed model is applied to examine the weather data collected by IMD, Pune, and PAU, Ludhiana (case study) for decision making in a crop water model.

Chapter 7: Decision Support System for Irrigation Scheduling using AI (DSS-IS)

We have explored artificial intelligence and evolutionary algorithms for estimation of the Net Irrigation Requirement, Gross Irrigation, and Pumping Time Discharge: Using Deep learning and PSO based Deep Neural Network developed using a case study of Border Irrigation and Sprinkler Irrigation systems for Ludhiana station. We have divided each dataset further into sub-parts. Therefore, we got higher accuracy in the case of border irrigation where we have applied three sub-parts of the dataset and the Pumping time discharge parameters has given higher accuracy using the proposed model.

Chapter 8: Conclusion and Future Research

The motivation behind future research work, well-informed and a comprehensive examination, and realistic analysis of deploying machine learning successfully in

industries have been given. Data analytics has great potential in the agriculture era, if the water irrigation management research is combined with the modeling approach using machine learning analytics and big data, the research level will be achieved at different levels of the agricultural development sector.

Finally, this chapter is the conclusion of the Thesis. We present a technical summary of all results and outline current limitations together with future research directions.

1.9 Summary

This chapter provides an overview of the irrigation water scheduling. It presents the concept of reference evapotranspiration and crop evapotranspiration for crop-water modeling. It also presents the various methods of irrigation scheduling and address the need of decision support system and its various approaches that lead to irrigation water management. In the last, this chapter highlights the contributions and thesis organization.

Chapter 2

Literature Review

This chapter provides a literature review in the domain of Statistics, Machine Learning, Deep Learning, and Decision Support System of an Irrigation Scheduling. In the first section, concept of reference evapotranspiration (ET_o), crop evapotranspiration (ET_c) and irrigation water scheduling methods have been described. Furthermore, the existing literature with various empirical models and analytics approaches have been presented.

2.1 Overview

Smart Climate based agriculture provides a necessary contribution to farming in terms of quantity and quality of crop production. Change in climate and variations in meteorological aspects influence the water requirement of crops, and evapotranspiration. However, there are some difficulties in monitoring the growth of crop and irrigation management system.

In climate of arid and semi-arid, accurate estimation of evapotranspiration (ET_o) can provide a scientific basis for developing irrigation scheduling. Quantification of transpiration (T) to evapotranspiration (ET_o) from crop is critical in the irrigation scheduling. However, few investigations have explored the seasonal variability of (T), (ET_o), and provides the comparison of ET_o with different methods using various climate conditions [1] [29] [61].

Several computer simulation techniques, and decision support systems have been developed to estimate ET_o, ET_c and crop water requirement (CWR). It is also important to identify changes in the hydrological cycle when we want to predict the impacts of climate change. However, current studies on climate change must be expanded to cover the entire globe.

2.2 Estimation of Reference Evapotranspiration (ET_o)

The application of Evapotranspiration (ET_o) in irrigation scheduling is divided into different categories for literature section such as statistical, machine learning, evolutionary models, and decision support system [40], [62], [63]. We have presented a comprehensive review literature for reference evapotranspiration as follows:

ET_o is an imperative aspect of the hydrological cycle that is stirring water availability on the earth surface. It is one of the significant criteria of accurate quantification of crop water requirement that influence various hydrological processes, planing of water management and resources [51], and requirement of irrigation [64]. Traditionally, the ET_o is estimated at the field scale, but it consumes lot of time and is difficult to process by complex mathematical calculations with various climatic variables. Methods for measuring evapotranspiration from meteorological data include a number of climatology, and the field inputs which is directly estimated in weather stations. Some parameters are associated with measured data, where as others can be obtained directly or through empirical methods.

Empirical based methods

Since many years, various researchers have established reference evapotranspiration ET_o estimation with empirical methods. There are few categories of ET_o estimation methods: Temperature-based, Radiation-based, Empirical, Pan, and many more. Commonly, FAO-56 Penman-Monteith method is applied as the scientific, standard and temperature based method to estimate the ET_o [8] [65]. FAO-PM has been extensively adopted because to its positive outcomes in a variety of climates across the world. However, it needs a significant amount of meteorological data obtained from regular meteorological observation stations [66].

To overcome the existing limitation of the FAO-PM model, various attempts aiming to estimate ET_o with limited observed data have been made. A large number of studies have focused on estimating ET_o using empirical methods with limited ground-level data such as the Hargreaves and Samani equation, Priestley–Taylor equation, and Thornthwaite equation have been used for estimating (ET_o) by Tomas-Burguera et al. [67].

ET_o is estimated with simplified or empirical methods (e.g. Lysimetric measurements) and it is highly difficult to achieve more precise and robust approaches [68] [69]. HS equation is the most simple and accurate approach based on temperature [27], [8]. There are many empirical approaches to predict the ET_o using five mass transfer-based models (Ivanov, WMO, Penman, Trabert, and Mahringer), five radiation-based approach (Tu, PT, Ab, JH, and Mk), and five temperature-based

approach (HS, modified Hargreaves-Samani$_1$) (Th, BC, MHS$_1$, and MHS$_2$) [70]. Table 2.1 shows the empirical methods for estimation of ET_o.

Table 2.1: Estimation of evapotranspiration ET_o with empirical methods

Author	Method	Parameters
Malamos et al. [71]	FAO Penman-Monteith	T_{mean}, u_2, R_H, latitude ϕ, altitude (m), I_s
Tegos et al. [72]	FAO Penman-Monteith	T_{mean}, u_2, R_H,I_s, R_s, R
Ficklin et al. [73]	Penman, Monteith, Thornthwaite, Hamon	T_{mean}, u_2, R_H,I_s
Yang et al. [61]	FAO-56 Penman-Monteith, Hargreaves-Samani, Reduced-set Penman Monteith (RPM)	T_{mean}, T_{max}, T_{min}, RH_{mean}, u_2, I_s, VPD.
Heydari, et al. [74]	Blaney-Criddle	T_{mean}, RH_{min}, u_2, I_s, ρ, α, and β coefficients.

Malamos et al. [71] investigated the monthly Geo-spatial ET_o with FAO Penman-Monteith using line, polygon, and point through a geometry independent algorithm. They selected various climate parameters such as T_{mean}, u_2, R_H, latitude ϕ, altitude (m), I_s.

Tegos et al. [72] applied the radiation based model to calculate the daily potential evapotranspiration ET_o with FAO-56 Penman-Monteith using T_{mean}, u_2, R_H,and I_s parameters. The Potential ET_o is estimated for current and future drought condition using PDSI tool, Spatial, and Temperature based models such as Penman-Monteith, Thornthwaite, and Hamon methods. They have selected T_{mean}, u_2, R_H, I_s input variable, and analyzed on MATLAB GUI [73].

Yang et al. [61] analyses the daily reference evapotranspiration (ET_o) using short-term forecasting with FAO-56 Penman-Monteith equation, Hargreaves-Samani equation, and Reduced-set Penman-Monteith (RPM). The R language is used to simulate with various climate parameters including T_{mean}, T_{max}, T_{min}, u_2, RH_{mean}, I_s, and Vapor pressure (ea).

Machine learning and Evolutionary models

There have been many studies on hybrid models with machine learning, and evolutionary algorithms to estimation of ET_o with few climate parameters around the world.

Patil and Deka [75] investigated the performance of extreme machine learning (ELM) to quantify the weekly ET_o in the Thar Desert of India. Also, they have

showed the comparison of Artificial Neural Network (ANN) with three input variable. The ELM model gives slightly higher accuracy than empirical, and ANN models.

Wu et al. [76] proposed hybrid model using machine learning and soft-computing to estimate the monthly ET_o in south China with 26 data stations. The proposed (Kmeans-FFA-KELM) approach developed with the three approaches (K-means clustering, Firefly Algorithms, and Kernel Extreme Learning Machine model) found higher accuracy using three input variables ($Temperature_{max}$, $Temperature_{min}$, and R_a). Another study showed performance of six remote-sensing based ML models to predict the daily ET_o in the Andalusian. The two ELM and MLP models found higher accuracy than RF, SVM, GRNN, and XGBoost models [77].

Stacking and blending ensemble based ML models are used to calculate the daily ET_o with limited input variables. Two-layer ensemble model is build with RF, SVR, MLP, LR and KNN models and found higher accuracy in terms of R^2 ranged from (0.66 to 0.99) as compared to empirical models [78]. Another, ensemble based model is build with ANN, SVM and RF to estimate the ET_o with geno-types and optimize the ET_o with time series data, and found the correct results [79].

Bai et al. [80] proposed ensemble-based four ML models with RF, BMA, KNN, SVM and MLP to calculate the ET_o. The MLP-based ensemble model provides the efficient results in terms of $R^2$from (0.69 to 0.71) and RMSE(23.0 to 25.0). However, the ML and DL based models are proposed to estimate a urban ET_o with Flux Footprint, Remote Sensing and Geographic Information System (GIS) data. The RF model provides slightly better result in terms of (R^2 of 0.840 and RMSE of 0.0239 mm/h) than CNN model [81].

Adnan et al. [82] demonstrated the capability of different Neuro-Fuzzy methods to estimate the pan evaporation monthly using climatic inputs of different parameters obtained from Uttarakhand, (India). Recently, Adnan et al. [83] demonstrated the capability of dynamic evolving Neural-Fuzzy Inference System (DENFIS) and Least-Square Support Vector Regression with a Gravitational Search Algorithm (LSSVR-GSA) for estimating ET_o.

It has been shown that the extraterrestrial radiation or temperature-based LSSVR-GSA models are superior to DENFIS model for estimating monthly ET_o. They [84] forecasted the monthly and daily stream flows of poorly gauged mountainous watershed with Fuzzy Genetic Algorithm (FGA), Least Square Support Vector Machine (LSSVM), and M5 model tree (M5T) models.

Heddam et al. [85] estimated and compared daily reference evapotranspiration (ET_o) using the Online Sequential Extreme Learning Machine (OSELM) and Optimally Pruned Extreme Learning Machine (OPELM) in the Mediterranean region of Algeria. The OPELM models showed good performances as compared to OSELM

models.

Recently, Tikhamarine et al. [86] combined the Support Vector Regression and Grey Wolf Optimizer (SVR-GWO) to predict the monthly ET_o at Annaba, Algiers, and Tlemcen stations in North Algeria. Moreover, the proposed model is compared with the existing variants of SVR and showed that the performance of the SVR-GWO gives occasionally competitive and very promising results. Maroufpoorb et al. [30] proposed the concept of hybrid Artificial Neural Network-Gray Wolf Optimization (ANN-GWO) model and predicted the ET_o for Iran.

Further, the proposed model showed more efficient and accurate results as compared to ANN and LS-SVR. Mohammadi and Mehdizadeh [87] proposed a hybrid of two models Support Vector Regression and Whale Optimization Algorithm to predict the daily reference evapotranspiration at three stations in Iran. It has been shown that hybrid models outperformed the support vector regressions models. Kisi [54] obtained weather dataset from Turkish Meteorological Organization (TMO) for 2 stations from 1982 to 2006 and applied MARS, LSSVR, and M5-Tree to estimate ET_o.

Valipour et al. [70] collected data for the period of (1961-2010) with 50 climate parameters from 18 regions of Iran to estimate monthly ET_o using five models namely (mass transfer, radiation and temperature based).

Mattar [88] obtained 32 weather stations of data from United Nations Food & Agriculture Organization (UN-FAO) known as CLIMWAT for (2013 to 2015) and presented gene expression programming (GEP) and empirical models to estimate ET_o. Tao (2018) [89] presented the hybrid intelligent ET_o model using data of three meteorological stations during 1998 to 2012. They used the Adaptive Neuron Fuzzy Inference System (ANFIS), Firefly Optimization Algorithm with ANFIS (ANFIS-FA) and Penman-Monteith models.

Co-active Neuro Fuzzy Inference System (CANFIS) model is proposed for modeling the monthly evaporation of Lake Nasser, Egypt [90]. The Gene Expression Programming (GEP), Support Vector Machine (SVM), Classification and Regression Tree (CART), the Cascade Correlation Neural Network (CCNNs), and are proposed for estimating evaporation by Yaseen et al. [91].

Falamarzi et al. [92] estimated the daily ET_o for water resources with ANN and WNN models from the period of 2009 to 2012. They have applied RMSE, APE, N.S., R^2 metrics to check the model accuracy with three input parameters such as T_{min}, T_{max}, and u_2. Models LS-SVM, MARS, and M5 models have been applied to estimate the Pan evaporation for Reservoir and water resources management. They have applied four input variable, T_{mean}, R_s, u_2, and R_h with dataset of period 1986 to 2006 [93]. Another, ANN model is applied to forecast the ET_o for application of real-time irrigation scheduling with T_{mean}, R_s, u_2, and R_h input parameters using

dataset of (2011 to 2012) [94].

Yassin et al. [64] analyzed the performance of ANN, and GEP models to quantify the ET_o with various climate parameters using dataset from 1980 to 2010. The GeneXpro Tools 5.0, and Propagation version 2.2.4 were used to developed the model and also provide the accuracy on the basis of these metrics MAE, RMSE, R^2, and OI. Gocic et al. [62] forecasted the ET_o using SVM, FFA, DWT, ANN, and GEP models with climate parameters for the period of 1980 to 2010. Goyal et al. [95] explored the four ML models namely ANN, LS-SVR, FL, ANFIS, and Gamma Test to estimate the pan evaporation for the duration of 2000 to 2010. They have found the efficient results with the FG and LS-SVR models using various climate parameters on MATLAB platform.

However, Chen et al. [96] found the best performance of Bayesian Model Averaging Model to estimate the terrestrial ET_o using KGE and Cubist software. Mehdizadeh et al. [97] proposed the hybrid model to estimate the monthly ET_o with GEP, SVM-Poly, SVM-RBF, and MARS models for duration of 1951 to 2010. The performance of the applied models is compared with the empirical methods, where the MARS and SVM-RBF models give the most accurate results. The hybrid ELM model revealed a superior performance to estimate the daily ET_o at the four major countries of (US, Germany, Belgium, and Sweden) using 9 years of dataset [63].

Mohammadi et al. [87] proposed an approach that couples Support Vector Regression with Whale Optimization (SVM-WO) to estimate the daily ET_o.The T_{max}, T_{min}, RH_{mean}, u_2, R_h, and SSD parameters are used to build the model and found accurate result with SVM-WO model.

In the domain of agriculture, ML offered new predictive models for ET_o estimation, e.g. Generalized Neuro-Fuzzy Models (GNFM) [49], ANN model [50], ANFIS model [51], MLP-NN [52] [98], ELM algorithm [53] [99], M5 Model Tree [98], LS-SVR , MARS, ELM, WNN and GANN [55]. The DL model and ML model are applied in various domains such the COVID-19 analysis [100], proposed a novel method based on artificial intelligence (AI) to identify COVID-19 disease [101], developed genetically optimized Deep Neural Network [102], Tripathy et al. [103] investigated the performance of MARPUF approach and it is found to be better resistant to such modelling attacks, image classification using deep learning [104], Artificial Intelligence approaches used to classifying various types of cancer [105], enhanced the grip functionality of myoelectric hands based on deep learning [106], and classifiers for on-line handwriting recognition based on SVM and KNN algorithms [107], and a survey for software fault prediction [108]. Singh et al. [109] presented the efficient results and reliable algorithm for optimal design of water distribution networks.The summary of ML and EA are given in Table 2.2.

Table 2.2: Estimation of evapotranspiration ET_o based on ML and EA

Author	Purpose	Algorithm	Data	Parameters
Patil and Deka [75]	Weekly ET_o	ANN, ELM	T_R(1970–2005) T_S(2006-2010)	T_{max}, T_{min}, RH_{max}, RH_{min}, R_s, u_2
Wu et al. [76]	Monthly ET_o	Kmeans-FFA-KELM	T_R(1966–2000) T_S(2001–2015)	T_{max}, T_{min} ,R_a
Wu [78]	Daily ET_o	ELM, MLP, RF, SVR, GRNN, XGBoost	T_R(1989–2008) T_S(2009–2018)	T_{max}, T_{min}, RH_{mean}, R_a
Bai et al. [80]	Daily ET_o	RF, KNN, SVM, MLP	FLUXNET 2015 47 cropland	T_{mean} R_n, u_2, NDVI, EVI, VPD, DTsR, TR, Pd, and WSF
Adnan et al. [83]	Monthly ET_o	DENFIS, LSSVR-GSA	T_R(1961–1986) T_S(2000–2012)	T_{mean} R_a
Heddam et al. [85]	Daily ET_o	OSELM, OPELM	T_R(2001–2008) T_S(2009–2013)	T_{max}, T_{min}, RH_{mean}, R_s, u_2
Tikhamarine et al. [86]	Monthly ET_o	SVR-GWO, SVR-PSO	T_R(2000-2009) T_S(2009-2013)	T_{max}, T_{min}, RH_{mean}, R_s, u_2

Author	Purpose	Algorithm	Data	Parameters
Maroufpoorb et al. [30]	Daily ET_o	ANN-GWO, LS-SVR	T_R(2012–2016) T_S(2017)	T_{max}, T_{min}, RH_{mean}, S_h, u_2, P_e
Kisi [54]	Monthly ET_o	LS-SVR, MARS, M5 Tree	1982-2006	T_{mean}, RH_{mean}, S_h, u_2, R_s
Mattar [88]	Monthly ET_o	GEP	2013 to 2015	T_{max}, T_{min}, RH_{mean}, S_h, u_2
Tao (2018) [89]	Monthly ET_o	ANFIS, ANFIS-FA, Penman-Monteith	1998-2012	T_{max}, T_{min}, RH_{max}, u_2, R_s, VPD
Yaseen et al. [91]	Evaporation	CART, CCNN, GEP, SVM	1999 to 2009	T_{max}, T_{min}, RH_{mean}, u_2, S_h, RF
Falamarzi et al. [92]	Daily ET_o	ANN WNN	2009–2012	T_{min}, T_{max}, u_2
Kisi [93]	Pan Evaporation	LS-SVM MARS M5 Tree	1986–2006	T_{mean}, R_s, u_2, R_h

Author	Purpose	Algorithm	Data	Parameters
Ballesteros et al. [94]	Forecasting ET_o	ANN	2011–2012	T_{mean}, R_s, u_2, R_h
Yassin et al. [64]	ET_o	ANN, GEP	1980–2010	T_{mean},T_{max}, T_{min}, RH_{mean}, RH_{max}, RH_{min}, u_2
Gocic et al. [62]	Forecasting ET_o	SVM, FA, DWT, GEP	1980–2010	T_{max}, T_{min}, e_a, u_2, I_s
Goyal et al. [95]	Daily Pan Evaporation	ANN, ANFIS, LS-SVR, FL	2000–2010	R_r, T_{max}, T_{min}, RH_{max}, RH_{min}, I_s
Chen et al. [96]	Terrestrial ET_o	BMAM	1982–2009	T_{mean}, T_{max}, T_{min}, RH_{mean}, u_2, R_s, R_n, VPD
Mehdizadeh et al. [97]	Montly ET_o	GEP, SVM-Poly, SVM-RBF, MARS	1951-2010	T_{mean}, T_{max}, T_{min}, RH_{mean}, u_2, R_s
Dou et al. [63]	Daily ET_o	ELM, ANN, SVM, ANFIS	2001-2009	T_{max}, RH_{mean}, R_{net}, T_{soil}

Finally, it maybe concluded that the various machine learning, soft computing, evolutionary algorithms, and ensemble-based models have been investigated to analyze and predict the ET_o using many climate parameters. However, the performance of deep learning and ML-based ensemble models could provide efficient results to forecast the reference evapotranspiration ET_o with limited climate parameters.

2.3 Estimation of Crop Evapotranspiration (ET_c)

Crop evapotranspiration (ET_c) one the most essential element of the hydrological system for irrigation scheduling. The crop coefficient K_c method multiplied with (ET_o) is most widely-used to determine the (ET_c) Eq. (1.2). Different estimations and methods having their own advantages and disadvantages are available. For the estimation of ET_c using Machine Learning, Deep Learning and Evolutionary Algorithms, some potential literature work are presented in this section.

Machine learning and Evolutionary models

The Back-propagation Neural Network (BP-NN) model is proposed to evaluate the crop evapotranspiration ET_c with combination of various climate parameters (T_{max}, T_{min}, RH_{mean}, S_h, RF and crop coefficient K_c). It is observed that the combination of Eddy Covariance method and BP model achieved the best accuracy in terms of R^2 (0.87) and accuracy (91.44%) than MLR model [110]. Mehta et al. [111] estimated the ET_o, ET_c and K_c of Wheat and Maize crops of Gujarat using climate data. They applied the various temperature and radiation based empirical methods to calculate and estimate the crop water requirement.

It is observed that the accurate value of Kc for wheat crop is more efficient as compared to FAO-56 Penman-Monteith method results. Whereas in case of maize crop the outcomes were found less accurate at Surat and higher outcomes as compared to FAO method at Bharuch station.

Elbeltagi et al. [112] presented the deep learning model to estimate the Wheat ET_c from 1970-2017 and forecasting the future changes from 2022-2035 of Nile Delta in Egypt using Visual Gene Developer technology. For calibration R^2 of 0.95, 0.96, 0.97 and for testing R^2 of 0.94, 0.95, 0.95 have been found efficient result respectively. Russo et al. [113] presented the MCMC and Bayesian algorithms to analysis the irrigation requirements for ground water mass balance with soil tensiometer of Rice crop. They have optimized the management decisions on crop replacement and increased the irrigation efficiency. The NN model and regression model are explored to estimate a greenhouse tomato crop yield, its growth, and efficiency in use of water with CropAssist and NeuralWare platforms [114]. Maurya et al. [115] developed a novel fuzzy-based energy-efficient routing protocol based on automated irrigation

system for maize crop on MATLAB platform. The FIS-DSS (Flexible Irrigation Scheduling Decision Support System) is proposed to analyze the optimal allocation of water resources of irrigation system. Fuzzy-inference and knowledge based user-friendly optimization tool is developed for Wheat and corn crops [116]. Chauhan et al. [117] proposed a web-based DSS to enhance irrigation water management for peanut crop on APSIM platform.

Table 2.5: Estimation of ET_c based on empirical, ML and EA

Author	Algorithm	Crops
Han et al. [110]	BP-NN	Wheat
Mehta et al. [111]	Empirical	Wheat, Maize
Elbeltagi et al. [112]	DL	Wheat
Russo et al. [113]	Bayesian, MCMC	Rice
Ehret et al. [114]	NN, RA	Tomato
Maurya et al. [115]	Fuzzy-based, Hybrid routing	Maize
Yang et al. [116]	Fuzzy Inference Model	Wheat Corn, Cotton
Chauhan et al. [117]	DSS	Peanut
Gavilán et al. [118]	Radiation, Makkink FAO-24	Strawberry
Tabari et al. [119]	ANFIS, SVM	Potato
Yamaç et al. [120]	NN, ABM, KNN	Potato

Gavilán et al. [118] measured the daily greenhouse crop evapotranspiration for strawberry and found more accuracy using empirical methods and sensors based on soil moisture. Tabari et al. [119] explored a ANFIS and SVM model performance for Potato crop evapotranspiration ET_c using meteorological data.

Yamaç et al. [120] applied the four scenarios based on features subset to accurately estimate the ET_c of Potato crop using ANN, ABM and KNN models. Further, ANN and SVM models are also applied to estimate the garlic ET_c by Abyaneh et al. [121] and the outcomes are found accurate as compared to lysimeter performance.

The need of precise estimation of crop water is an crucial aspect of agricultural planning and there exists several methods for determining ET in crop land [122]. The field based estimations are required and appropriate for monitoring the crop-water status at the land-scale level [123] [124].

The FAO-Penman and Penman methods are applied to forecast the ET for rice crop using meteorological data [125]. They have found the crop coefficients for initial, middle and late stage as 1.39, 1.51, and 1.43. The derivation and development of crop K_c are identified for castor and maize crops of Rajendranganagar by Reddy et al. [126]. Ko et al. [127] conducted a analyses report to evaluate the crop water requirement for cotton, and wheat crops at Uvalde, TX, USA. Fang and Ping [128]

presented an optimal solution to estimate the ET_o using interval regression analysis, crop water production function and Penman–Monteith method with LINGO software.

Numerous experiments have been conducted in recent decades to investigate the possible effect of climate change on evapotranspiration ET_c. For efficient crop evapotranspiration ET_c modeling using VIP (Vegetation Interface Processes) for Wheat and Maize [129], durum wheat in Tunisia [130], APSIM-Maize model [131], SEBAL model for yield, WUE, IWUE and HUE for Wheat crop [132], weighing lysimeter for Kc and ET_c [133], farm-level operational services in smart agriculture [134], crop water model based on Crop2ML framework [135] have been used.

The ET_c estimation results proved that the ML and EA approaches performed better than existing classical methods. However, several studies have investigated the estimation of ET_c with empirical methods. But limited studies have reported the estimation of ET_c using ML, and EA models as shown in Table 2.3.

2.4 Decision Support System for Irrigation Scheduling

This section considers the Decision support system (DSS) based on research that have included ET_o, ET_c, and irrigation requirement. An irrigation management system can offer farmers with appropriate decision-making tools in order to regulate the amount of water supplied to crops.

A decision support system PETP V2.0.0 is developed to analysis and estimate the potential evapotranspiration ET_o using various empirical approaches namely Hargreaves, Jensen-Haise, Penman-Monteith, Priestley-Taylor, etc. Visual Studio 2010 software is used to build the computational tool to estimate the accurate results for water requirement of crop [136].

Navarro et al. [137] developed smart irrigation DSS for managing the irrigation scheduling. They purposed 2 ML techniques i.e. PLSR and ANFIS. Maximum and minimum relative humidity, temperature, and direction of wind, global radiation, vapour pressure deficit, rainfall, dew point are the variables used to predict the daily ET_o with FAO Penman–Monteith method. Zizhong and Zenghui [138] presented a single irrigation system that enhanced higher corn production and also provided efficiency in use of water in Northeast China. They include climate parameters namely average, max and min of air temperatures, max and min of relative humidity, wind speed, and sunshine hours from 1980 to 2012. Penman–Monteith approach is used to determine the soil evaporation and ET_c.

The knowledge of the irrigation management has an impact on crop water requirements, maintaining water balance and is the practical considerations to enhance

productivity of crop [139]. Various research work in Punjab have demonstrated the requirements of crop water irrigation, irrigation water based on ET and pan evaporation, [140], but few studies have focused on Soil Water Deficit (SWD) [141].

Paraskevopoulos and Singels [142] investigated the integrated content of soil water recordings of real-time field into the MyCanesim system to estimate its use in 15 sugarcane fields of South Africa for supporting irrigation system. It is used to determine the decision making for irrigation scheduling based on the status of forecasts of crop, soil water, and the next irrigation date. Ying, Shaoyuan and Xianfang [143] represented the evaluation for summer wheat and winter maize cropping system for optimal irrigation scheduling. Further, they described topical versions of the SWAP and Wofost models for crop growing simulation and obtaining efficiency in use of water.

Afzal, Battilani, Solimando and Ragab [144] improved water resources management using different irrigation strategies and water qualities by field, and modeling study. To deficit irrigation PRD and RDI methods are used to estimate the effects of waste and fresh water on salinity distribution, soil moisture, and crop yield of potato, maize in Italy, Bologna through field experiments.

The fuzzy, evolutionary, and machine learning models are used to develop a DSS model for irrigation scheduling. Gaiqiang et al. [145] developed a FIS-DSS software based on knowledge, interface for user, and an inference engine for wheat, corn and cotton crops. It is a fuzzy interval programming model having multiple objectives and constraints, flexibility of model, data processing, and an alternative solving algorithm. The main objective is to maximize the economic-based benefits for cropland in China. The NN model is used to train the model with precipitation and historical climate parameters.

Giusti and Marsili-Libelli [146] developed a fuzzy-DSS to schedule the daily irrigation need of crop based on soil moisture as a predictive model and an inference model as irrigation decision maker. This model determines the actual need of water for kiwi, corn, and potato crops with past irrigation soil moisture, climatic parameters, and ET_c. They used meteorological data including temperature, solar radiation, wind, rain, etc. The objective of FDSS model is to reduce the water usage and provide the efficient result in terms of saving water up-to 13.55, 18.3, and 72.95 water units for irrigating three crops respectively. Sahoo et al. [147] proposed fuzzy multi objective linear programming approach for planing of land-water-crop system. The meteorological data like daily rainfall, evaporation, temperature, solar radiation, daily sunshine hours, humidity, wind velocity, and albedo are collected from the Central Rice Research Institute, Cuttack. The objective function is to optimize, maximized crop production, net return, and to minimize the labor requirement for various vegetables and pulses.

Reddy and Kumar [148] demonstrated a multi-objective method for the optimal crop pattern and multi-crop irrigation reservoir scheme by several procedures. Adeyemo and Otieno [149] explored a method to solve the multi objective crop planning model by an evolutionary algorithm. They have found excellent results in minimizing total water irrigation, maximizing the yield productivity, and net income from farming. Table 2.4 shows the different types of irrigation systems.

Table 2.6: Different types of irrigation systems

Name	Advantages	Application	Technology	Ref
Drip Irrigation	Efficient system, Saves water, Reduces nutrient leaching	Fuzzy Control for automatic greenhouse irrigation, DIDAS software for linearized water flow and scheduling	Matlab, Delphi (Embarcadero, Version XE3)	[150] [151]
Sprinkler Irrigation	Automatic Irrigation, Smooth fertigation and chemigation, No labor requirements	Multi-agent system for garden irrigation, Crop model AquaCrop for the optimization	Agent based simulation, AquaCrop Simulator	[152] [153]
Flood Irrigation	Usable on shallow soils, Low cost	The risk and sensitivity analysis of water, Energy and emissions in IR	Risk Software	[154]
Border Irrigation	Easy to design and maintain, Simple operations of the system, Natural drainage	Improved understanding of IR Models and measures basis for improving IR	SISCO	[155]
Furrow Irrigation	Accomplished, Minimal erosion , Adaptable to a large array of land slopes	Irrigation and fertigation in isolated furrow networks	C-language	[156]
Basin Irrigation	Small fields, Well suited for crops	Modelling and multi-criteria analysis of water saving	ISAREG, SRFR, SIRMOD	[157]

Irrigation water management is numerically intensive for computations and provides model interpretation and discretization. Neural networks and evolutionary algorithms demonstrated to estimate the irrigation volume and also determined the effectiveness to diminish irrigation application and maximize production [158]. Ortega Álvarez et al. [159] proposed a non-linear model to recognize yield schemes and water irrigation management plannings using the genetic algorithms. Further, they estimated crop yield, gross margin and production as a function of irrigation depth. Schmitz et al. [160] simulated 92 percent greater production for corn using evolutionary algorithm as compared to dynamic programming.

Application and web based DSS models are developed for the irrigation water scheduling by various researchers. Recently, a web-based DSS is proposed to es-

timate the soil-water balance for irrigation system with limited input parameters such as (dual crop-coefficient and meteorological). The irrigation parameters are computed through soil moisture and requirement of water. A web-based irrigation decision support system is introduced with limited inputs (WIDSSLI) for summer corn and winter wheat irrigation management in North China Plain (NCP). [161].

Table 2.7: Estimation of evapotranspiration ET_o, ET_c, and Irrigation with DSS systems

Author	Estimation	Developed Software	Approach	Crop
Cesar et al. [136]	ETo	PETP V2.0.0	Empirical Methods	-
Navarro-Hellín et al. [137]	ET_o, IR	DSS	PLSR, ANFIS	-
AL et al. [142]	IR	MyCanesim	-	15 Sugarcane
Ma et al. [143]	Optimal IR	DSS	SWAP, Wofost	Summer Wheat, Winter Maize
Yang et al. [145]	Crop-land Model	FIS-DSS	Fuzzy Interval Programming, Neural Network	Wheat, Corn, Cotton
Giusti et al. [146]	ETc, IR	fuzzy-DSS	Fuzzy	Kiwi, Corn, Potato
Alvarez et al. [159]	Yield, Gross margin	Non-linear Model	Genetic Algorithms	
Li et al. [161]	Soil-water balance IR	Web-based DSS	Dual crop-coefficient	Summer Corn, Winter Wheat
Rowshon et al. [162]	ETc	Climate-Smart-DSS	GCM	Rice

Table 2.5 presents the estimation of ET_o, ET_c, and irrigation with DSS systems. Antonopoulou et al. [163] presented an appropriate decision support system for crops which is implemented on web-based software. They introduced this approach by using the Java and PHP technologies for specific irrigation technique and soil improvement instructions. Dutta et al. [164] developed a mobile application based on sustainable irrigation DSS. They proposed cloud sensors based approach to evaluate the ground water usage and availability of water. This approach includes the CSIRO sensor based cloud computing organization and integrated big data that includes machine learning technologies. Bonfante et al. [165] proposed an irrigation water supply management tool to obtain the maximum yield of maize with W-tens, IRRISAT, and W-Mod approaches. W-Mod and IRRISAT models found more accurate results as compared to W-Tens in terms of irrigation water use

efficiency.

A Climate-Smart Decision-Support System (CSDSS) tool is proposed to evaluate the requirement of rice crop in Malaysia. They determine a daily crop-water balance based on input data (2010 to 2099) and (1976 to 2005) from Global climate models (GCM) and integrate with evapotranspiration using MATLAB simulator [162].

Ragab et al. [166] proposed a SALTMED model which includes the to partial root drying or deficit irrigation, subsurface irrigation, soil nitrogen fertilizer application, fertigation, dry matter production, plant nitrogen uptake, and nitrate leaching.

For the model calibration and endorsement the statistical measurements are used such as R2 coefficient, RMSE, and percentage error. The DIDAS software package for irrigation system decision-making strategies of drip irrigation systems is developed. [151].

A DSS framework have been introduced which includes 22 ET_o estimation approaches using user-friendly GUI (Microsoft Visual Basic 6.0) of 133 selected stations of India [167]. Potential-ET_o and FAO56-PM ET_o are used to estimate the ET_o in the Geisenheim Irrigation Scheduling (GS) for vegetable crops using sprinkler irrigation [168]. Ballesteros et al. [169] estimated the ET_o using FORETO software with Hargreaves Samani (HS) equation or the Penman Monteith (PM) and Artificial neural networks (ANNs) models. Modern platforms (.NET and Java) software applied to calculate the daily/monthly ET_o using meteorological parameters [170]. There are various crop simulation based models exists such as CropSyst [171], STICS (Brisson et al. 1998) [172], EPIC [173], DSSAT [174], VegSyst simulation model [175], and CERES [176].

A single approach, that can address all operational circumstances (weather information, crop growth monitoring, field data, agricultural expertise and infrastructure etc.), as well as the associated expenses for farmers which may limit the usage of these systems. Therefore, DSS's for irrigation scheduling have been developed to integrating various approaches in terms of (data collection from meteorological), pre-processing techniques and modeling based on empirical or artificial intelligence.

A bibliometric perspective of Irrigation Scheduling: The number of articles from digital database such as Science direct is considered. The irrigation scheduling, reference evapotranspiration and crop evapotranspiration articles are demonstrated from 1995 to 2021 as shown in Fig 2.1. We have selected only elsevier science direct library to find out research articles, where we found 34,083 results in area of irrigation scheduling, 32,775 papers in reference evapotranspiration and 14,261 results articles in crop evapotranspiration (crop water need).

Irrigation scheduling, ET_o and ET_c trends are represented by Green, Blue, and Yellow colored lines respectively. Fig 2.2 shows the publication year in the field of Irrigation Scheduling, ET_o and ET_c. Fig 2.3 shows the article type in which review

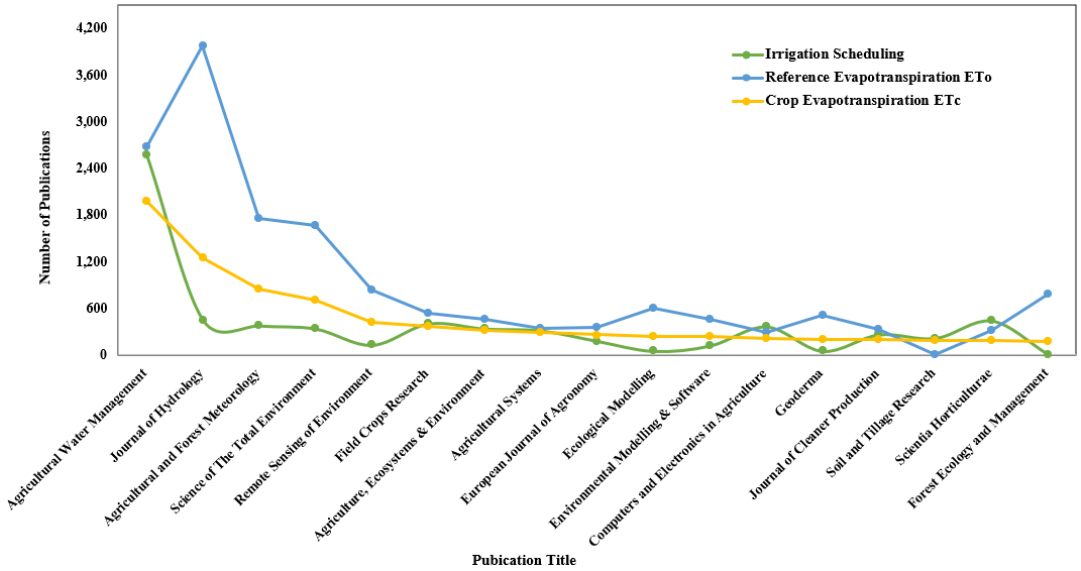

Figure 2.1: Number of publication in different journal type

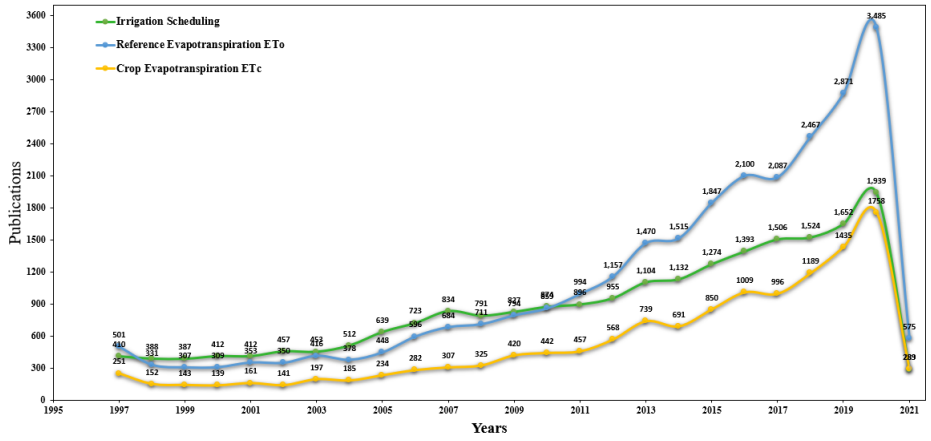

Figure 2.2: Timeline of publication year in domain of irrigation scheduling and evapotranspiration

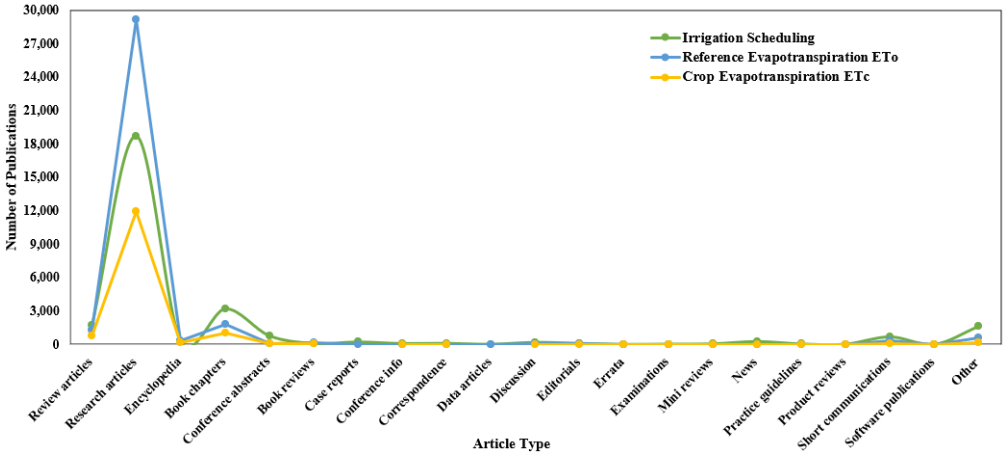

Figure 2.3: Number of publication articles types

reports, case reports, data articles, mini review, and many more articles.

2.5 Research Gaps

Based on the literature survey, some of the following research gaps have been identified:

- Internet of things sensors provides information about crop growth, yields, rainfall, soil moisture, soil nutrition, pest infestation, monitor their smart irrigation equipment's to farmers and connected harvesters to farmers. It generates lot of agriculture internet of things data and low usage of water efficiency, deficit and diffusion are the huge interrogations of presently improvement of irrigated agriculture.

- For the development purpose large areas of irrigation plantation is an onerous job. However, Irrigation scheduling techniques are introduced to solve these problems, which are based on parameters such as soil, crop and weather conditions. In traditional approaches, irrigation system is practiced by using manual control strategies and analyzing without real streaming dataset. Thereby, it is essential to analysis data on real-time streaming weather and effective rainfall dataset. However, after applying the appropriate approaches, irrigation decision system is still affected by soil monitoring, crops and weather information..

- Irrigation scheduling methods are based on soil moisture measurements, soil water balance calculations/meteorological approaches, plant water status monitoring or computer simulation. Several irrigation scheduling models using soil water balance calculation/meteorological approach, evaporation pan measurements, soil measurements are available. Other studies have also determined the potential use of crop water stress index for irrigation timing. However, irrigation scheduling using crop water models is limited.

- Effective irrigation requires management decisions that ensure an accurate estimation of crop and irrigation water requirements as well as an allocation irrespective of the irrigation scheduling methods. Accurate Evapotranspiration estimates are important in determining crop water requirements for appropriate irrigation scheduling.

- Crop water Irrigation scheduling has been investigated by various researchers. A lot of works is required to explore the potential of this irrigation scheduling in crop water management. There are incredible opportunities for soft computing

models and machine learning in irrigation. However there is need to explore advanced data analytics based on statistical methods, soft computing method and machine learning algorithms for calculation or prediction that need to obtain the optimal Irrigation scheduling.

- Today, DSS is required for crop on-farm irrigation water management due to the use of computer to control quality conditions of historical weather, crop, soil water, and effective rainfall. It can be used for determining and analyzing how much water is needed and when the irrigators should provide irrigation later.

- A DSS need to deliver a statistics methods, machine learning and precision farming using cloud based applications for big data analytics, visualization tool for allowing farmers to easily interrogate and analyze agriculture database, greatly enhancing their understanding of the key factors that lead to effective irrigation water management, for saving water and crop yield increasing.

2.6 Motivation

As technology rapidly spread in a few decades, precision agriculture is the key to fostering a new revolution in Irrigation scheduling. The United Nations statistical data indicate that agriculture consumes 70% of the overall use of water worldwide, compared with 20% for industry and 10% for domestic use [177]. To ensure the proper use of water supplies in irrigation we need more effective technologies. Automatic irrigation scheduling techniques replaced manual irrigation which was based upon crop water estimation. The crop evapotranspiration can be measured by weather indicators such as max-min temperature, humidity, solar radiations, wind speed and also the crop parameters such as the crop height, stage of growth, and the soil properties for the development of irrigation scheduling. The machine learning and deep learning advanced technologies provide direction and motivation to propose a novel application on crop water modeling.

2.7 Summary

In spite of the vast literature available, the subject of irrigation water management and crop water modeling for machine learning techniques are yet in its emerging phase. Although there is wide literature available on statistics, machine learning, decision support system for general manifolds. To estimate crop water modeling on general manifolds, we have different approaches available in literature.

All the approaches suffer from specific drawbacks such as model complexity while implementing approaches and the model-based DSS for irrigation scheduling. DSS is one of the key components for irrigation to increase agricultural water productivity by focusing on water applications during critical periods of crop water use. Various sensors and tools are used to validate the model accuracy and to improve via feedback using model-based DSS. Moreover, the water balance is well-entrenched approach for estimating irrigation amount and time (i.e. irrigation frequency) in irrigation scheduling. This approach is simple to use, typically inexpensive and very effective approach to estimates the ET_o, and ET_c.

The major objective is to adopt the several approaches to develop a flexible system that supports irrigation water requirement system, which may fit into diverse fields of operational activities (weather information, field data collection, crop coefficients etc.) This study consist of three main methodologies for Irrigation Scheduling, which are based on ET_o, ET_c and Irrigation Scheduling using DSS (Weather parameters, Crop Coefficients, Irrigation Frequency, Net Irrigation and Gross Irrigation etc.).

Chapter 3

Reference Evapotranspiration ET_O using H2O Framework

This chapter presents H2O model framework to determine the daily reference evapotranspiration ET_o using Hoshiarpur station and Patiala station. The effects of four supervised learning algorithms: Random Forest(RF), Gradient-Boosting Machine (GBM), Deep Learning-Multilayer Perceptrons (DL), and Generalized Linear Model (GLM) models and to predict future ET_o. Analysis of these four models, perform in H2O framework.

3.1 Overview

The accurate calculation of ET_o is necessary to evaluate the requirement of Net Irrigation, regional water resources planning and management, and to model the climate change effect. A more comprehensive decryption can be found in previous literature for the ET_o estimation using sensitivity analysis or empirical based within or India. Similarly, several studies have examined the predictive accuracy of Genetic Programming (GP), and Gene Expression Programming (GEP) in the modeling of ET_o.

3.2 Study Area and Datasets

Punjab is called as "*India's bread basket*", located from 29°32 N to 32°32 N latitude and 73°55 E to 76°50 E longitude) and has 4.20 million ha land under cultivation. The area under Wheat cultivation is 3.4 million hectare (ha) with a production of 14.9 million [178] and the state contributes 60% of the Wheat crop to the central pool. Since the central zone of Punjab is facing problem of declining water table, increased energy cost for pumping, and scanty rainfall, there is a need to manage the available surface water and ground water resources optimally to sustain agriculture.

Table 3.1: Statistical parameters of available meteorological variables and ET_o of Hoshiarpur and Patiala stations

Climate	Dataset	Max		Min		Mean		Std.		SK		K	
		HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL
T_{max} (°C)	All	47.8	50.6	5.1	7.7	30.17	30.2	7.45	7.04	-0.29	-0.29	-0.59	-0.6
	Training	47.8	50.6	5.1	7.7	30.17	30.19	7.47	7.06	-0.31	-0.29	2.43	2.42
	Validation	47.3	45.9	5.4	9.1	30.24	30.19	7.42	6.95	-0.27	-0.27	2.3	2.36
	Testing	47	47	8.7	9.8	30.03	30.26	7.43	7.14	-0.28	-0.3	2.28	2.39
T_{min} (°C)	All	33.5	33.7	0	0.1	16.29	17.47	8.21	7.8	-0.21	-0.22	-1.26	-1.3
	Training	33.5	33.7	0	0.1	16.34	17.49	8.19	7.78	-0.23	-0.22	1.75	1.7
	Validation	32.5	31.7	0.1	0.6	16.21	17.37	8.17	7.81	-0.18	-0.2	1.72	1.69
	Testing	31.7	31.4	0	0.6	16.3	17.61	8.36	7.86	-0.23	-0.25	1.68	1.71
R_H(%)	All	99.1	100.1	11.1	0.1	68.14	76.26	16.52	17.01	-0.72	-0.79	0.12	-0.07
	Training	99.1	100.1	11.1	0.1	68.23	76.34	16.53	16.93	-0.73	-0.78	3.15	2.92
	Validation	99.1	100.1	13.1	19.1	67.94	76.13	16.53	16.98	-0.68	-0.79	3.07	2.93
	Testing	99.1	100.1	14.1	15.1	68.24	76.22	16.5	17.39	-0.73	-0.82	3.12	2.96
u_2 (km h^{-1} day^{-1})	All	32.1	37.1	0	0.1	2.95	4.56	2.28	2.96	2.01	1.17	9.54	2.88
	Training	29.1	37.1	0	0.1	2.91	4.57	2.24	2.99	1.87	1.23	10.61	6.47
	Validation	28.1	28.1	0	0.1	3	4.58	2.29	2.92	1.95	1.11	11.78	5.37
	Testing	32.1	20.1	0	0.1	2.97	4.52	2.4	2.92	2.53	1.03	19.05	4.46
I_s (h)	All	19.4	15.7	0	0	7.75	7.33	3.35	3.14	-0.69	-0.88	0.029	0.01
	Training	19.4	12.9	0	0	7.74	7.33	3.39	3.13	-0.67	-0.88	2.97	3.01
	Validation	15	13.4	0	0	7.74	7.3	3.3	3.15	-0.73	-0.89	3.09	2.99

Climate	Dataset	Max		Min		Mean		Std.		SK		K	
		HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL	HSP	PTL
R_s (MJ m^{-2} day^{-2})	Testing	15	15.7	0	0	7.83	7.38	3.34	3.14	-0.68	-0.88	3.09	3.06
	All	38.9	33.2	4.7	4.9	17.79	17.47	6.14	5.81	0.04	-0.05	-0.68	-0.79
	Training	38.9	29.3	4.7	4.9	17.79	17.48	6.19	5.81	0.04	-0.05	2.3	2.21
	Validation	32.3	30.1	4.7	4.9	17.73	17.39	6.03	5.81	0.03	-0.05	2.33	2.21
	Testing	32.3	33.2	4.7	4.9	17.88	17.57	6.17	5.81	0.06	-0.04	2.29	2.21
ET_o (mm)	All	7.31	6	0.5	0	3.1	2.71	1.55	1.51	0.27	0.18	-0.97	-1.07
	Training	7.31	6	0.5	0	3.11	2.72	1.55	1.51	0.26	0.19	2.02	1.94
	Validation	7.22	6	0.52	0	3.08	2.69	1.53	1.5	0.26	0.19	1.99	1.92
	Testing	7.31	6	0.66	0	3.12	2.73	1.58	1.51	0.29	0.14	2.05	1.91

Over-irrigation can cause to waterlogging, water wastage, nutrient leaching in the soil, and polluting the groundwater resources. To avoid such negative impact, proper policies and measures are needed for the quantification of crop growth as well as water. This section describe the study area and various types of dataset.

3.2.1 Geographical Conditions

Punjab is classified into five Agro-Climatic Zones (ACZ) i.e. Central Plain, Sub-Mountain, Western plain, Undulating Plain, and Western. The climatic data of the district Patiala, Hoshiarpur, Ludhiana, and Amritsar stations have been received from India Meteorological Department (IMD), Pune, India. The Hoshiarpur station is located at 31.5°E latitude and 75.9°S longitude. The Patiala station is located at 30.33°E latitude and 76.38°S longitude. The elevation of the Hoshiarpur and Patiala stations are (296 m and 351 m) above sea level. Ludhiana is centrally located of Punjab at (30° 54' N latitude, 75° 48' E longitude), and an 247m elevation. The location of the climatology station, and its map of the study area as shown in Fig. 3.4. The climate of Hoshiarpur area has a hot, dry summers and very cold in winters. The soil types are calcareous sand, silt, and sandy loam. The Patiala station climate has hot and semi-arid which is mainly dry in summers and winters are on their peak except during monsoon. Also, Patiala soil types are arid-brown and tropical arid-brown. Ludhiana has semi-arid climate with hot summer and mild winter seasons. The proposed model have been developed using three different types of datasets.

- Weather Dataset: The proposed model is developed for the estimation of ET_o using daily weather dataset of Punjab from IMD, Pune. Daily meteorological data during (1970-1999 and 2007-2016) of Thirty-eight years for Patiala station and 1978-1999 and 2007-2016 of Thirty-one years for Hoshiarpur station is considered for study. Six attributes such as Solar Radiation (R_s), Wind Speed (u_2), Temperature maximum (T_{max}) and minimum (T_{min}), Average Relative Humidity (R_H), and Sunshine Hours (I_s) are used for this study.

- Crop Dataset: In this study two crops are considered Wheat and Maize, and its characteristics such as crop type and variety, planting and harvesting period, and total crop growth days with different number of stages are obtained from PAU package of practice for crops and vegetables, Ludhiana [179]. The four stages of crop growing period from date of sowing and their correspondence crop coefficient (K_c) values are taken from PAU, Ludhiana.

- Soil Dataset: The soil pH and electrical conductivity (EC) of the soil are 7.99 and 0.31 DSM values are obtained from the Department of Soil Science, PAU, Ludhiana respectively. Table 3.1 shows the statistical parameters of

meteorological variables at Hoshiarpur and Patiala sites. Table 3.2 and Table 3.3 shows the correlation coefficients of the two dataset respectively.

Table 3.3: Cross-correlation matrix of dataset at Hoshiarpur

	T_{min}	T_{max}	R_H	u_2	I_s	R_s	ET_o
T_{min}	-	**0.84**	-0.18	**0.21**	0.14	0.57	0.81
T_{max}	**0.84**	-	-0.37	0.17	0.31	0.66	0.82
R_H	-0.18	-0.37	-	-0.16	-0.31	-0.41	-0.29
u_2	0.21	0.17	-0.16	-	-0.05	0.14	0.20
I_s	0.14	0.31	-0.31	-0.05	-	0.84	0.58
R_s	0.57	0.66	**-0.41**	0.14	0.84	-	**0.91**
ET_o	0.81	0.82	-0.29	0.20	0.58	**0.91**	-

Table 3.4: Cross-correlation matrix of dataset at Patiala

	T_{min}	T_{max}	R_H	u_2	I_s	R_s	ET_o
T_{min}	-	**0.85**	-0.30	0.06	0.16	0.59	0.81
T_{max}	**0.85**	-	**-0.56**	0.11	0.34	0.68	0.81
R_H	-0.30	-0.56	-	**-0.27**	-0.38	-0.51	-0.42
u_2	0.06	0.11	-0.27	-	0.13	0.20	0.16
I_s	0.16	0.34	-0.38	0.13	-	0.82	0.58
R_s	0.59	0.68	-0.51	0.20	**0.82**	-	**0.91**
ET_o	0.81	0.81	-0.42	0.16	0.58	**0.91**	-

3.3 Methods

3.3.1 FAO-56 Penman-Monteith

In this section, the estimation of ET_o is done by the PM method, which is introduced by the Land and Water Development Division of Food and Agriculture Organization (FAO) of the United Nations.

The FAO-56 Penman-Monteith (PM) is broadly used to estimate ET_o from meteorological factors (Eq 3.1) and it is suggested as the standard technique by FAO [13]. FAO56-PM method is renowned for being the most accurate model and it can be used for hourly or daily time stamps. For hourly time stamps, the equation is as shown [8]:

$$ET_0 = \frac{0.408 \cdot \Delta \cdot (R_n - G) + \gamma \cdot \frac{900}{T_{mean}+273} \cdot u_2 \cdot (e_s - e_a)}{\Delta + \gamma(1 + 0.34u_2)} \tag{3.1}$$

where, (FAO56-PM) method is applied to calculate the reference crop evapotranspiration ET_o (mm/day^{-1}) using Eq. (3.1).

3.3.2 Random Forest (RF)

Random forest (RF) provides the many key advantages over other well-known classifier models, because of it's non-parametric nature. RF technique, is relatively new approach developed from the decision tree that aggregates the predictions from many decision trees on different subsets of data [180]. It combines an ensemble of decision trees to improve prediction accuracy and demonstrate a more robust capacity in terms of incurring the over-fitting problem and resisting noise data. Fig. 3.1 shows the structure of Random Forest.

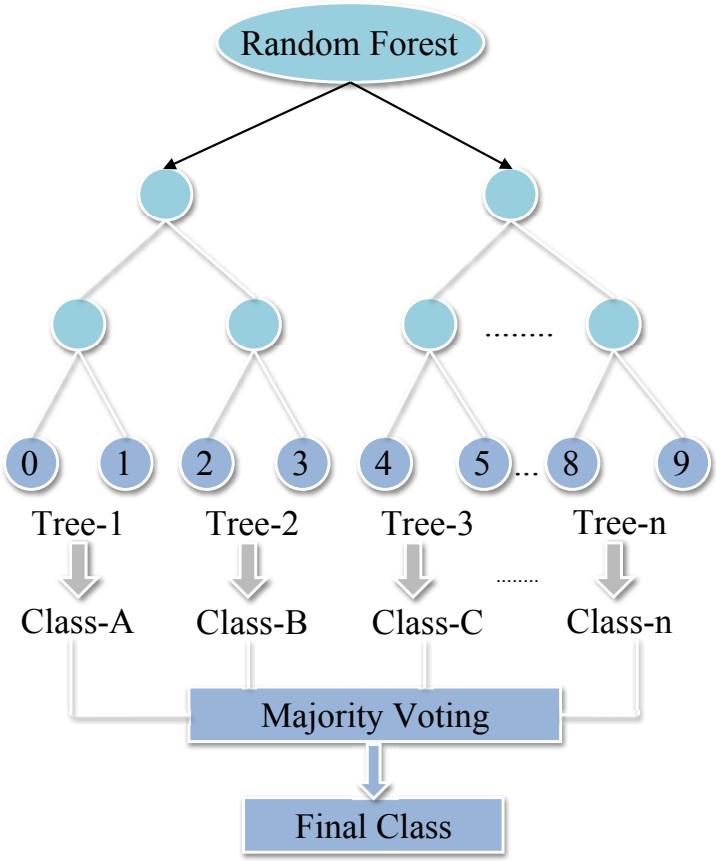

Figure 3.1: Structure of random forest

The study presents a detailed methodological insights to model setup and validation of the underlying machine learning approach with RF. Further, a detailed methodological insights to set up model and validate the underlying machine learning approach with the RF method is presented. RF is an ensemble learning technique, in which the performance of several weak learners is boosted via a voting scheme. It also provides higher classification accuracy than the other classifier models. Moreover, RF is a powerful classification (in the case of a nominal response) and regression (the case of a numeric response) approach, which is currently an active research area and successfully solves problems in many domains. It has been applied in many applications such as remote sensing, image classification, medicine,

neuroscience, bio-informatics etc [181]. The steps of the random forest algorithm are summarized in Algorithm 3.1.

Algorithm 3.1: The steps of random forest algorithm

1. *Bootstrap sampling:* Generate *ntree* bootstrap sample sets from the original training dataset.
2. *Random-forest tree growing:* Grow a tree for each bootstrap for the dataset, at each node, select the best split among a randomly selected subset of input variables (*mtry*), which is the tuning parameter of random forest algorithm.
   (a) Randomly select sample "k" variables from all the input variables "m".
   (b) Pick the best split predictor from those k features, where $k < m$, and m is a number of all input variables.
3. *Ensemble averaging:* Predict new output by averaging the outputs of n classification trees when new inputs are fed into RF. Assemble information from the *ntree* for new dataset prediction such as majority voting for classification.
4. *Out-of-bag error estimation:* Compute an out-of-bag (OOB) error rate by using the data, not in the bootstrap sample.

The number of trees is referred to as *ntree*. For each split in a tree, *mtry* randomly drawn predictors are assessed as candidates for splitting and the predictor that yields the best split is chosen. The default values for *mtry* are $\sqrt{p}$ for classification and $p/3$ for regression. For regression models, the prediction error is returned as a mean squared error (MSE).

3.3.3 Gradient Boosting Machine (GBM)

Gradient Boosting Machine (GBM), is introduced by Friedman [182]. The idea of gradient boosting originated in the observation by Breiman [183] and stated boosting can be interpreted as an optimization algorithm on a suitable cost function. GBM is a machine learning technique that combines two powerful tools: Gradient-based optimization and Boosting. Gradient based-optimization uses gradient computations to minimize a model's loss function with respect to training data and it also used for predictive results for regression or classification. It is also an ensemble learning method and provides high-prediction accuracy and often outperforms many competing methods, such as linear regression/classification, bagging [180].

GBM can be used in multiple predictive fashions such as regression or tree based classification. In GBM [184], the learning process constantly fits new models to yield a more accurate estimate of the response variable. The main perception behind this algorithm is to develop the new base-learners to be maximally correlated with the

negative gradient of the loss function and combine with the whole ensemble. The five parameters are used to build the GBM model such as the number of trees, the depth of the tree, the learning rate, the sample rate, and distribution. Moreover, other tuning parameters are defined at their given default values in GBM booklet [184]. GBM for classification is defined in Algorithm 3.2.

Algorithm 3.2: The steps of gradient boosting machine algorithm

1. *Initialization:*
   Initialize $f_{k0} = 0$, k = 1, 2, 3, ..., K
2. *Calculate Error loss:*
   For m = 1 to M
   (a) *Set* $= p_k(x) = \frac{e^{f_k(x)}}{\sum_{l=1}^{K} e^{f_l(x)}}$ *for all* $k = 1, 2, 3, ...K$
   (b) *For* $k = 1$ *to* K
      i. $Computer_{ikm} = y_{ik} - p_k(x_i), i = 1, 2, ..., N$
      ii. Fit a regression tree to the targets $r_{ikm}, i = 1, 2, ..., N$, giving terminal regions $R_{jkm}, i = 1, 2, ..., J_m$
      iii. *Compute* $y_{jkm} = \frac{K-1}{K} \frac{\sum_{x_i \epsilon R_{jkm}} (r_{ikm})}{\sum_{x_i \epsilon R_{jkm}} |r_{ikm}|(1-|r_{ikm}|)}, j = 1, 2, ..., J_m$
      iv. Update $f_{km}(x) = f_{k,m-1}(x) + \sum_{j=1}^{J_m} y_{jkm} I(x \epsilon R_{jkm})$
   (c) END
3. Output $f_k(x) = f_{kM}(x), k = 1, 2, 3, .., K$

   END

The above algorithm considers multi-class classification k-regression trees, where one tree represents each target class. The index m tracks the number of weak learners to be added to the current ensemble. Within this outer loop, there is an inner loop across each of the K classes. The first line initializes optimal constant model, which is just a single terminal node. Within this inner loop is to compute the residuals, r_{ikm}, which are the actually gradient values for each N bins in the CART model. A regression tree is then becomes fit to these gradient computations. This fitting process is distributed and paralleled. Further, inner loop is to add the current model to the fitted regression tree to improve the accuracy of the model during the inherent gradient descent step. After M iterations, the final "boosted" model can be tested out on new data.

3.3.4 Generalized Linear Model (GLM)

Generalized linear model (GLM) extend the concept of the well understood linear regression model which is widely used in field of hydrology and meteorology [185]

[186], using a machine learning technique. An application of GLMs for simulation of daily sequences of potential evaporation for hydrological applications model is given [187]. The theory of generalized linear model (GLM's) has been reviewed thoroughly [188].

GLM model can be fitted by solving the maximum likelihood optimization problem. In this chapter, GLM is implemented and fitted with elastic-net penalties [189]. The elastic net penalty is used for parameter regularization. The model fitting computation is distributed, extremely fast and scaled well for models with a limited number ($\sim$ low thousands) of predictors with a non-zero coefficient.

GLM is the model for solving the following likelihood optimization with parameter regularization:

$$max_{\beta,\beta_0}(GLMLog - likelihood–RegularizationPenalty). \tag{3.2}$$

where β is the maximum likelihood, β_0 represents the vector coefficient and y is the independent observations vector. The Eq. (3.2) presents the algorithm for the elastic-net, which include the least absolute shrinkage and selection operator (LASSO) and Ridge regression functionalities. GLM model penalties have been used to avoid over fitting, deal with corrected predictors and reduced the variance of the prediction error. The Ridge regression and LASSO are the popular penalized models that provide a feasibility and an elastic-net which combines both penalties for predicting the model. These penalized parameters are known as *L1* and *L2* penalties and these are parameterized by the α and λ arguments presented [182]. Elastic net regularization penalty is the weighted sum of the *L1* and *L2* norms of the coefficients vector.

By using this approach, α handles the elastic net penalty distribution between the *L1* and *L2* norms. However, it can have any value in the [0; 1] range or a vector of values (which triggers grid search). If $\alpha = 0$, it determines the GLM model using ridge regression. If $\alpha = 1$, the LASSO penalty is used. In case of λ, it handles the penalty strength in which the range is computed as a positive value or a vector of values. However it, models the likelihood of an observation belonging to an output category from given data (for instance $Pr(y = c|x)$) and refer as Eq. (3.2). The likelihoods are demonstrated as follows:

$$\hat{y}c = Pr(y = c|x) = \frac{e^{x^T\beta_c+\beta_{c0}}}{\sum_{k=1}^{K}(e^{x^T\beta_k+\beta_{k0}})} \tag{3.3}$$

When the output response variable y has K $>$ 2 levels, the Eq (3.3) and Eq (3.4) can be generalized to a multinomial model. The multinomial model extends the binomial when the number of classes is more than 2.

The regularized multinomial classification probabilities with elastic net are de-

fined as follow:

$$\begin{aligned} -[\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{i,k}(x_i^T\beta_k+\beta_{k0}))-log(\sum_{k=1}^{K}e^{x_i^T\beta_k+\beta_{k0}})] \\ +\lambda[\frac{(1-\alpha)}{2}||\beta||_F^2+\alpha\sum_{j=1}^{P}||\beta_j||_1] \end{aligned} \tag{3.4}$$

,where the Eq (3.4) presents the GLM parameter β is determined by maximum likelihood estimation and is a p X K matrix of coefficients, β_k refers to the k^{th} column (outcome category), β_j the j^{th} row (vector of K coefficients for variable j). β_c is a vector of coefficients for class c, j denotes the penalty factor for the j^{th} variable and $y_{i,k}$ is the k^{th} element of the binary vector developed by extending the response variable using one-hot encoding (i.e. $y_{i,k} == 1$ if the response at the i^{th} observation is k; otherwise is it 0.)

The regularization parameters α and λ are tune in order to find the optimal one and it provides a grid search over α and a certain form of grid search called "lambda search "over λ.

However, the value of parameter $\alpha = 0.5$, $\lambda =$ 1.0E-5 is considered. Lambda search enables efficient and automatic search for the optimal value of the lambda parameter.

3.3.5 Deep Learning (DL)

Deep feed-forward neural networks are based on Multi-layer Perceptrons (MLPs) [190]. A Neural network (NN), or Artificial Neural Network (ANN), is an information processing system that consists of a directed graph with activation functions associated to each node. DL is based on high-level artificial neural networks whose parameters are optimized via stochastic gradient decent using back-propagation. A multi-layer DL model has been implemented by considering a number of multiple hidden layers and rectified linear active function. Thus, our model is trained with stochastic gradient descent using back-propagation. Fig. 3.2 shows the structure of the Deep Learning model. The DL structure, it consists input layer, first hidden layer h_j, the second hidden layer h_k, the third hidden layer h_L and final output layer.

Deep learning approach supports many advanced features for learning, optimization and to overcome the over-fitting of the model. [191]. A perceptron's consist of a vector of weights w=[w_1, w_2....w_n], one for each input x_i= [x_1, x_2....x_n], a distinguished weight b, called bias, so the weighted sum Y(net) can be computed by Eq. (3.5).

$$Y(net) = \sum_{i=1}^{n} w_i * x_i + b \tag{3.5}$$

In addition, the activation function is applied to the input variable (x_i) which is then multiplied with the (w) and added to the (b) function.

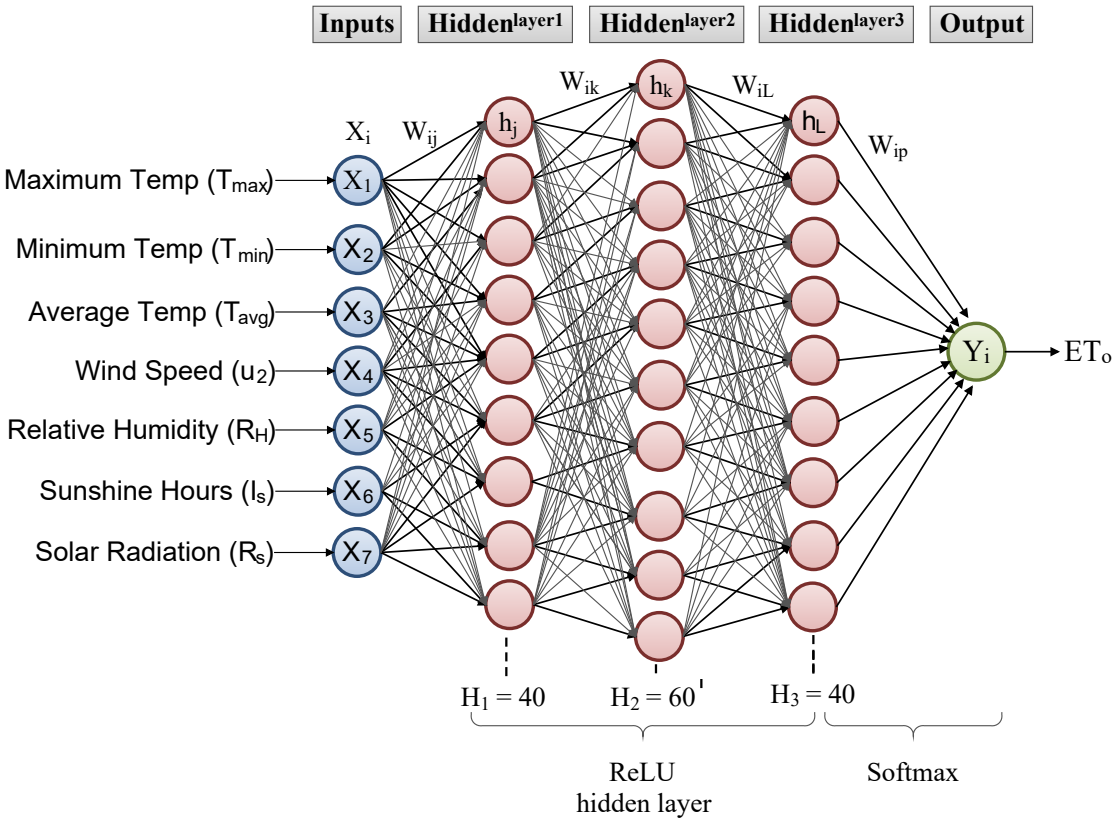

Figure 3.2: Architecture of deep learning

The weighted combination of input signals is aggregated, and then an output signal is transmitted by the connected neuron. The function $f(x)$ represents the non-linear activation function used throughout the network, and the bias b accounts for the neuron's activation threshold. The weights of each node are learned through the process of forward propagation and backward propagation changes the weights based on a computed output and update the weights to minimize the error.

The Rectified Linear Unit (ReLU) function has demonstrated high performance on image recognition tasks. It is more biologically accurate and computationally cheaper. The ReLU is expressed by Eq. (3.6):

$$f(x) = max(0, x) \tag{3.6}$$

The derivative of ReLU is:

$$f'(x) = \left\{ \begin{array}{ll} 1, & \text{if } x > 0 \\ 0, & \text{otherwise} \end{array} \right\}$$

The multinomial classification for calculating loss function is defined as Cross Entropy and is calculated by Eq. (3.7):

$$L(W, B|j) = -\sum_{y\epsilon O}^{n}(ln(O_y^{(j)}) * (t_y^{(j)}) + ln(1 - O_y^{(j)}) * (1 - t_y^{(j)})) \tag{3.7}$$

A three-layer deep learning model algorithm steps are summarized as follows:

From input layer to first hidden layer $f(h_j)$ with relu activation, which can be expressed by Eq. (3.8)

$$f(h_j) = ReLU(\sum_i (x_i * w_{ij}) + b_j) \tag{3.8}$$

From first hidden layer $f(h_j)$ to second hidden layer $f(h_k)$, which can be expressed by Eq. (3.9)

$$f(h_k) = ReLU(\sum_j (h_j * w_{ik}) + b_k) \tag{3.9}$$

From second hidden layer $f(h_k)$ to third hidden layer $f(h_L)$, which can be expressed by Eq. (3.10)

$$f(h_L) = ReLU(\sum_k (h_k * w_{iL}) + b_L) \tag{3.10}$$

From third hidden layer $f(h_L)$ to output soft-max layer pL, which can be expressed by Eq. (3.11)

$$pL = softmax(\sum_L (h_L * w_{ip}) + b_p) \tag{3.11}$$

Soft-max activation function is a type of sigmoid function, which can handle several classes in the classification problem. The output of soft-max function is equivalent to a probability distribution.

$$Y = softmax(x_i = \frac{exp^{(x_i)}}{\sum_{k=1}^{k}(exp^{(x_i)})}) \tag{3.12}$$

,where x_i represents i th element of the input to soft-max, which correspond to class i and k that are the number of classes.

The output layer error between the target and the observed output is calculated by Eq. (3.13) to Eq. (3.17). Here, the equations show that stochastic gradient descent has been applied to update the weights and bias via back-propagation for minimizing the prediction error. Standard (SGD) can be summarized as follows, with the gradient $\bigtriangledown\ L\ (W, B|j)$ computed via back-propagation and the partial derivatives.

$$w_{jk} = w_{jk} - \alpha \frac{\delta L(W, B|j)}{\delta w_{jk}} \tag{3.13}$$

$$b_{jk} = b_{jk} - \alpha \frac{\delta L(W, B|j)}{\delta b_{jk}} \tag{3.14}$$

$$net = \sum_{i} (w_{jk} * x_i + b_{jk}) \tag{3.15}$$

$$Y = activation(net) \tag{3.16}$$

$$E = error(Y) \tag{3.17}$$

where, updated weights $w_{jk} \in W$, biases $b_{jk} \in B$, α represents a learning rate to compute the parameters W and b with iterations which is expressed by Eq. (3.13) and (3.14). Y = output of layers or neurons; x_i = input layer i to hidden neurons; w_i, w_k, w_L, w_p = weight connection among input and a hidden layer or neurons from input i to layer p; b_j, b_k, b_L, b_p = bias of hidden neurons of each hidden layer.

3.4 Proposed Model

3.4.1 H2O Framework

H2O provide a Java-based platform for more scalable, fast and open source machine learning, recently applied in more than 2000 corporations [191] [189] [184]. H2O framework is focused on scalable machine learning as the API for big data applications. It is an open source platform which mainly focuses on scalable machine learning and interactive data science.

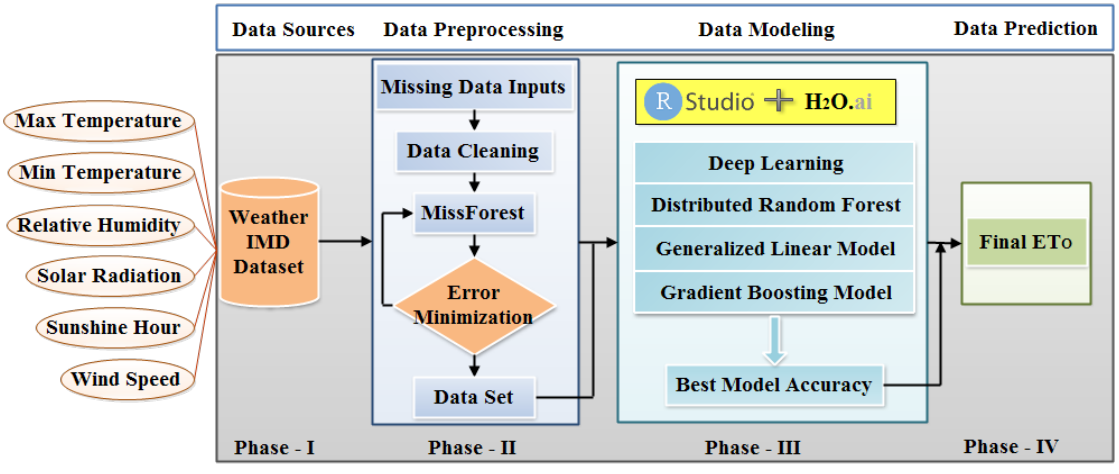

Figure 3.3: Reference evapotranspiration ET_o estimation framework

H2O uses in-memory Map-Reduce paradigm to distribute work. One of its advantages is the clever distribution of data chunks. When H2O imports a file, it does so in distributed fashion, i.e., each node tries to load it in parallel (number of cores, nthreads). When the data file is parsed, it gets split into several data chunks. H2O tries to have more data chunks than CPU cores. Since it uses Map-Reduce paradigm, it is often useful to look in different chunks in map phase than the one map worker is given. H2O assumes that such lookups are more likely to data chunks that are nearer to the given chunk. Due to this reason, H2O tries to keep these chunks in the same node. Another benefit of using H2O is the simplicity of deployment.

The proposed H2O framework model consist of four phases to collect, process and analyze the dataset. Fig 3.3 represents the proposed H2O framework for prediction the ET_o of two stations. Our methodology consists of four steps. It shows the four phases of the framework (i) Data sources, (ii) Data Pre-processing, (iii) Data Modeling, and (iv) Data Prediction.

Data Sources:

Data collection is same as the case study presented in section 3.2.1. In this section, we briefly discuss about the study area and climate datasets as shown in Fig 3.4.

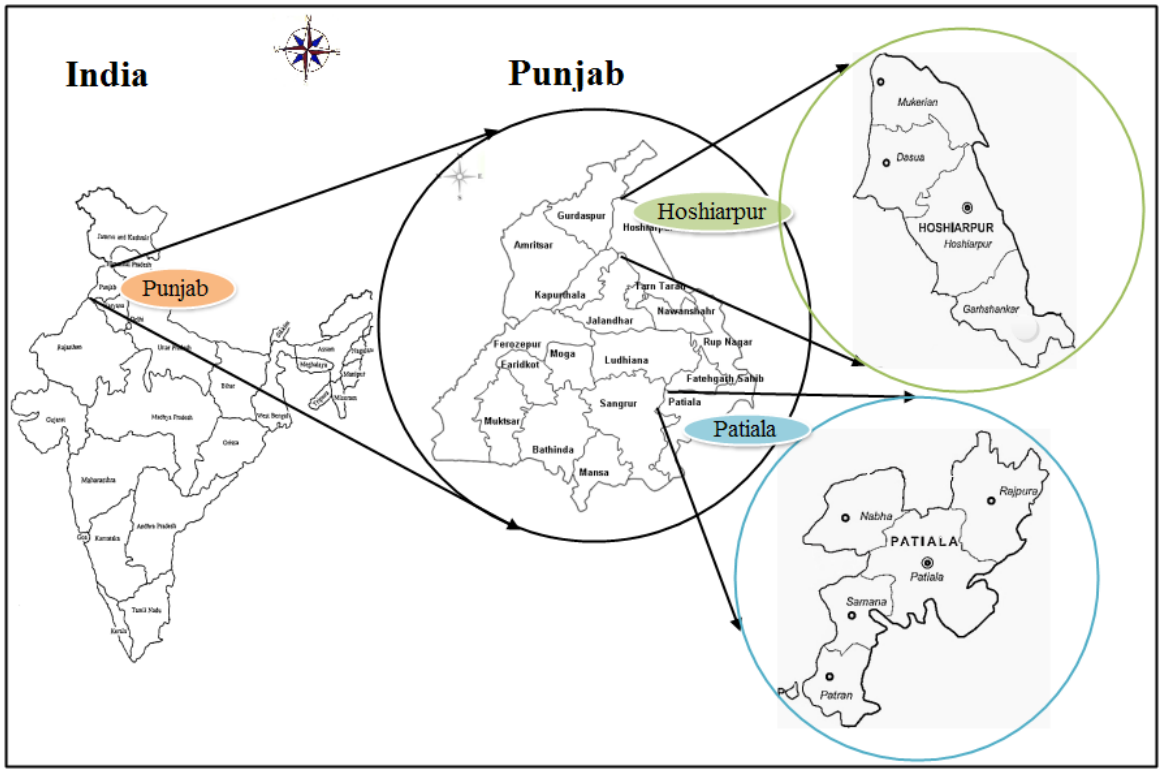

Figure 3.4: Study area's in Hoshiarpur and Patiala

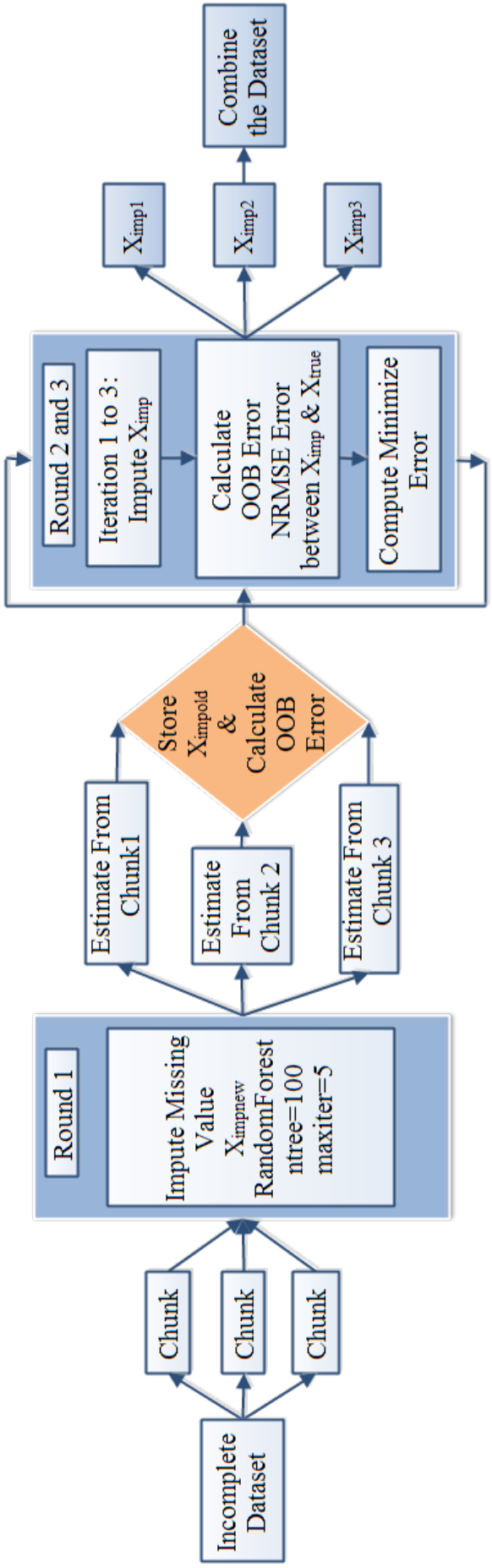

Figure 3.5: Missing data treatment using missforest algorithm

Data Pre-processing:

In this section, the data is processed using imputation method, such as MissForest package. The missForest package is a non-parametric imputation method. Missing data treatment (MDT) has been generally mentioned in the data-driven studies of psychology, social science, transportation, biology, and behavioral science [192]. However, the missing values in datasets have negative impacts on the estimation of accurate predictive accuracy and therefore, could lead to inconsistent results.

MissForest is a random forest-based method to impute phenomic data [193]. In this study, MissForest calculation is utilized to estimate the missing information as shown in Fig. 3.5. This process is carried out in three rounds.

Round 1.	esoph.imp2 ← missForest(esoph.mis1, verbose= TRUE, maxiter = 5, $ntree = 100$)
Round 2.	esoph.imp3 ← missForest(esoph.mis1, xtrue = esoph.imp2$ximp, verbose = TRUE)
Round 3.	esoph.imp4 ← missForest(esoph.mis1, xtrue = esoph.imp3$ximp, verbose = TRUE)

In the first process of iteration, the imputation of missing values have been divided by the data into three chunks, i.e. $chunk_1$, $chunk_2$, and $chunk_3$. For each Sample X dataset missing values are imputed by first fitting a random forest with response Y_{obs} and predictors x_{obs}; Secondly, predicting the missing values y_{mis} by applying the trained random forest to x_{mis}. MissForest uses an iterative imputation scheme by training a RF on observed values for each variable, followed by predicting the missing values, and then proceeding iteratively until the stopping criteria is met. Besides, it can be run parallel to save computation time and the OOB (out-of-bag) imputation error to evaluate the continuous and categorical parts of the imputed datasets. OOB is a method for calculating prediction error in random forest approach.

The performance of this evaluation is observed by comparing the absolute difference between true imputation error (error_{true}) and OOB imputation error (error_{OOB}) in all simulation iterations. The method is repeated until the imputed values reach convergence.The difference for the set of continuous variables N in the data set is computed by Eq (3.18):

$$\triangle_N = \frac{\sum_{j \epsilon N}(X^{imp}_{new} - X^{imp}_{old})^2}{\sum_{j \epsilon N}(X^{imp}_{new})^2} \tag{3.18}$$

The normalized root mean squared error (NRMSE) is calculated by Eq. (3.19):

$$NRMSE = \sqrt{\frac{mean((X_{true} - X_{imp})^2)}{var(X_{true})}} \tag{3.19}$$

where X_{true} − is the total data matrix, X_{imp} − is the imputed data matrix and

mean/var − are empirical mean and variance computed over the continuous missing values.

In this section, the algorithm demonstrated that how missing data imputed in the dataset.The algorithm steps are summarized in Algorithm 3.3. The process of MissForest to impute the missing values is as follows:

Algorithm 3.3: The steps of missforest algorithm

1. X $\longrightarrow$ n $\times$ p matrix, divide X sample into chunks $X_{1,2,3...N}$.
2. $X_{imp} \longleftarrow$ Calculate the number of variables that have missing values and Impute the values.
   (i) for X in Chunk = 1,2,3, do and Iteration = Round 1 to Round 3,
   (ii) Store the previously imputed values in X_{old}^{imp}
      (a) Fit a random forest: $Y_{obs}^{(c)} \sim X_{obs}^{(c)}$,
      (b) Predict Y_{mis} using X_{mis}
      (c) Update the new imputed values in X_{new}^{imp}
      (d) end for
3. Compute Error between OOBerror and NRMSE error X_{true}(Complete data) and X_{imp} (Imputed data).
4. Return the imputed matrix X_{imp}
5. Combine the Dataset of Chunks

Where, X_s is observed values denoted by Y_{obs} and missing values denoted by $Y_{mis}X_{old}^{imp}$ and X_{new}^{imp} denotes new and previously imputed data matrix using predicted Y_{mis}. Error results obtained by MissForest for Hoshiarpur and Patiala is shown in Table 3.4.

Table 3.5: Estimated error results of two dataset using missforest

NRMSE	Chunk1		Chunk2		Chunk3	
	HSP	PTL	HSP	PTL	HSP	PTL
1. OOBE	0.1879	0.1577	0.1520	0.1357	0.1743	0.1533
2. OOBE	0.1874	0.1567	0.1495	0.1368	0.1734	0.1524
Error	0.0656	0.0387	0.0686	0.0416	0.0823	0.0459
3. OOBE	0.1872	0.1565	0.1517	0.1362	0.1737	0.1524
Error	0.0470	0.0413	0.0605	0.0591	0.0402	0.0444

During simulation, it is observed that MissForest performed the best in each chunk of dataset. The imputation procedure is repeated until a stopping criterion is defined. Furthermore, data division is an essential phase in the modeling process. Various researchers divided the data into training and testing dataset, based on domain knowledge or arbitrarily.

In this study, the dataset is split into three subsets (Training, Validation, and Testing) datasets. The training set consists of 55%, validation set consists of 30% and testing set consists of 15% of the original dataset. These divisions are selected based on several iterations. Then, the optimal values have been selected to avoid the over-fitting problem. The procedure of H2O framework shown in Fig. 3.6.

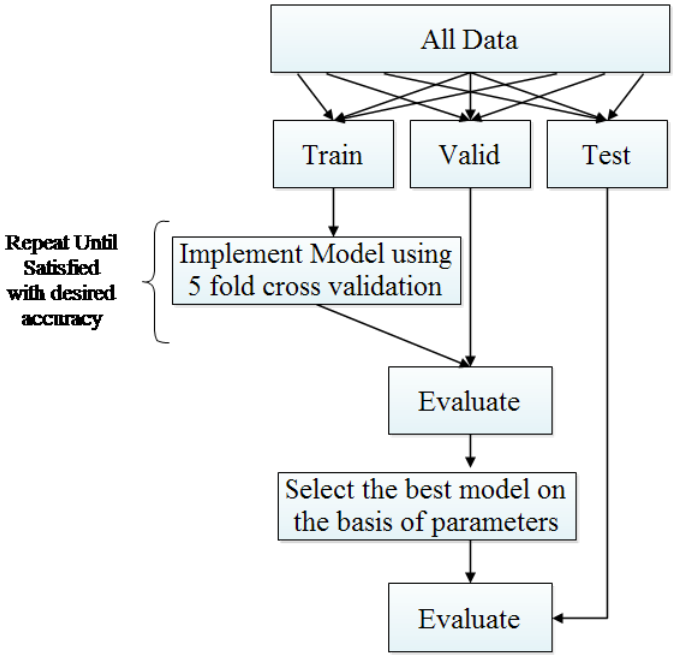

Figure 3.6: The procedure of H2O model

Data Modeling and Prediction

In this stage, the DLs, GLMs, GBMs, and RFs models are trained by using training dataset. After receiving the best accuracy from the training dataset, validation dataset is applied to evaluate the model accuracy. After achieving best accuracy from training and validation datasets, the model is applied to the testing dataset. In the fourth step, the prediction values are evaluated. The flowchart of proposed ET_o estimation model is shown in Fig. 3.7.

3.5 Simulation

Based on the simulation parameter settings described in section 3.4, this section presents the prediction performance of the applied models. This study has been conducted with H2O software and R-studio [194]. Four data-driven methods(DL, GLM, GBM and RF) have been developed in this study to simulate and predict the daily ET_o for Hoshiarpur and Patiala stations. The essential parameters used for considered models are illustrated in Table 3.5.

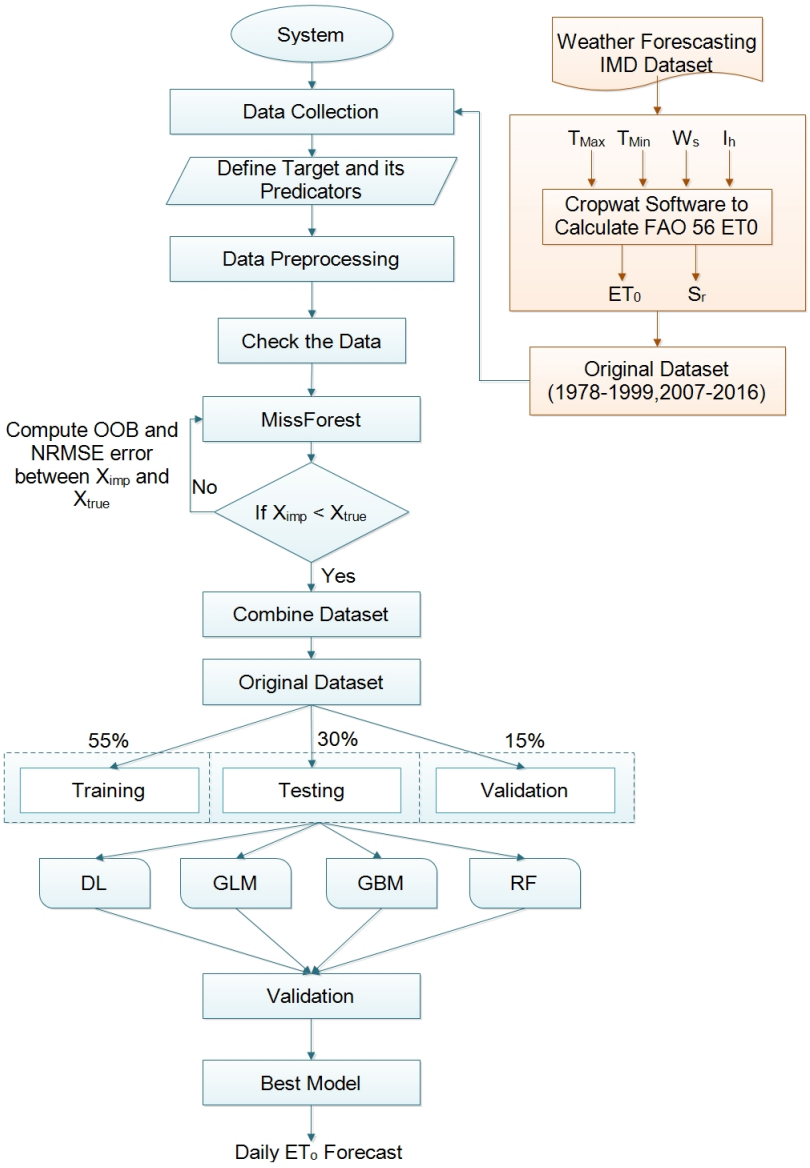

Figure 3.7: Flowchart of ET_o estimation based on H2O framework

Table 3.6: Model Hyper-parameters for ET_o model

Model-Name	Parameters
DL	hidden layer= (40,60,40), actfunction= Rectifier, epoches= 100, L1= 1e-5, Distribution= multinomial
GLM	lambda= TRUE, solver= IRLSM, and nlambdas = 100
GBM	$ntree$ = 200, max_depth = 30, sample_rate = 0.2 and learning_rate = 0.1
RF	$ntree$ = 100, max_depth = 50, mtries = 5 and score_each_interval = TRUE, distribution= multinomial

3.5.1 Evaluation of Model Parameters

In this section, to evaluate the performance of models, the following statistical indicators have been selected using Mean Square Error (MSE), Root Mean Square Error

(RMSE), Normalized Root Mean Square Error (NRMSE), Root Mean Square Logarithmic Error (RMSLE), Pearson Correlation (R), Coefficient of Determination(R^2), Logloss(LL), Nash-Sutcliffe Efficiency (NSE), Accuracy (ACC) and Mean Per-Class Error (MCE). Here, R^2 and ACC present with largest values and RMSE, MSE, NRMSE, and LL with lowest values in terms of mean higher model efficiency.

(i) MSE: The Mean Squared Error of an estimator observed for estimating actual is and if MSE observed is smaller, then it is a better estimator of actual and calculated by Eq. (3.20).

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(ET^i_{o,PM} - \widehat{ET^i}_{o,M})^2 \tag{3.20}$$

(ii) RMSE: The Root Mean Squared Error determines the precision of the model by comparing the deviation between the predicted and real data. The RMSE has always a positive value and is calculated by Eq. (3.21).

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(ET^i_{o,PM} - \widehat{ET^i}_{o,M})^2} \tag{3.21}$$

(iii) NRMSE: The Normalized Root Mean Squared Error, normalizing the RMSD facilitates the comparison between datasets or models with different scales.For RMSE we can find these kinds of normalization and is calculated by Eq. (3.22).

$$\begin{aligned} NRMSE &= \sqrt{\frac{\frac{1}{N}\sum_{i=1}^{N}(ET^i_{o,PM} - \widehat{ET^i}_{o,M})^2}{nval}} \\ nval &= \begin{cases} sd(O_i), norm = \text{"standard deviation"} \\ O_{max} - O_{min}, norm = "max\ min" \end{cases} \end{aligned} \tag{3.22}$$

(iv) RMSLE: The Root Mean Square Logarithmic error computes the root mean squared log error between two numeric vectors and is calculated by Eq. (3.23).

$$\begin{aligned} RMSLE = &\sqrt{\frac{1}{N}\sum_{i=1}^{N}(log(ET^i_{o,PM} + 1)} \\ &\times \sqrt{-log(\widehat{ET^i}_{o,M} + 1))^2} \end{aligned} \tag{3.23}$$

(v) R : The R represents the Pearson correlation between actual and predicted

value and is calculated by Eq. (3.24)

$$(r) = \frac{\sum_1^N (ET^i_{o,PM} - \overline{ET^i_{o,PM}})(\widehat{ET^i}_{o,M}) - \overline{\widehat{ET^i}_{o,M}})}{\sqrt{\sum_1^N (ET^i_{o,PM} - \overline{ET^i_{o,PM}})^2 \sum_1^N (\widehat{ET^i}_{o,M} - \overline{\widehat{ET^i}_{o,M}})^2}} \tag{3.24}$$

(vi) R^2: The coefficient of determination is computed by taking the square of R (Correlation) and calculate by Eq. (3.25).

$$r^2 = r * r \tag{3.25}$$

(vii) LL: Log-loss measures the performance of a classification model where the prediction input is a probability value between 0 and 1. It is calculated by Eq. (3.26).

$$LL = -\frac{1}{N} \sum_i^N \sum_j^P ET^i_{o,PM_{ij}} log(\widehat{ET^i}_{o,M}) \tag{3.26}$$

(viii) NSE: The Nash-Sutcliffe Efficiency is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information") and calculated by Eq. (3.27).

$$NSE = 1 - \frac{\sum_{i=1}^N (ET^i_{o,PM} - \widehat{ET^i}_{o,M})^2}{\sum_{i=1}^N (ET^i_{o,PM} - ET^i_{mean})^2} \tag{3.27}$$

(ix) Accuracy: Number of classes correctly identified as either truly positive or truly negative out of the total number of classes and defined as Eq. (3.28)

$$ACC = \frac{TP + TN}{N} \times 100 \tag{3.28}$$

(x) RAE: Relative Absolute Error, is the residual sum of square, where actual bar is a actual^{mean}, that the total absolute error and normalizes it by dividing it by the total absolute error as Eq. (3.29)

$$RAE = \frac{\sum_{i=1}^N |\widehat{ET^i}_{o,M} - ET^i_{o,PM}|}{\sum_{i=1}^N |\overline{ET^i_{o,PM}} - ET^i_{o,PM}|} \tag{3.29}$$

where, $ET^i_{o,PM}$ is observed/actual values, $\overline{ET^i_{o,PM}}$ is actual^{mean}, $\widehat{ET^i}_{o,M}$ is simulated or predicted value and N as total number of data points, TP and TN is True Positive and True Negative, FP and FN is False Positive and False Negative.

Table 3.7: Performance comparison of selected models for Hoshiarpur

Models	Training					Validation				Testing			
	MSE	RMSE	LL	ACC	Time	MSE	RMSE	LL	ACC	MSE	RMSE	LL	ACC
DL	0.1215	0.3486	0.508	0.85	1.74	0.0382	0.1956	0.140	**0.94**	0.0369	0.1921	0.126	**0.95**
RF	0.0702	0.2649	0.309	0.908	4.02	0.0645	0.254	0.232	0.914	0.0662	0.257	0.276	0.917
GBM	0.0009	0.0310	0.007	**0.99**	2.01	0.0557	0.2362	0.213	0.927	0.0522	0.2285	0.218	0.932
GLM	0.0645	0.253	0.237	0.932	5.60	0.0612	0.2475	0.226	0.934	0.0641	0.253	0.235	0.934

Table 3.8: Performance comparison of selected models for Patiala

Models	Training					Validation				Testing			
	MSE	RMSE	LL	ACC	Time	MSE	RMSE	LL	ACC	MSE	RMSE	LL	ACC
DL	0.0724	0.2691	0.2718	0.904	1.01	0.0477	0.218	0.1911	**0.932**	0.0442	0.210	0.1622	**0.940**
RF	0.065	0.255	0.4921	0.909	0.61	0.063	0.252	0.3641	0.914	0.0605	0.245	0.2864	0.917
GBM	0.0042	0.064	0.02	**0.995**	3.49	0.0539	0.232	0.2152	0.932	0.0502	0.224	0.1951	0.934
GLM	0.0594	0.243	0.2193	0.934	7.90	0.063	0.252	0.2301	0.924	0.059	0.243	0.2158	0.927

3.5.2 Results and Discussion

The results of Hoshiarpur and Patiala stations are categorized into two sections. The first result section describes the Hoshiarpur station results through tables and figures.

3.5.3 Estimation of ET_o using Heuristic Models at Hoshiarpur

The performance indices, including MSE, RMSE, LL, ACC are conducted for determining the high-accuracy of deployed models of daily ET_o prediction during the training, validation and testing periods in Table 3.6. The values of RMSLE, R, R^2, NRMSE, NSE, and RAE during the training, validation and testing datasets for DL, RF, GLM and GBM models are summarized in Table 3.8 for Hoshiarpur site.

Table 3.9: Performance comparison of DL, RF, GLM and GBM models for Hoshiarpur

	Dataset	DL	RF	GLM	GBM
RMSLE	Training	0.0853	0.1150	0.1112	0.0128
	Validation	0.0764	0.1069	0.1123	0.0972
	Testing	**0.0693**	0.09529	0.1023	0.0894
R	Training	0.97	0.96	0.96	0.99
	Validation	0.98	0.96	0.96	0.98
	Testing	**0.98**	0.96	0.96	0.98
R^2	Training	0.97	0.96	0.97	0.99
	Validation	0.98	0.95	0.97	0.97
	Testing	**0.99**	0.96	0.97	0.97
NSE	Training	0.97	0.97	0.97	0.99
	Validation	0.98	0.96	0.96	0.97
	Testing	**0.98**	0.96	0.97	0.97
RAE	Training	0.0468	0.0602	0.0602	0.00107
	Validation	0.0376	0.0662	0.0644	0.0540
	Testing	**0.0346**	0.0621	0.0567	0.0508
NRMSE	Training	16.10%	18.20%	18.20%	2.40%
	Validation	14.40%	20.30%	18.90%	17.30%
	Testing	**13.90**%	18.70%	17.80%	16.80%

As evident from Table 3.8, the Deep learning (DL) provided the best accuracy results (MSE= 0.0369, RMSE= 0.1921, LL= 0.126, ACC= 0.95, MCE= 0.04187). Although, it presents global average performance of heuristic models DL, RF, GLM, and GBM during the testing period. It can be seen that DL model provided the most accurate results among the considered models with average (RMSLE= 0.0693, R= 0.98, R^2= 0.99, NSE= 0.98, NRMSE= 13.90 %, RAE= 0.0346) for Hoshiarpur station. GBM model is a slightly better than GLM model with average (MSE= 0.0522, RMSE= 0.2285, LL= 0.218, ACC= 0.932, MCE= 0.0842 and RMSLE= 0.0894, R= 0.98, R^2= 0.97, NSE= 0.97, NRMSE= 16.80%, RAE= 0.0508), but GBM gives over-fitting in case of training dataset.

Generally, GLM had slightly better accuracy than RF with (MSE= 0.0641, RMSE= 0.253, LL= 0.235, ACC= 0.93, MCE= 0.0848) and (R = 0.96, R^2= 0.97, NSE= 0.97, NRMSE= 17.80% and RAE=0.0567).

The results obtained by four models for Hoshiarpur Station are shown in Fig 3.8, Fig. 3.9 and Table 3.8. Violin plots represents the obtained results for predicted and actual ET_o with training, validation and testing dataset as shown in Fig 3.8. This plot presents density of data distribution and a box plot (median, inter-quartile range, and adjacent values) plot. Violin plots presents the comparison of distributions between multiple group of models. Fig. 3.12 presented predicted and actual estimated ET_o of the models during the testing period for Hoshiarpur stations.

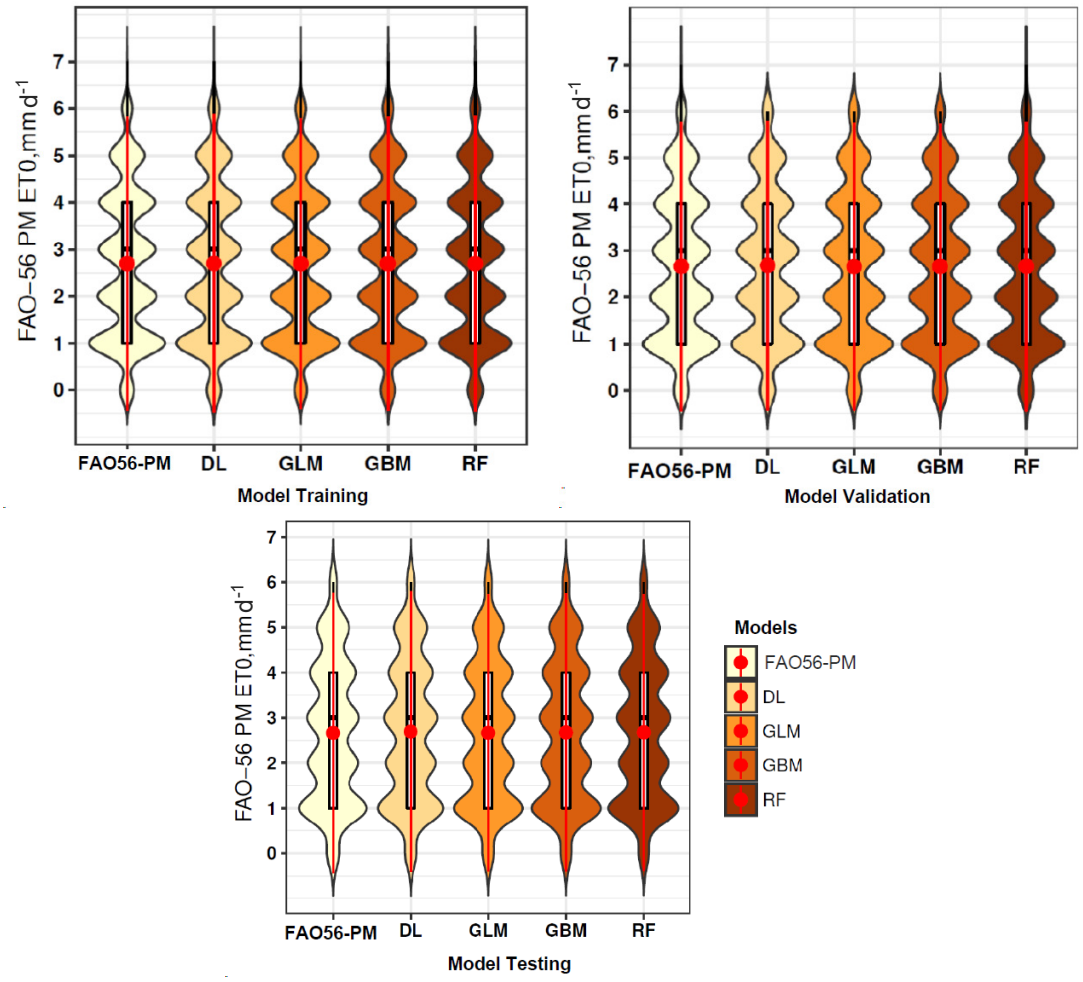

Figure 3.8: Estimated ET_o for actual and predicted results of Hoshiarpur

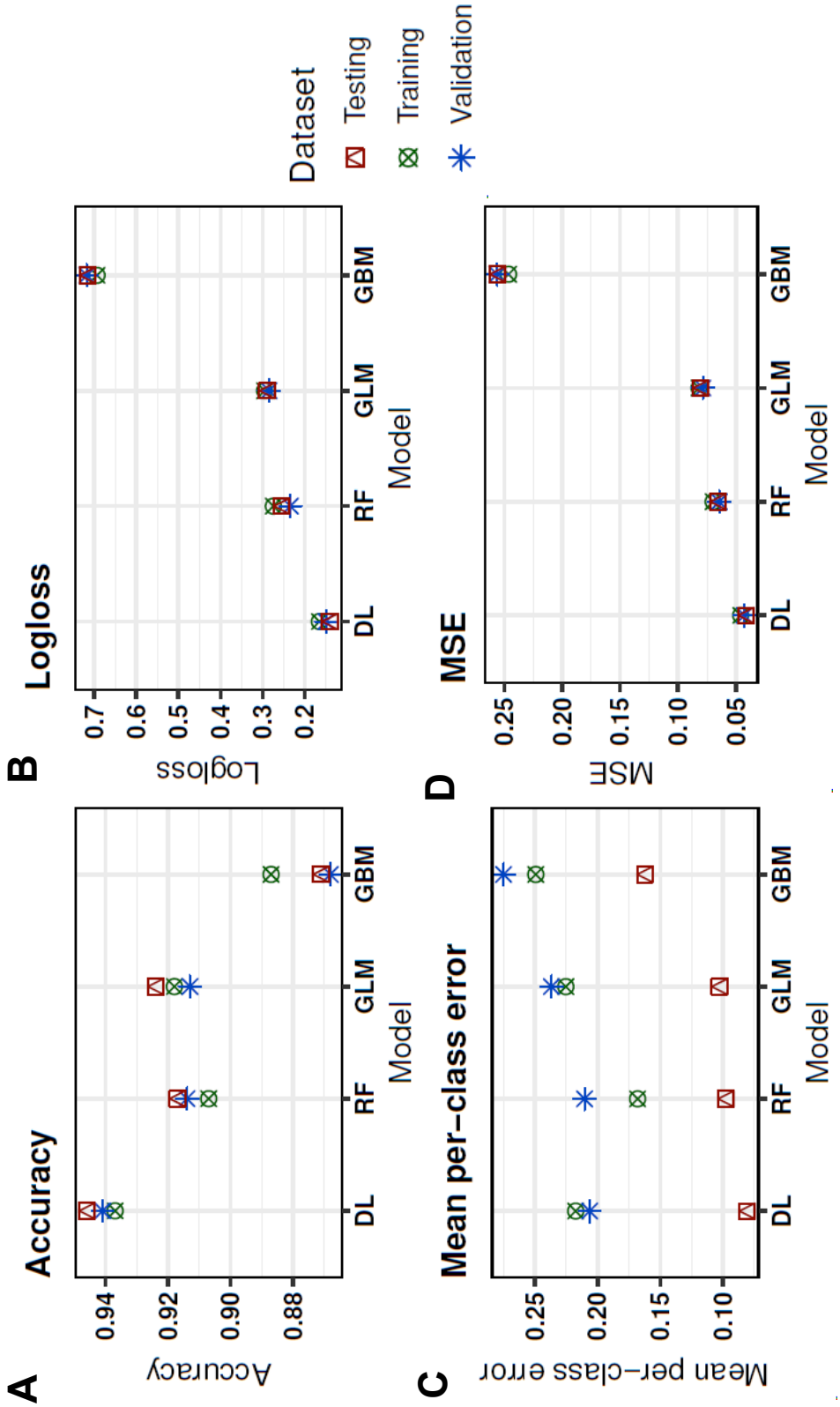

Figure 3.9: Results for DL, RF, GBM, and GLM of Hoshiarpur

3.5.4 Estimation of ET_o using Heuristic Models at Patiala

The DL, RF, GLM and GBM models have been evaluated using training, validation and testing datasets (0.55%, 0.30%, and 0.15% splitting). The performance of the DL is found to be better than the other models in estimating ET_o for Patiala station as shown in Table 3.7 and Table 3.9.

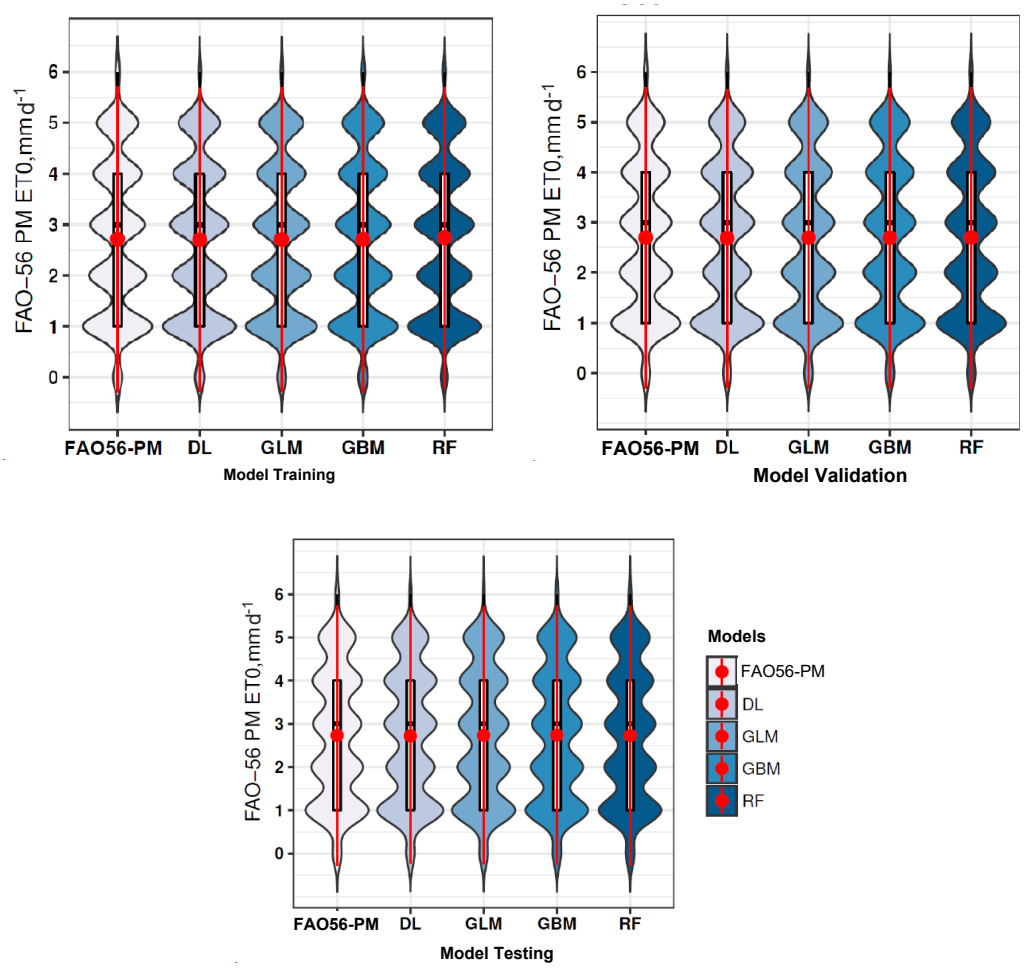

Figure 3.10: Estimated ET_o for actual and predicted results of Patiala

The values of RMSE, MSE, LL, ACC, MCE results during the testing period for the applied models are given in Table 3.9. The DL model values of (MSE= 0.0442, RMSE= 0.210, LL= 0.162, ACC= 0.94, MCE= 0.0594) and (R= 0.98, R^2= 0.97, NSE= 0.97, NRMSE= 16.20%, RAE= 0.0454) shown in Table 3.9.

Similarly, the values of (MSE= 0.0502, 0.0590 and 0.0605, RMSE= 0.0224, 0.243 and 0.245, LL= 0.1951, 0.2158 and 0.2864, ACC= 0.93, 0.92 and 0.91) (0.55%, 0.30% and 0.15% splitting) respectively obtained for the GBM, GLM, and RF models during the testing period. Comparisons of daily ET_o between measured and modeled by DL, RF, GBM and GLM models in the testing period for Patiala site is shown in Fig. 3.11.

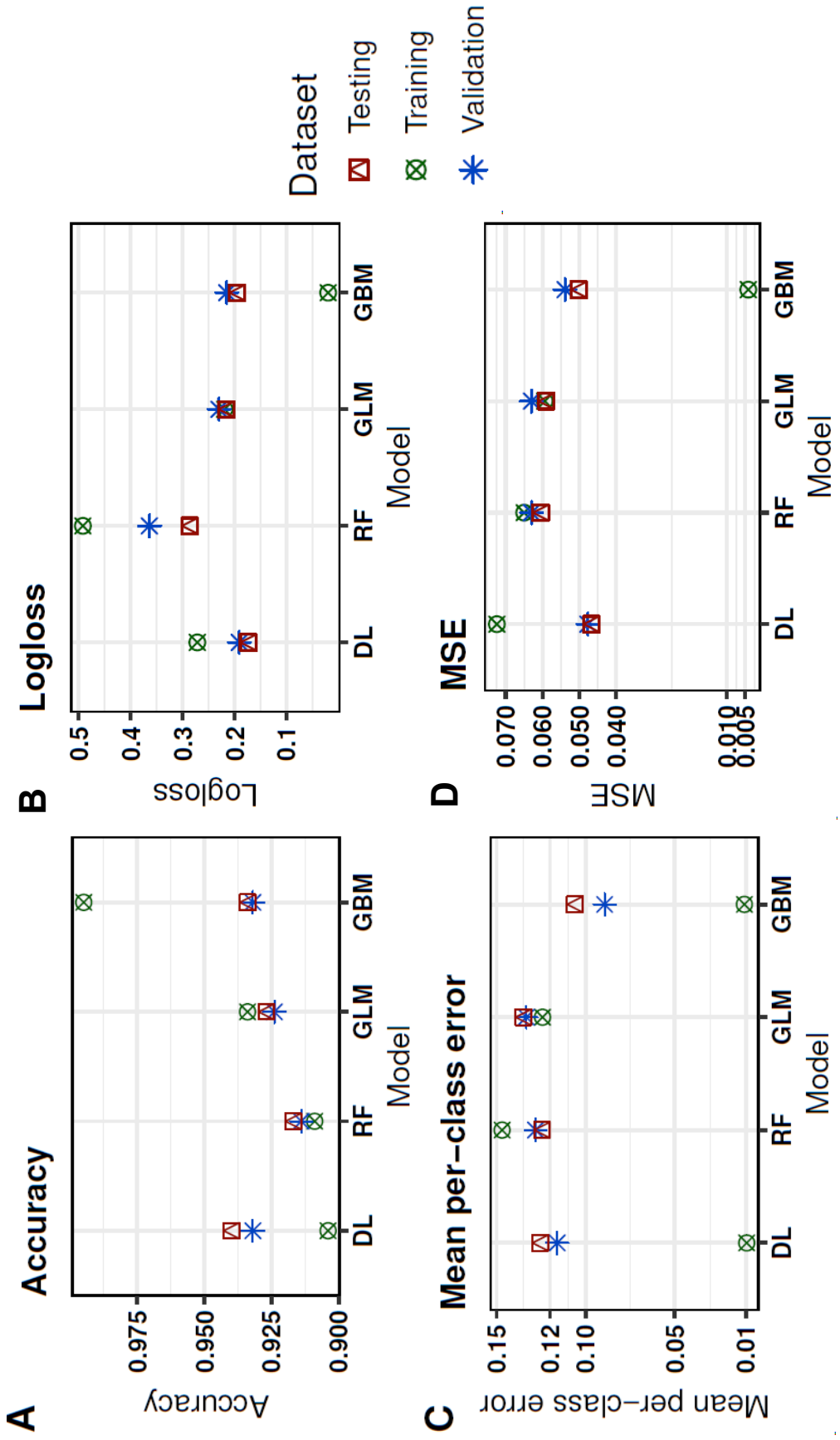

Figure 3.11: Results for DL, RF, GBM, and GLM of Patiala

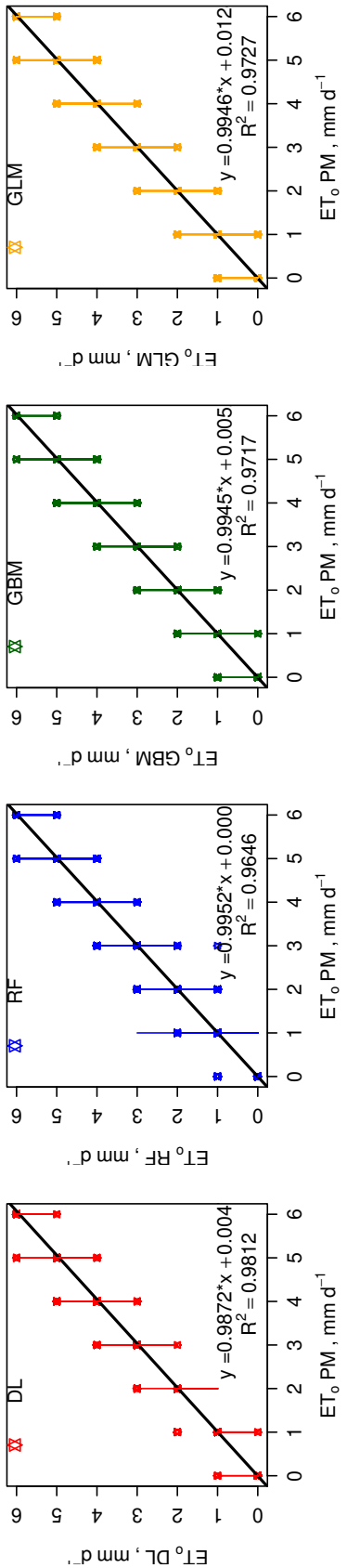

Figure 3.12: Estimated vs Observed daily ET_o in the testing time at Hoshiarpur

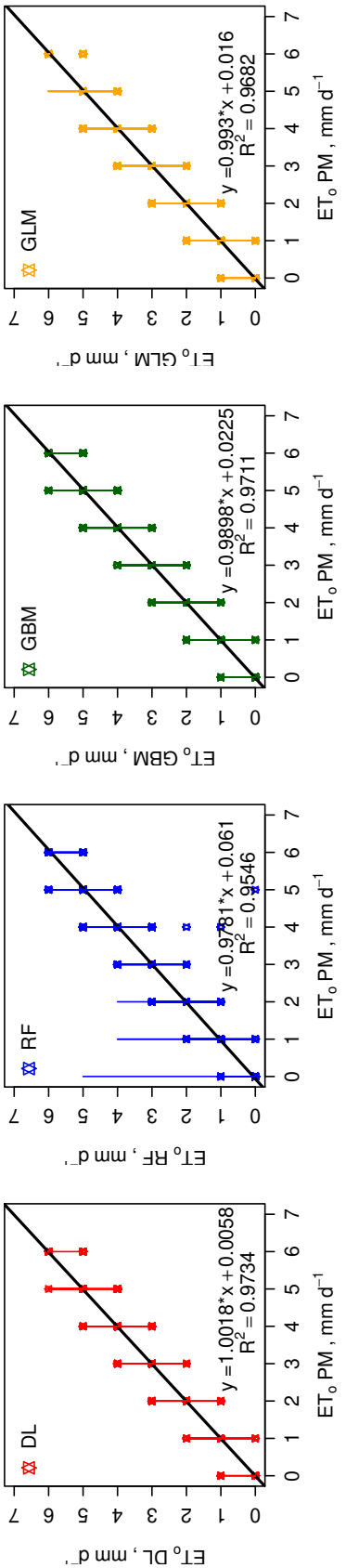

Figure 3.13: Estimated vs Observed daily ET_o in the testing time at at Patiala

Table 3.10: Performance comparison of DL, RF, GLM and GBM models for Patiala

	Dataset	DL	RF	GLM	GBM
RMSLE	Training	0.0738	0.0088	0.0939	0.0256
	Validation	0.0863	0.1046	0.0971	0.0862
	Testing	0.0839	0.1055	0.0928	0.0837
R	Training	0.98	0.99	0.98	0.99
	Validation	0.98	0.96	0.96	0.96
	Testing	0.98	0.96	0.96	0.98
R^2	Training	0.9783	0.9998	0.9713	0.9981
	Validation	0.9706	0.9537	0.9665	0.9700
	Testing	0.9738	0.9542	0.9682	0.9707
NSE	Training	0.97	0.99	0.97	0.99
	Validation	0.97	0.95	0.97	0.97
	Testing	0.97	0.95	0.97	0.97
RAE	Training	0.0376	0.0003	0.0499	0.0031
	Validation	0.0510	0.0684	0.0579	0.0515
	Testing	0.0454	0.0667	0.0552	0.0507
NRMSE	Training	14.7	1.3	16.9	4.4
	Validation	17.1	21.5	18.3	17.3
	Testing	16.2	21.4	17.8	17.1

It is clear from the table 3.9 that the DL model has the lowest RMSE, MSE, LL and the highest ACC as compared to RF, GLM and GBM models. The DL model is superior to the other models in the estimation of ET_o in the testing period, with the highest values of ACC= 0.95 for Hoshiarpur and ACC= 0.94 for Patiala station and the lowest values of MSE= 0.0369 for Hoshiarpur and MSE= 0.0442 for Patiala. Results obtained by four models for Patiala station are shown in Fig 3.10, Fig 3.11, and Table 3.9.

Comparisons with the standard PM method and four models DL, RF, GBM and GLM for predicting daily ET_o in the training, validation and testing of Hoshiarpur and Patiala sites shown in Fig 3.12 and Fig 3.13.

3.6 Summary

In this chapter four data-driven models (DL, GBM, GLM, and RF) have been analyzed under H2O framework for evaluating daily ET_o at Hoshiarpur and Patiala sites in India. The five-fold cross-validation test has been deployed to estimate the performance of considered models. The following conclusions are drawn from the study-

- The combination of six input variables i.e. T_{min}, T_{max}, R_H, u_2, I_s and R_s, have filled the appropriate missing values using MissForest for estimation of daily ET_o for Hoshiarpur and Patiala.

- The newly developed DL model showed great capabilities for ET_o estimation and performed much better than the original and calibrated RF, GLM and GBM models. One of the most significant results to be noted is that the Deep Learning model performed very well and showed very accurate results in comparison of RF, GBM and GLM.

- Moreover, the DL model has avoided the over-fitting issue by giving more accuracy on training, validation, and testing respectively. It has shown higher robustness than conventional approaches.

- DL model provided the most accurate results among the considered models with average (RMSLE= 0.0693, r= 0.98, R^2= 0.99, NSE= 0.98, NRMSE= 13.90 %, RAE= 0.0346) for Hoshiarpur station.

- In this respect, our analysis depicts that models present high performance for modeling daily ET_o (e.g. NSE= 0.95-0.98, R^2= 0.95-0.99, ACC= 85-95, MSE= 0.0369-0.1215, RMSE= 0.1921-0.2691).

Chapter 4

Reference Evapotranspiration ET_o using MPS

Matrix product state (MPS) is a well-designed class of tensor network states that plays an important role in processing quantum information. The MPS, as a one-dimensional array of tensors, can be used to classify classical and quantum data. In this chapter, we have investigated matrix product state models for classification and to predict the reference evapotranspiration ET_o.

4.1 Overview

Quantum computing is a winsome field that deals with theoretical computational systems (i.e., quantum computers) combining visionary ideas of Computer Science, Physics, and Mathematics. It concerns with the behavior and nature of energy at the quantum level to improve the efficiency of computations. Quantum computing relies upon the quantum phenomena of entanglement, superposition and interference to perform operations, which are generally considered as resources for this speed up.

Although, the full influence of quantum computing is probably more than a decade away. But, it has the potential to transform the information processing and promises a wide range of applications in the area of quantum chemistry, high energy physics and condensed matter, which are not tractable on classical computers.

In the last decade, the simulation of open and closed quantum systems has got overwhelming response. The study of tensor network theory taking a central role in quantum physics and beyond. It is simply a countable group of tensors associated by contractions. Tensor network states are a new language, based on entanglement, for quantum many-body systems. Tensor network states are classified on the basis of dimensions along which the tensors are traversed. It is widely used to simulate strongly entangled correlated systems and to represent quantum states and circuits.

Tensor network methods is the term associated with the tools, which are widely employed in experimental and quantum theoretical applications of machine learn-

ing. The matrix product state (MPS) is the most prominent example of tensor network states which is maximally unbalanced. It can be observed by the maximum entanglement entropy without even forfeiting one-dimensional distributions expressiveness. Matrix product tensor networks has the ability to surround the whole input or output state space.

By using classical resources, tensor networks have shown impressive results for supervised and unsupervised learning tasks. Recently, MPS method have been introduced to compress the weights of neural network layers and classify the images. Matrix product state (MPS) as one-dimensional array of tensors can be used to classify classical and quantum data.

Quantum computing can also bring revolution through its ability to handle experimental data, which can produce numerous solutions in various areas such as healthcare, smart city, smart agriculture etc.

We have performed binary classification of agriculture dataset encoded in a quantum state. Further, this study investigated the performance by considering different parameters on the ibmqx4 quantum computer and proved that MPS circuits can be used to attain better accuracy.

Further, the learning ability of MPS quantum classifier is tested to classify evapotranspiration (ET_o) for Patiala meteorological station located in Northern Punjab (India), using three years of historical dataset (Agri). Furthermore, the performance metrics are used to measure its capability.

4.2 Matrix Product State (MPS)

Matrix product state concedes the extent of entanglement in bond dimensions. It is a method of tensor network, where the tensors are connected in a one-dimensional geometry. In fact, any pure quantum state can be described by substituting the coefficients. In MPS, a pure quantum state $|\phi\rangle$ is given in Eq. (4.1):

$$|\phi\rangle = \sum_{\sigma_1,\sigma_2,...\sigma_L}^{d} Tr[M_1^{\sigma_1} M_2^{\sigma_2} ... M_L^{\sigma_L}] \, |\sigma_1, \sigma_2, ...\sigma_L\rangle \tag{4.1}$$

where $M_i^{\sigma_i}$ are complex square matrices, d is dimension, σ_i represents the indices i.e. {0, 1} for qubits and $Tr()$ denotes trace of matrices [195].

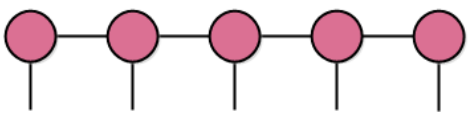

Figure 4.1: Representation of MPS with five sites

Figure 4.1 shows the MPS as one-dimensional array of tensors and an instance

of finite system of 5 sites. It provides an efficient approximation of realistic local Hamiltonians and can be produced sequentially by tensors.

4.3 Encoding of Classical Data

In quantum mechanics, the N independent systems can be combine by performing tensor product operation on their respective state vectors [193] [196]. The feature map is consider as shown in Eq. (4.2)

$$\phi^d(x) = \phi^{s_1}(x_1) \otimes \phi^{s_2}(x_2) \otimes ... \otimes \phi^{s_N}(x_N) \tag{4.2}$$

where s_j are indices run over the local dimension d such that $d=\{s_1, s_2, ..., s_N\}$. Therefore, each state vector x_j is mapped to full feature map $\phi(x)$ in a d-dimensional space. Fig 4.2 shows the tensor diagram of full feature map $\phi(x)$.

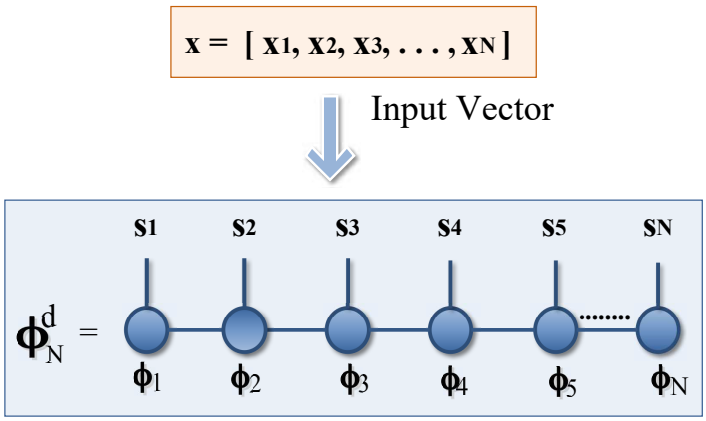

Figure 4.2: Mapping of input vector to order N tensor

Before illustrating the MPS tensor network, it is very crucial to encode classical machine learning dataset into quantum state. Consider a classical dataset $S = \{(x^d, y^d)\}_{d=1}^{D}$ for binary classification, where $y^d \in \{0, 1\}$ are class labels for N-dimensional input vectors such that $x^d \in \mathbb{R}^N$. The input vectors are normalized to lie in $[-\pi, \pi]$. Thus, the qubit ϕ is represented as Eq. (4.3), and Eq. (4.4).

$$\phi_n^d = cos(x_n^d)\,|0\rangle + sin(x_n^d)\,|1\rangle \tag{4.3}$$

$$\phi_n^d = \begin{bmatrix} cos(x_1^d) \\ ssin(x_1^d) \end{bmatrix} \otimes \begin{bmatrix} cos(x_2^d) \\ sin(x_2^d) \end{bmatrix} \otimes ... \otimes \begin{bmatrix} cos(x_N^d) \\ sin(x_N^d) \end{bmatrix} \tag{4.4}$$

The N-dimensional input vectors $x^d \in \mathbb{R}^N$ are mapped to a product state on N qubits by using the feature map as shown by Eq. (4.2). The full quantum data is represented as tensor product $\phi^d = \otimes_{n=1}^{N} \phi_n^d$ [197] [198]. Thus, the preparation of quantum state is efficient as it only needs single-qubit rotations to encode each segment of classical dataset $n = \{1, 2, ...N\}$ in the amplitude of a qubit. Overall,

there is no relevant cost for such encoding. Similar to classical dataset for binary classification, quantum data set for binary classification is denoted as a set $S_q = \{(\phi^d, y^d)\}_{d=1}^{D}$, where $y^d \in \{0, 1\}$ are class labels for 2^N-dimensional input vectors such that $\phi^d \in \mathbb{C}^{2^N}$. It can be easily checked that quantum data as a output of quantum circuit is in superposition state.

4.4 Quantum Circuit Classifier

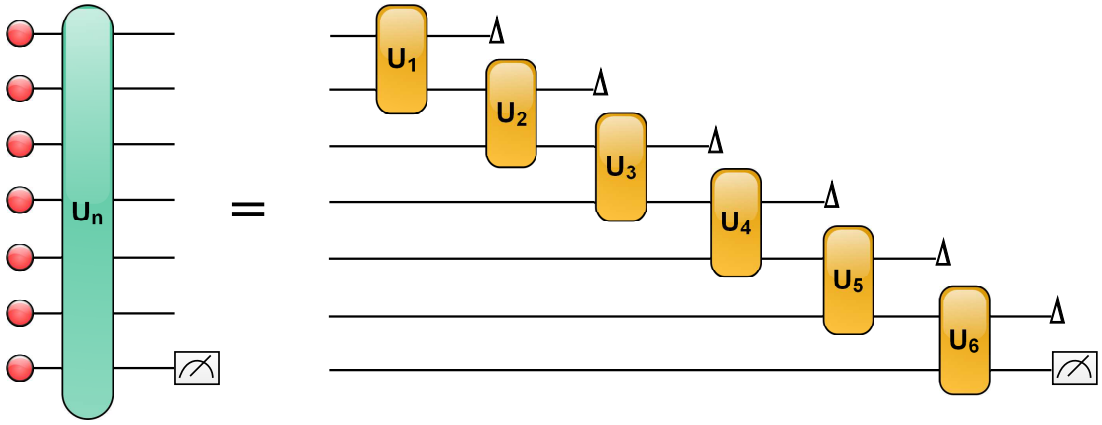

Figure 4.3: Matrix product state quantum classifier

In this section, the MPS quantum circuit classifier is constructed for classification of quantum data, which is made up of unitaries. An iterative approach is followed by keeping positive trace values from N-qubit input space to output qubits. The unitaries composed of single qubit rotations are applied around y-axis. Further, the CNOT gate is applied to the input set, and one of the qubits (unobserved) is discarded from each unitary. Therefore, the qubit is split into two parts for the next layer of the circuit. This process continues till the last qubit is left to be measured.

The unitary blocks in Fig 4.3 consists of input dataset with ancilla qubit which is initialized to zero. It can be easily traced out. Ancilla qubit can be used to execute large class of non-linear operations. In Fig 4.3, circles denote inputs prepared in a product state and triangles indicate unobserved qubits. At the end, when all unitary operations interpreted in the circuit have been executed, then one qubit is observed and used as the output qubit. The measurement on particular qubit is carried out by applying Pauli operators in particular direction. The output qubit determines the predicted value of the input, i.e. the class label values assigned. In order to calculate the most likely state of output qubit, the quantum circuit can be evaluated for number of iterations considering the same input to determine their probability distribution in the computational basis. The MPS quantum circuit for 7 qubits is consisting inputs, unitary blocks $\{U_i\}_{i=1}^{6}$ and measurement operator on

last qubit. Here, single-qubit rotations in the y-direction are followed by a CNOT gate and discard a qubit for next layer of the quantum circuit.

In order to assess the quality of actual and predicted values of dataset, the cost function is needed to calculate. It measures the difference between actual and predicted values of dataset. It is given as Eq. (4.5):

$$J_\theta = \frac{1}{D}\sum_{d=1}^{D}(M_\theta(x^d) - y^d)^2 \tag{4.5}$$

where x^d and y^d are the input and class labels respectively, M is qubit operator, θ represents the set of parameters to define the unitaries and D is total number of data points. It calculates the average amount that the model's predictions differ from the actual values. The goal is to minimizing the cost function i.e. it must be close to zero. Although, there exists various procedures to carry out optimization. The conjugate gradient (CG) method is applied to to optimize the large datasets. It is iterative and effective method to optimize the results. But, it can be break down over multiple iterations when the function to be optimized is noisy. Alternatively, stochastic gradient descent method can be applied which is resilience to noise.

Different parameters are used to measure the performance of MPS quantum classifier such as accuracy (ACC), sensitivity (Sens), specificity (Spec) and gini coefficient. Accuracy is computed to measure the correctness of classifier, Specificity refers the the ability of classifier to identify the negative results. Sensitivity defines the true positive rate i.e. correctly identified by the classifier, Gini coefficient determines the inequality in the distribution and it should be between 0 and 1.

(i) Accuracy:

$$ACC = \frac{TP + TN}{N} \times 100 \tag{4.6}$$

(ii) Sensitivity:

$$Sens = \frac{TP}{TP + FN} \tag{4.7}$$

(iii) Specificity:

$$Spec = \frac{TN}{TN + FP} \tag{4.8}$$

(iv) Gini:

$$Gini = 1 - \sum_{i=1}^{N} p^2(c_i) \tag{4.9}$$

, where N is the total number of data points, p(c_i) is the probability of class, TP is True positive, TN defines True negative, FP and FN represent False positive and False negative respectively.

4.5 Proposed Model

Data Collection and Pre-processing

The MPS based model is developed for classification using agriculture dataset. The daily meteorological data for Patiala during (2014, 2015 and 2016) has been utilized. It consists following parameters: maximum and minimum air temperature (T_{max}) (T_{min}), relative humidity (R_H), wind speed (u_2), solar radiation (R_s), sunshine hours (I_s), evapotranspiration (ET_o), stand-deviation (SD), skewness (SK), and kurtosis (K). The statistical parameters of meteorological variables at Patiala are given in Table 4.1. Fig 4.4 presents the climate Parameters of IMD dataset.

Table 4.1: Statistical parameters of available meteorological variables and ET_o of Patiala station

Parameters	Max	Min	Mean	SD	SK	K
T_{min} (°C)	30.7	2.3	18.71	7.50	-0.27	-1.28
T_{max} (°C)	44.4	9.8	30.39	7.10	-0.47	-0.36
R_H(%)	100	0	73.30	17.64	-0.77	-0.02
u_2 (km h^{-1} day^{-1})	16	0	3.23	2.18	1.46	3.16
I_s (h)	12.2	0	6.24	3.53	-0.52	-0.98
R_s (MJ m^{-2} day^{-2})	28.2	4.9	16.15	6.14	-0.01	-0.98
ET_o (mm)	6	0	2.49	1.48	0.17	1.01

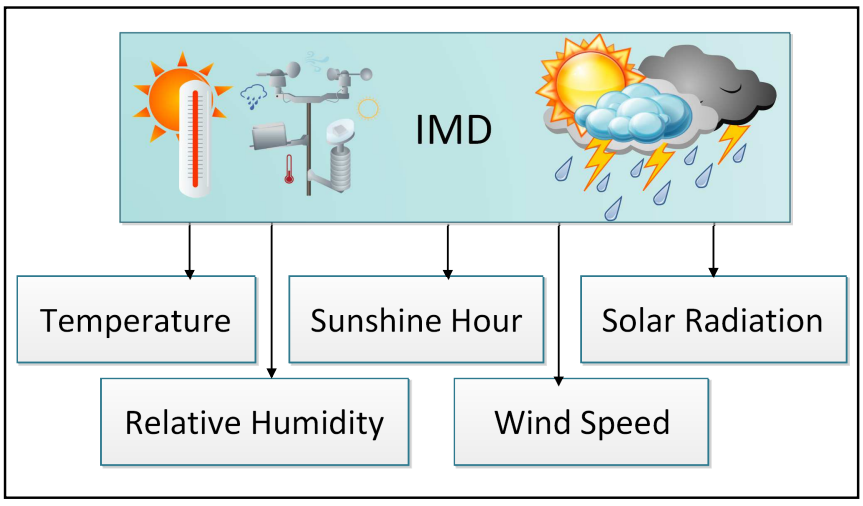

Figure 4.4: Parameters of IMD weather dataset

In agriculture dataset, the ET_o varies from 0 to 6 classes. In order to perform binary classification, each class is divided into three categories, the set {0, 1} comes under LOW , {2, 3} is MEDIUM and set {4, 5, 6} is categorized to HIGH as shown in Table 4.2. The three binary dataset is generated for classification tasks as C_1, C_2 and C_3. Then, the input vectors are re-scaled and element-wise lie in $[-\pi, \pi]$ and applied binary classification to class labels.

Table 4.2: Low, medium and high categories for the estimation of ET_o

ET_o	Categories	Classes	Samples
0 1	(0-1)→ LOW	C_1	$Agri_1$ (C_1 (0) & C_2 (1))
2 3	(2-3)→ MEDIUM	C_2	$Agri_2$ (C_2 (0) & C_3 (1))
4	(4-6)→ HIGH	C_3	$Agri_3$ (C_1 (0) & C_3(1))

After normalization process, the combination of samples ($Agri_1$, $Agri_2$ and $Agri_3$) are formed into pairwise on the basis of classes such as $Agri_1$ consists data belong to class labels C_1 and C_2 (now encoded as 0 and 1). Similarly, $Agri_2$ and $Agri_3$ consist data belong to class labels C_2, C_3 and C_1, C_3 respectively. Further, each sample is divided into training and testing sets. The training set consists of 80% and testing set consists of 20% of the original dataset. Further, MPS quantum classifier is applied to the training dataset. It is repeated for number of iterations considering the same input. After achieving best accuracy, the trained model is applied to the testing dataset (unseen) and results are analyzed.

Development Phases

The training and testing sets are mapped into tensor network state by using Eq. (4.2). The input vectors are encoded into quantum state for classification of classical data on a quantum computer using Eq. (4.3). Finally, taken the tensor product of each input quantum state to form complete quantum data so that it can be ready to use in a quantum circuit using Eq. (4.4).

MPS Quantum Classifier

In this stage, qubit efficient MPS quantum classifier is trained using unitary parameters and qubit rotations in chosen direction. At the end, one or more qubits are measured using Pauli operators. Fig 4.5 shows the seven stages of proposed methodology for model development.

Outcome

In final stage, determined the predicted value of the input, i.e. the class label values assigned for training set after assigning it to the quantum circuit. In order to calculate the most probable state of output qubit, the above stage is repeated for number of iterations considering the same input. Finally, the classification results for each sample are obtained.

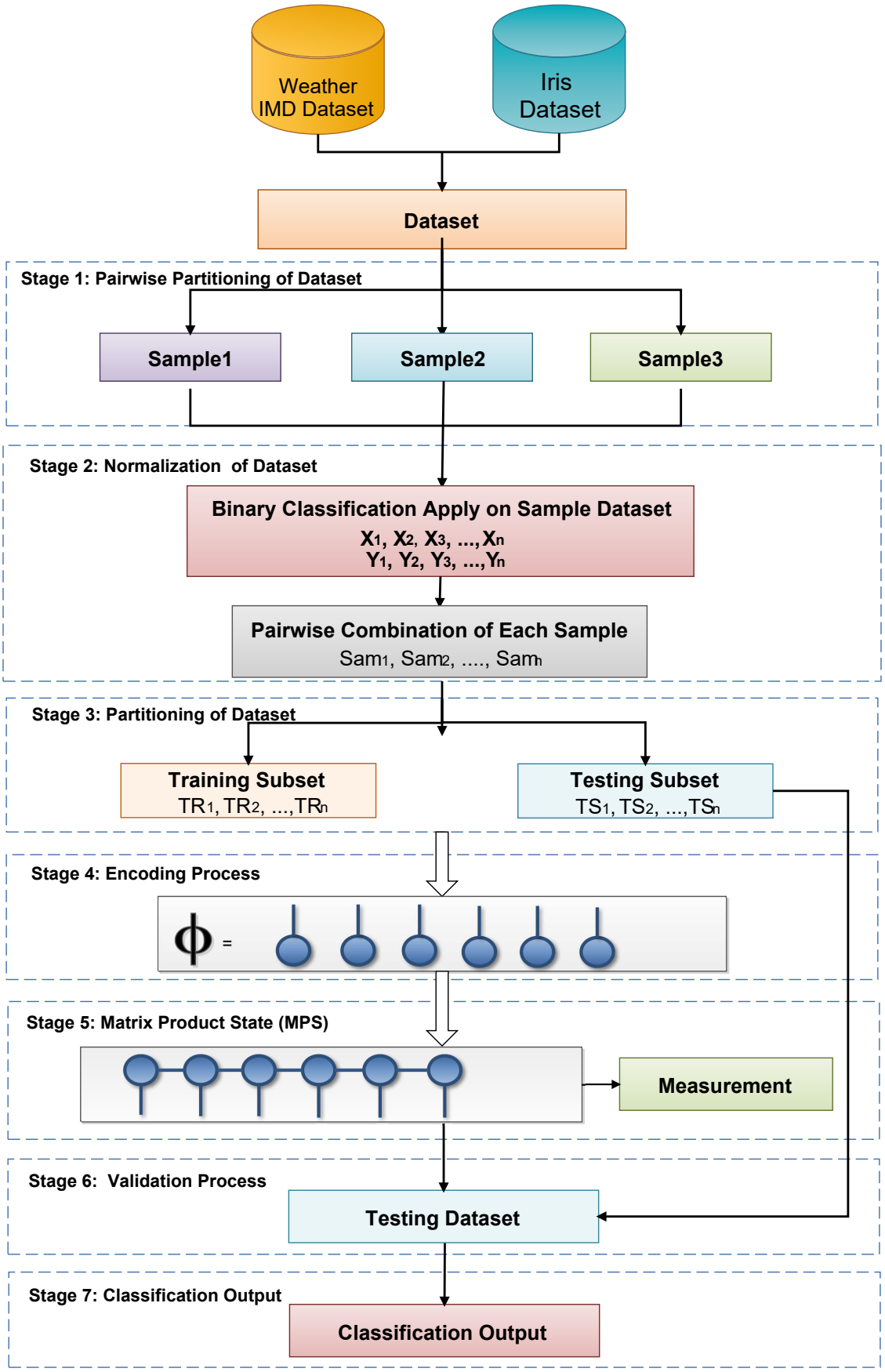

Figure 4.5: Matrix product state for classification

4.6 Simulation

In this section, the data pre-processing, encoding and managing of quantum data that are implemented on real-time quantum device ibmqx4, open source software and a programming language python 3.6.5 with installed Qiskit i.e. an open-source software development kit (SDK) for working with the IBM Q quantum processors.

The ability of MPS quantum classifier is tested to classify the reference evapotranspiration. In experimentation, we have given qubit rotations in y-direction to real values and parameterized the unitaries using ancilla qubit. It is represented as four-qubit input gate consisting an ancilla qubit. It can be traced out in order to execute non-linear operations on dataset. In order to analyze the performance of MPS quantum classifier, the dataset in divided into three samples on the basis of pairwise combination of class labels. Each sample consists 2/3 of the dataset. Further, the dataset is split into a training set and testing test with 80:20 ratio to compute the accuracy.

Here, the experiments are performed with quantum circuits of $N = 4$ qubits. The datasets consisting 2500 quantum states for each of the classes $y \in \{1, 2, 3\}$. The each quantum state is selected as an input into the quantum computer where the MPS quantum classifier is implemented. The four-qubit are used as input gates including ancilla qubit which is set to $|0\rangle$.

Each sample of Agri dataset consists statistical parameters (given in Table 4.1) formed by pairwise combination of classes. In order to test the ability of MPS classifier, the dataset is trained with training set as an input. The process is repeated considering the same input. Further, the testing set i.e. unseen data is given to quantum classifier. The performance of the proposed model during the training and testing period for each sample is given in Table 4.3. Compared with the each samples of training datasets, the accuracy of testing dataset of Agri_1 is just slightly greater. It can be easily checked that training accuracy of Agri_2 and Agri_3 samples is marginally higher than testing samples respectively. In case of training dataset of Agri_1 sample, the specificity is 0.98 approx i.e. MPS classifier identifies more negative results as compared to testing set 0.76. Therefore, the true positive value of training set is less than testing in case of Agri_1.

The estimated ET_o actual and predicted values (in %) for training and testing sets of Agri_1 (Sample1), Agri_2 (Sample2) and Agri_3 (Sample3) are plotted in Fig 4.6.

Fig 4.7 describes the results of training and testing set Accuracy, Cost, Spec, Sens and Gini coefficient for each sample and shows the consistency in accuracy of MPS quantum classifier.

The similarity between actual and predicted values is expressed on the basis

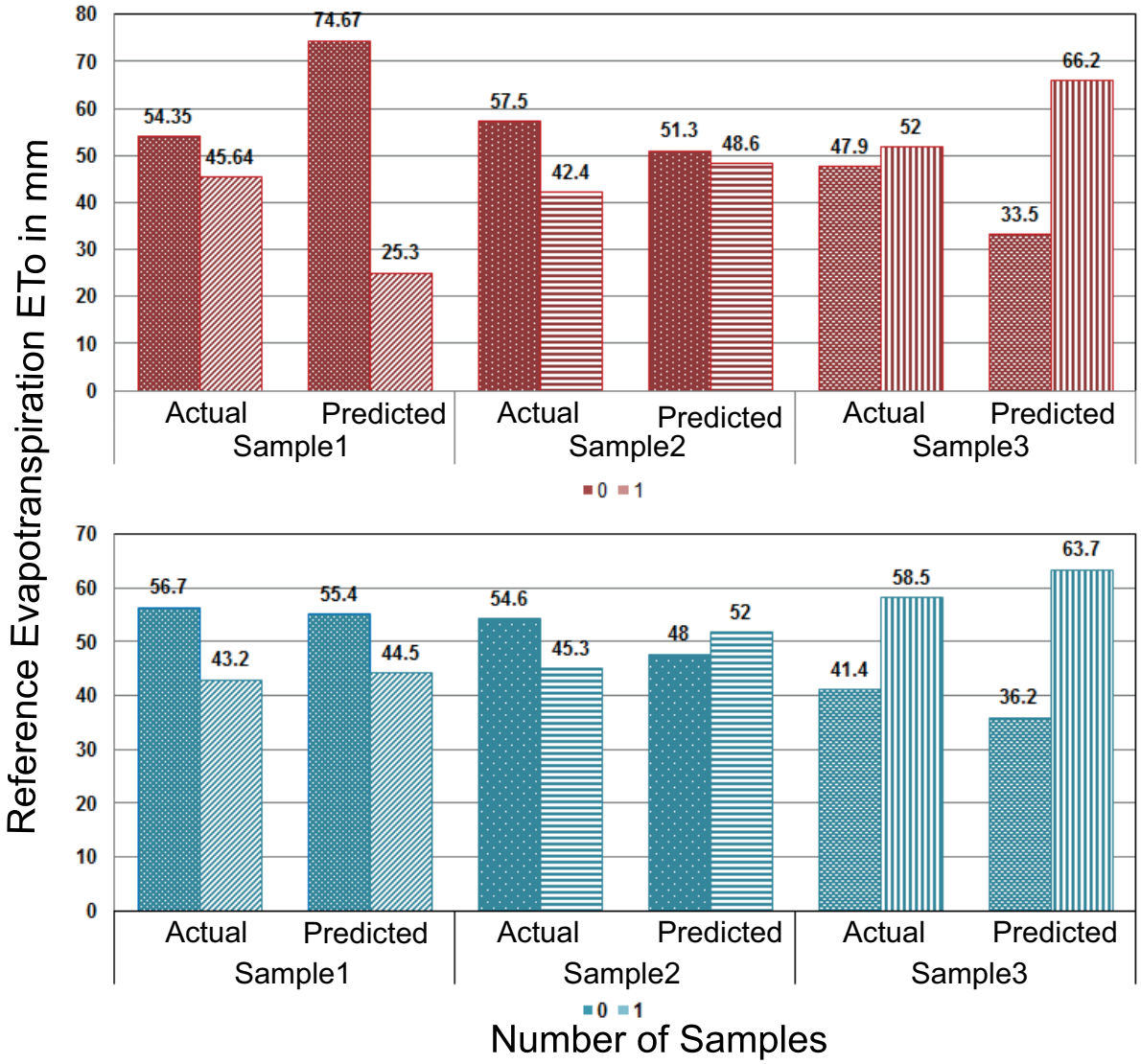

Figure 4.6: Estimated ET_o actual and predicted values (%) using training (in red), and testing (in blue) datasets with three samples

Table 4.3: Performance comparison of MPS for each samples of agriculture dataset

Sample	**Training**					**Testing**				
	Cost	ACC	Spec	Sens	Gini	Cost	ACC	Spec	Sens	Gini
$Agri_1$	0.20	79.03	0.98	0.72	0.52	0.19	80.65	0.76	0.83	0.53
$Agri_2$	0.24	75.34	0.68	0.82	0.51	0.26	73.33	0.67	0.79	0.50
$Agri_3$	0.21	78.73	0.73	0.89	0.61	0.22	77.04	0.77	0.75	0.53

of centered root-mean-square difference, correlation and their variations in amplitude i.e. standard deviation. Fig 4.8 represents the Taylor diagrams to graphically outline the degree of correspondence among values. It have been extensively used to investigate the performance of models to study aspects of climatic environment. The colors indicate actual and predicted values of different samples for Agriculture datasets.

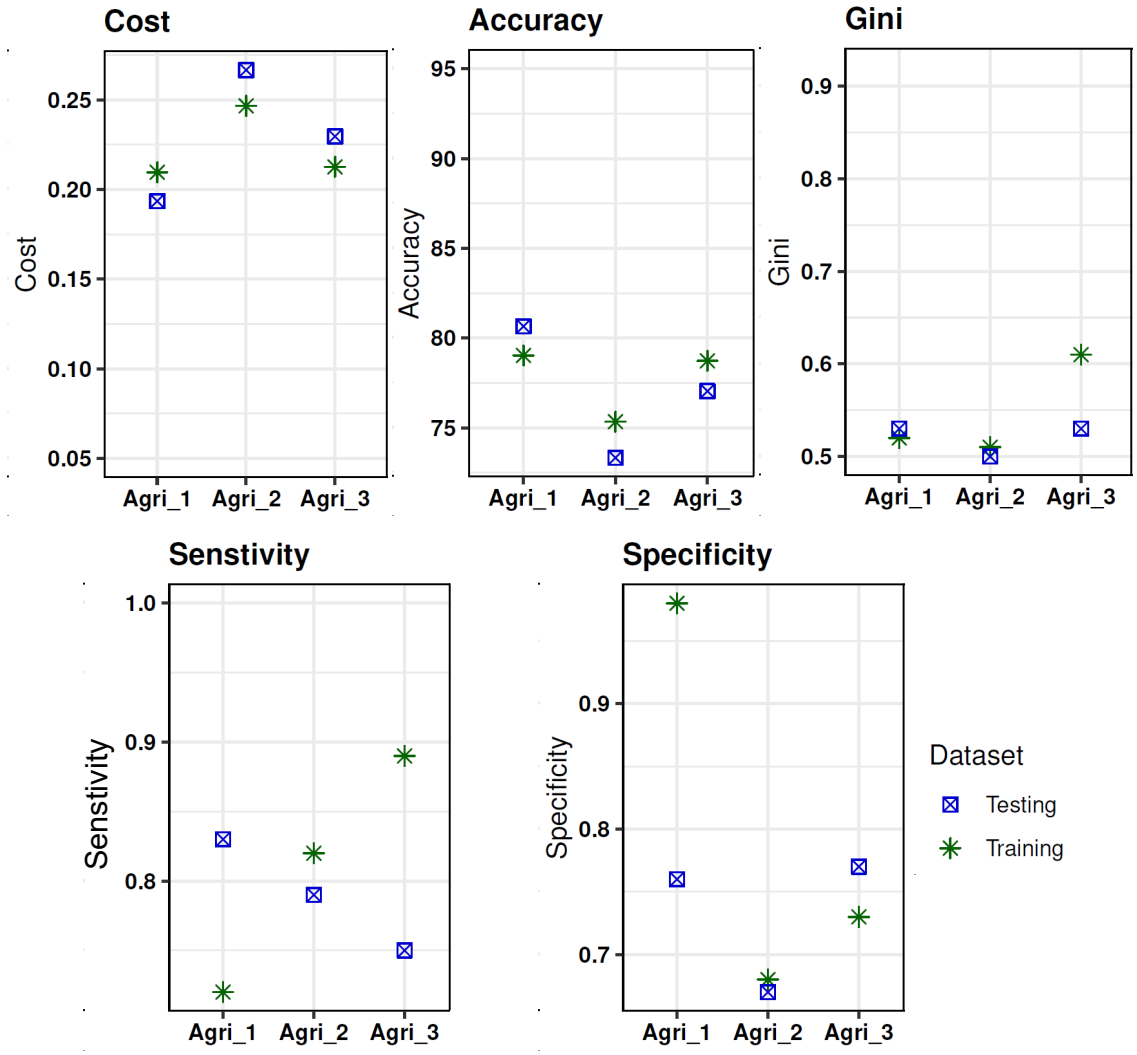

Figure 4.7: Comparison of forecasting ET_o results with MPS model (a) Cost, (b) ACC, (c) Gini, (d) Sens and (e) Spec

The main advantage of MPS quantum classifier is that training can be implemented with high efficiency. The mapping of classical data into MPS form i.e. (highly dimensional) is really beneficial for generating high-order correlations between classes. The bond dimension of MPS manages the parameters of the machine learning model. It is easy to compute and can be selected adaptively. Usually, the bond dimension exists between (10-10,000) or more. It follows that larger dimension of bond results in higher accuracy. In fact, on selecting a extremely large bond dimension, the model can also result in over-fitting, which is not in our case. MPS quantum classifier have been used to avoid over-fitting as well as under-fitting, deal with corrected predictors and reduced the variance of the prediction error. It has been found to perform very effectively and efficiently handled big datasets. We believe that it can be adopted for many other machine learning tasks to scrutinize its power. It has shown great learning capability for ET_o estimations in Agri dataset. Figure indicating the statistical summary of how well patterns match each other in

terms of their correlation, RMSD and the ratio of their variances.

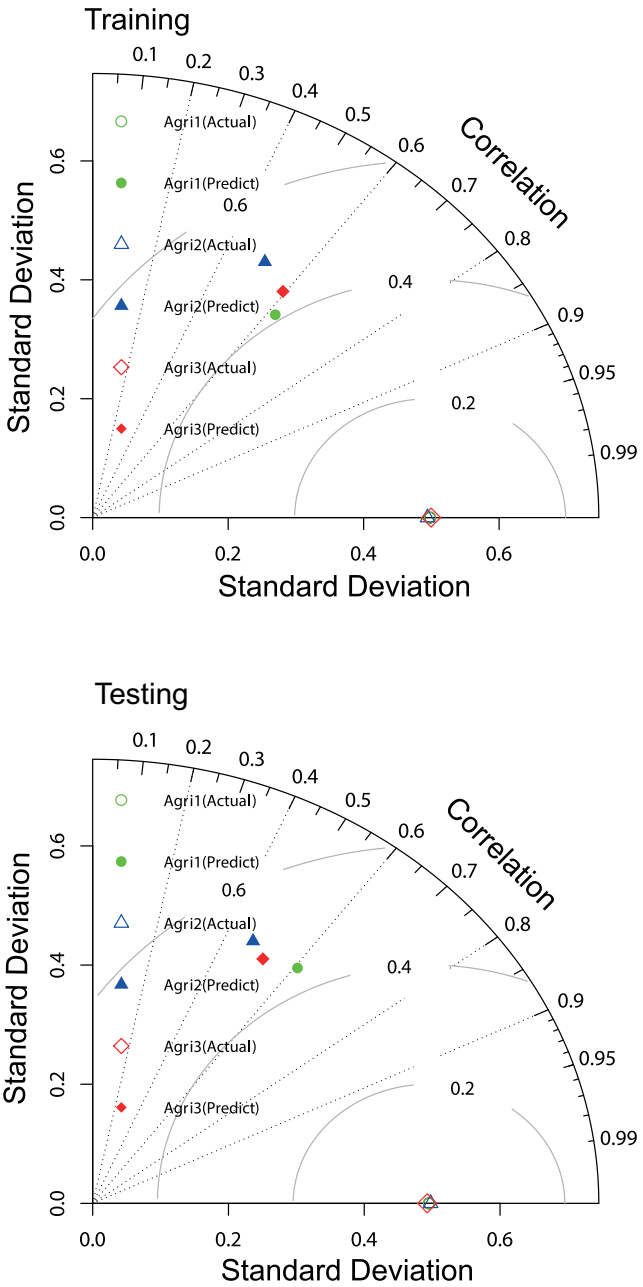

Figure 4.8: Representation of degree of correspondence between each sample of agriculture dataset training$_{Agri}$, and testing$_{Agri}$ using taylor diagrams

4.7 Summary

In this chapter the Matrix Product State (MPS) is used as quantum classifier to predict the ET_o of Patiala station. Mainly focused on MPS quantum circuit augmented with ancilla qubit that is implemented on quantum computer with restric-

tion on qubit rotations to be real only. The key advantage of executing classification with MPS quantum circuit is that it can be executed efficiently with small number of qubits. The following conclusions are drawn from the study-

- In order to analyze the performance of MPS quantum classifier, the dataset is divided into three samples on the basis of pairwise combination of class labels. Compared with the each samples of training datasets, the accuracy of testing dataset of Agri_1 is just slightly greater. It can be easily checked that training accuracy of Agri_2 and Agri_3 samples is marginally higher than testing samples respectively. In case of training dataset of Agri_1 sample, the specificity is 0.98 approx i.e. MPS classifier identifies more negative results as compared to testing set 0.76. Therefore, the true positive value of training set is less than testing in case of Agri_1.

- MPS quantum classifier has been used to avoid over-fitting as well as under-fitting, deal with corrected predictors and reduced the variance of the prediction error. It has been found very effective and efficient in handling big datasets. Moreover, it has shown great learning capability for ET_o estimations in Agri dataset.

Chapter 5

Reference Evapotranspiration ET_O using Multi-Ensembling

The main overarching goal of this chapter is to investigate the abilities and applicability of Three supervised machine learning models: Extreme Machine Learning, Multi-layer Perceptrons-Neural Network, Support Vector Machine to model daily ET_o process. Further, a three-layer multi-model ensemble machine learning approach is presented to predict evapotranspiration (ET_o).

5.1 Overview

Quantification of transpiration (T) to evapotranspiration (ET) from crops is critical in Irrigation Scheduling [1]. However, several research work have compared the ET methods and analyzed the the seasonal variability of (ET_o) and (T) under different climate conditions. The study of general atmospheric and geographic conditions of Punjab (India) received an overwhelming response from research communities. The serious decline in the ground water resources has threatened the success of Green Revolution in terms of crop yield. Moreover so many due to climatic change, a variation in meteorological aspects influences the water requirement for crops, evapotranspiration, and water allocation of agro-meteorological and agriculture. Accurate estimation of Evapotranspiration (ET_o) has great importance to improve the utilization of water efficiently and irrigation scheduling. Till now, the development of more precise and reliable techniques for estimating evapotranspiration has attained attention by several hydrologists.

However, many scientific disciplines advocate routine adoption of machine learning (ML) methods, which will have differing levels of success. Various ML algorithms have been available for decades such as most notably neural networks [199]. Gautam et al. [200] proposed the IoT based smart system to monitor and forecast the water consumption in an urban areas.

ML approaches are of two types: Supervised and Unsupervised. Supervised

learning (SL) is used to predict the target outcome based on mathematical and statistical modeling for regression and classification problem. Regression analysis is used to predict a continuous outcome and classification for predicting the class individuals. In this chapter, we have investigated the some regression based models for modeling the ET_o for Patiala station. For advanced machine learning methods, three machine learning models and multi-ensembling technique are used to build the model accurately.

Recently, ML techniques like Extreme Machine Learning (ELM), Multi-Layer Neural Network (ML-NN), Support Vector Machines (SVM), have been applied to estimate the evapotranspiration. Various researchers applied the ANN model to overcome the limitations of empirical based methods. In the next section these algorithms are discussed in detail.

5.2 Extreme Machine Learning (ELM)

ELM model is developed for generalized single-hidden layer feed-forward neural networks (SLFNs) [201]. The architecture of ELM model consist of three layers: Input layer, Hidden layer, and Output layer [202].

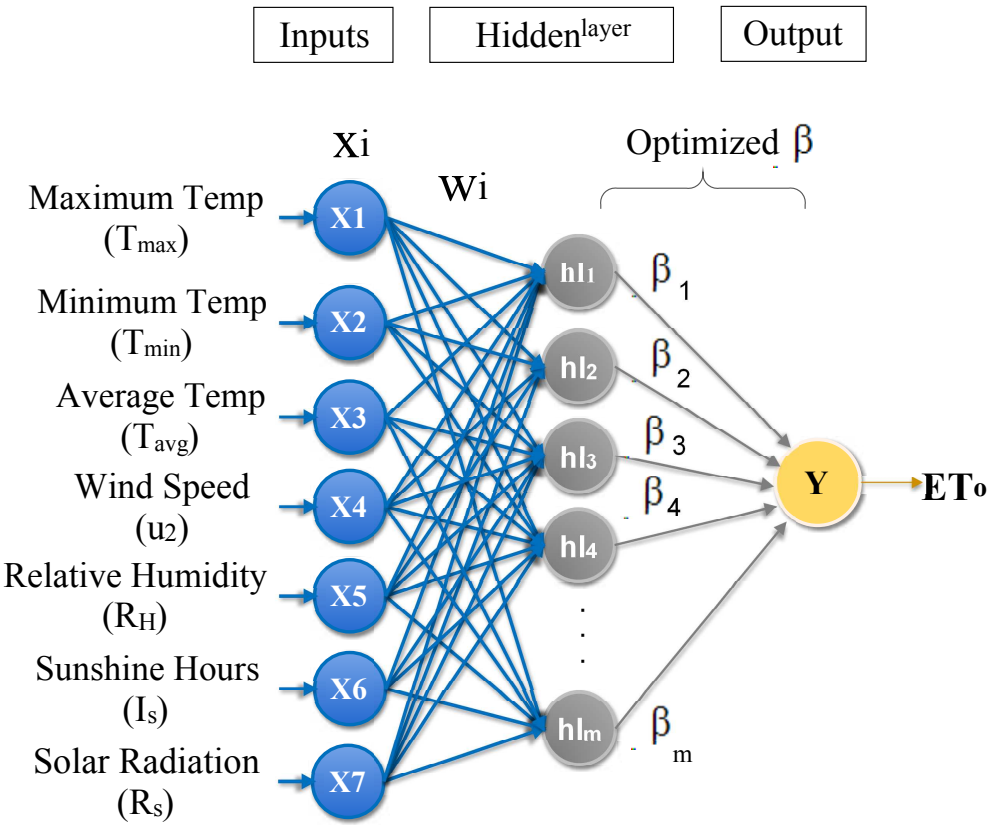

Figure 5.1: Basic structure of the ELM model

The basic structure of ELM is shown in Fig. 5.1. Consider a V arbitrary distinct

samples (x_t, y_t) with $x_t \epsilon R^d$ and $y_t \epsilon R$ then, a SLFN with V hidden nodes is modeled by 5.1.

$$Y_j = \sum_{t=1}^{V} \eta_t f(w_{it}, x_j + b_t), j = 1, ..., V \tag{5.1}$$

where, w_t is the weights, b_t is the bias, f being the activation function, η is the output weight matrix which can be written in compact form by Eq. (5.2).

$$Z\eta = Y \tag{5.2}$$

where, Z is the output matrix of hidden layer and expressed by Eq. (5.3):

$$Z(\tilde{w}, \tilde{x}, \tilde{b}) = \begin{bmatrix} f(w_1, x_1, b_1 & ... & f(w_L, x_1, b_L)) \\ : & & : \\ : & & : \\ f(w_1, x_v, b_1 & ... & f(w_L, x_v, b_L \end{bmatrix}_{VxL} \tag{5.3}$$

where, $\tilde{w} = w_1, w_2 ... w_L$, $\tilde{x} = x_1, x_2 ... x_V$, $\tilde{b} = b_1, b_2 ... b_L$ η denotes the training output matrix by Eq. (5.4)

$$\eta = \begin{bmatrix} \eta_1^T \\ \eta_2^T \\ \eta_3^T \\ : \\ \eta_L^T \end{bmatrix}_{L \times N} \quad and \quad \eta = \begin{bmatrix} \eta_1^T \\ \eta_2^T \\ \eta_3^T \\ : \\ \eta_L^T \end{bmatrix}_{L \times N} \tag{5.4}$$

The activation function *f(x)*: Re $\rightarrow$ Re for the additive hidden nodes such as Positive Linear Function (Poslin), and Linear Function (Purelin) are given by Eq. (5.5)

$$f(w_t, x, b_t) = g(w_t * x + b_t), b_t \in \text{Re} \tag{5.5}$$

The main process of the ELM can be summarized as the following three steps:

(i) Randomly generates input weights w_t and $bias_t$ before training, $1 \leq t \leq V$.

(ii) Compute the output matrix Z with hidden layer.

(iii) Compute the output weight matrix as $Y = Z^{\dagger}\eta$.

where, $Z^{\dagger}$ is referred as a Moore–Penrose generalized inverse of the matrix Z [203].

5.3 Multi-layer Perceptrons Neural Network (MLP-NN)

Multi-layer Perceptrons is defined as a feed forward artificial neural network (FF-ANN) which consists at-least three-hidden layers of nodes and it is calculated by Eq. (5.6).

$$f(x) = \sum_{i=1}^{m} w_{ij} * x_i + b \tag{5.6}$$

It is an efficient NN and significantly used to solve different tasks such as regression, time-series modeling, and classification. It is organized as hierarchical NNs consisting an input layer, hidden layer(s), and the output layer (Wang et al., 2016) [13]. It consist one or more hidden layers, which are connected by neurons between the input and output layers (i.e. Synaptic weights, Biases, and Activation functions).

Each node takes the input value(s) from an input parameter and determines a weighted sum of input values progress through the transfer function, which produces the outcome of nodes.

(i) Input layer is expressed by Eq. (5.7):

$$x \rightarrow x_1, x_2....., x_m \tag{5.7}$$

(ii) Calculate the sum of product and output of each hidden layer in order from $h\ \ell_1$ to $h\ \ell_{m-1}$ and shown by Eq. (5.8) to Eq. (5.10):

$$h\ell_1 = f(\sum_{i=1}^{m} w_i * x_i + b_i) \tag{5.8}$$

$$h\ell_2 = f(\sum_{j=1}^{m} h\ell_1^T w_j + b_j) \tag{5.9}$$

$$h\ell_3 = f(\sum_{k=1}^{m} h\ell_2^T w_k + b_k) \tag{5.10}$$

(iii) Calculate the output layer and expressed by Eq. (5.11):

$$\begin{aligned} y = f(\sum_{k=1}^{m} h\ell_3^T w_m + f(\sum_{k=1}^{m} h\ell_2^T w_k + f(\sum_{j=1}^{m} h\ell_1^T w_j + \\ f(\sum_{i=1}^{m} w_i * x_i + b_i b_j b_k b_m)))) \end{aligned} \tag{5.11}$$

where, Y is the output , x_i expressed as inputs i, w_i, w_j, w_k, w_m are weight connection among input and hidden layer or neurons from input i to layer m, b_i, b_j, b_k, b_m = bias, f is a transfer function for hidden neurons from i, j, k to m hidden layers respectively.

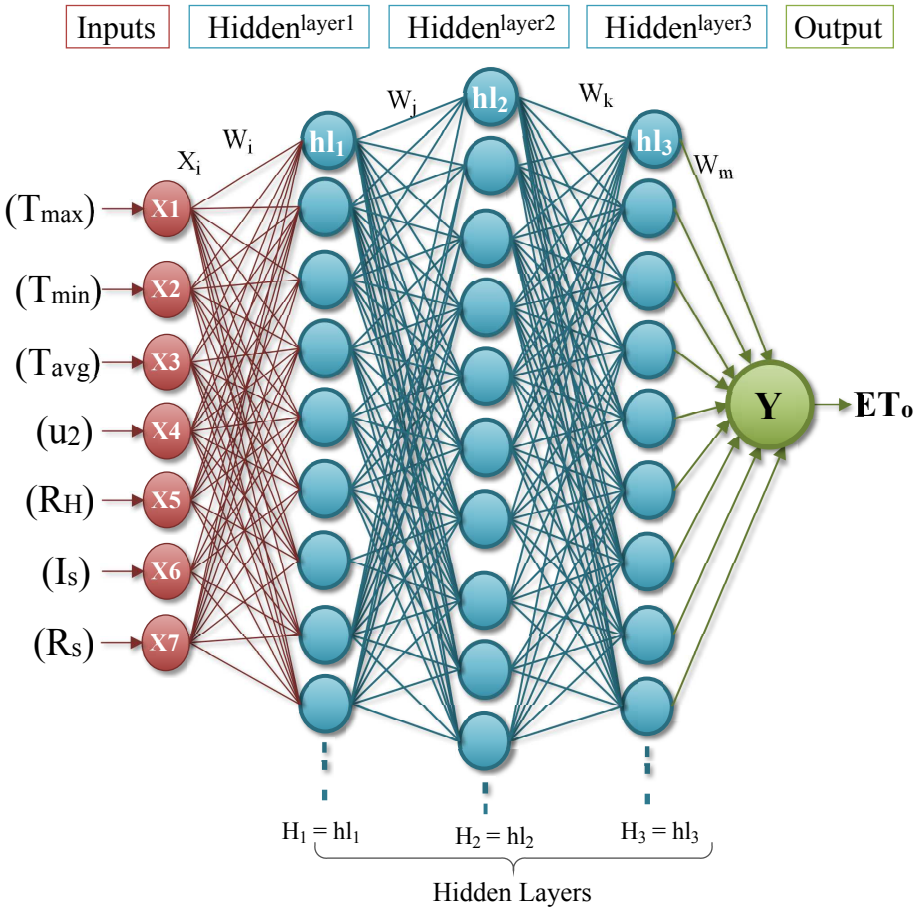

Figure 5.2: Basic structure of MLP-NN model

The MLP structure is shown in Fig 5.2, it includes an input layer, three hidden layers ($h\ \ell_1$, $h\ \ell_2$, $h\ \ell_3$), and the output layer. A backpropogation MLP used gradient descent technique to minimize the error function.

5.4 Support Vector Machine (SVM)

SVM is a machine learning approach which creates input-output mapping functions from a collection of labeled training data. It has three layers: Input layer, Hidden layer, and Output layer as shown in Fig. 5.3. Classification and regression problems can be used as mapping functions.

Initially, SVM began as an application of Vapnik's Structural Risk Minimization (SRM) principle [204], which is recognized for low generalization error and also avoid the over-fitting to the training dataset. It consider various types of kernel functions such as Radial Basis function (RBF), Linear Function and Sigmoid Function [205].

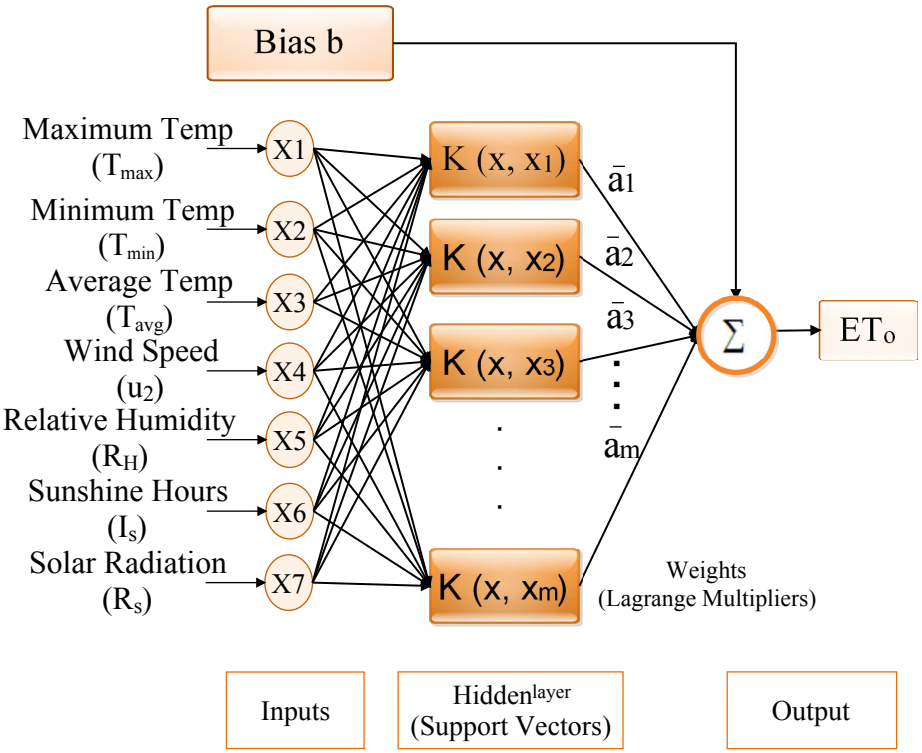

Figure 5.3: Basic structure of SVM model

Several fields have shown success with SVM as a classifier including (Smart Agriculture, Image or Pattern recognition [206], and Probability-based decision) [207]. The object function for the SVM method and the regression function can be expressed as Eq. (5.12) [208]:

$$f(x) = (w.x_i) + b \tag{5.12}$$

A set is given (x_1, y_1), (x_2, y_2),....., $(x_m, y_m) \epsilon (X, Y)$ where X = denotes the input set of parameter, Y = denotes the output, Y refers to an unseen x ϵ X. $(x_i, y_i) \subset$ X x Re, i= 1, 2, 3,...,m and it is calculated by Eq. (5.13).

$$Y = f(x) = \sum_{k=1}^{m} \bar{\alpha}_k.K(x, y_k) + b \tag{5.13}$$

To predict the value among the different kernel functions linear, sigmoid, and RBF is applied: *Linear*: The Linear kernel is the simplest kernel function defined as Eq. (5.14):

$$Linear : K(x_i, y_i) = x_i^T y_i \tag{5.14}$$

Radial: The Gaussian kernel is called as RBF, where σ signifies a window width defined as Eq. (5.15):

$$Radial : K(x_i, y_i) = exp(-||x_i - y_i||^2/2\sigma^2) \tag{5.15}$$

Sigmoid: The sigmoid kernel is expressed by Eq. (5.16):

$$Sigmoid : K(x_i, y_i) = tanh(k(x_i y_i) + b) \tag{5.16}$$

RBF kernel provides an efficient performance as compare to other kernels [209]. The kernal (sigma) and the C parameters were used to develop the SVM model. In this study, three common types of kernel functions, i.e. Linear (Eq. (5.14)), Sigmoid and RBF (Eq. (5.15, 5.16)) are employed in modeling process of ET_o as above.

5.5 Proposed Model

The three-layer multi-ensemble machine learning approach is proposed to predict evapotranspiration ET_o. The first layer consists of different statistical models to produce individual forecasts. The blending approach is employed to create an ensemble of the forecasts generated by the initial layer to produce probabilistic forecasts. In the second layer, three ensemble models ($Ensemble_{ELM}$, $Ensemble_{MLP}$, $Ensemble_{SVM}$) are trained for the prediction of ET_o by using the previous layer predictions with training data. In the third-layer, accuracy of the ET_o is estimated by tuning the parameters of second layer ensemble model.

The proposed framework of multi-level ensembling ET_o forecasting model (MLE-ET_o) is presented in Fig 5.4. It consists five phases: (i) Input dataset, (ii) Partitioning of dataset, (iii) Model development, (iii) Multi-level ensembling for ET_o, and (v) Quantified prediction output.

In first phase, as depicted in Fig. 5.4 dataset are collected from IMD, Pune for Patiala site, then the ET_o and R_s were computed using CROPWAT 8.0 software.

The dataset is divided into three portion: Training, Validation and Testing datasets as shown in second phase. Our training dataset range contains 7027 data vectors with (50%), for validation contains 4096 data vectors with (30%), and the testing dataset contains data vectors 2756 with (20%). In fact, the validation and testing datasets are employed to make sure that the applied models do not over-fit to the training datasets.

In third phase, the major objective behind the development of ensemble approach is to introduce a predictive model by combining homogeneous various machine learning models. The nine machine learning models are trained, validate and tested and then blend them into an ensemble. Three different machine learning models (ELM_1,.., ELM $_3$), (MLP_1,.., MLP_3), (SVM_1,.., SVM_3) as learner algorithm are developed.

After combining the aforementioned models, ensembling is performed to design the homogeneous multilevel model and to generate a prediction of each model. To maximally utilize potential and to ensure the greater diversity of models, three ELM

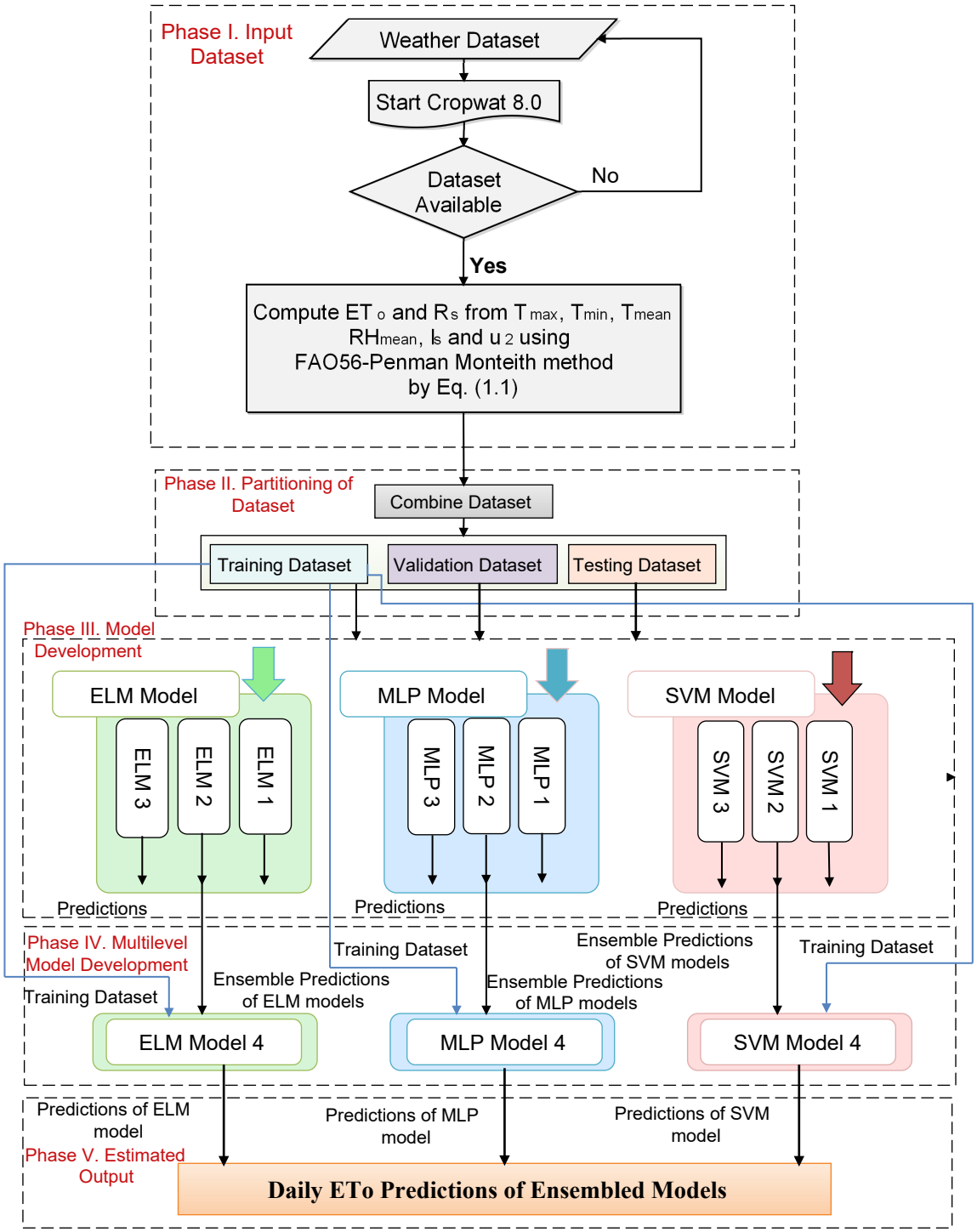

Figure 5.4: Framework of proposed MLE-ET_o forecasting model

models are applied with different activation function and hidden nodes. Similarly, we have created three MLP models with different multi hidden layer and also created three SVM with different kernel functions, i.e. cost, gamma and degree parameters.

In fourth phase, data-driven models SVM, MLP, and ELM have been applied to estimate the daily ET_o. We have blended the each model predictions with training dataset and prepared a new dataset to train the next level models such as (ELM_4,

MLP_4, and SVM_4) models. In another words, to obtain accurate forecasts, multiple machine learning models from the third layer are included in the ensembles. In fourth layer, considered models are used to train a third layer of models and it generate a prediction output by ensemble the models and use their ensemble predictions outputs (i.e. probabilities) to train a fourth layer of models such as (ELM_4, MLP_4, and SVM_4).

In fifth phase, the new prediction ET_o is estimated using the ensemble of each models of fourth layer. The fourth phase model is trained such as (ELM_4, MLP_4, and SVM_4) with training dataset and ensemble prediction of each models. The newly modeled ET_o predictions have been verified using accuracy of each model with the training, validation and testing set. Then, the model with best accuracy is determined

Using a data blending and Machine Learning approach, this study seeks to develop and evaluate a prediction model that is able to predict the daily ET_o values.

5.6 Simulation Results and Discussion

This work presents the performance of twelve machine learning models such as (ELM_1,..,ELM_3), (MLP_1,..,MLP_3), (SVM_1,..,SVM_3), and ($Ensemble_{ELM}$, $Ensemble_{MLP}$, and $Ensemble_{SVM}$) to predict the ET_o. Table 5.1 presents the detail of applied model parameters and essential packages. The "caret" and "e1071" packages of R software [210] is used for the SVM modeling.

Fig. 5.6 and Fig. 5.7 presents the comparison among considered models in the validation and testing datasets respectively. As shown in Fig. 5.8, the $Ensemble_{ELM}$, $Ensemble_{MLP}$, and $Ensemble_{SVM}$ models presented best forecasted ET_o in the training, validation, and testing periods for Patiala station.

The $Ensemble_{MLP}$ model shows better performance on validation and testing datasets as compare to training dataset in terms of ACC (96.4, 98.9 and 99.1). While the training dataset generates highest accuracy as compare to validation and testing periods (99.41 and 99.72), (99.0 and 99.5), (99.1 and 99.4) of $Ensemble_{ELM}$ and $Ensemble_{SVM}$ models respectively. Fig. 5.5 presents the comparison of the obtained ET_o through ELM, MLP, and SVM models and observed ET_o for training dataset.

The comparative study of the results is shown in Table 5.2. For comparison three performance metrics such as Mean Square Error (MSE), Accuracy (ACC), and Coefficient of Determination (R^2) parameters are used. Results shown in this table indicate the superiority of ELM_2 model in terms of ACC, MSE, and R^2. It is also observed that ELM_2 model takes very less time 0.01(s) during training comparison to the SVM_2 model 27.26(s).

Table 5.1: Model parameters

Model Name	Packages	Parameters
ELM 1	elmNN, caret	actfunction= poslin, nhid= 10, regression
ELM 2	elmNN, caret	actfunction= purelin, nhid= 10, regression
ELM 3	elmNN, caret	actfunction= poslin, nhid= 10, regression
MLP 1	mlpML, caret	hidden layer 1= 10, layer 2= 10, layer 3 = 10
MLP 2	mlpML, caret	hidden layer 1= 20, layer 2= 40, layer 3 = 60
MLP 3	mlpML, caret	hidden layer 1= 10, layer 2= 50, layer 3 = 10
SVM 1	e1071, caret	kernalfun= sigmoid, type= eps-regression, cost= 1000, gamma= 0.0001, eplison= 0.001
SVM 2	e1071, caret	kernalfun= radial, type= eps-regression, cost= 1000, gamma= 0.0001, eplison= 0.001
SVM 3	e1071, caret	kernalfun= linear, type= eps-regression, cost= $2\hat{}(2{:}8)$, gamma= 1, eplison= 1, degree= 1
En_ELM 1	elm, caret	actfunction= poslin, nhid= 50, regression
En_MLP 2	mlp,caret	size= 10, tuneLenth=10, number=5, method= 'cv '
En_SVM 3	e1071,caret	kernalfun= svmRadial, sigma=c(.01, 0.15, 0.1, 0.2, 1.0, 2.0) , C=c(0.7, 0.9, 1, 1.1, 1.25, 2.0)

Table 5.2: Performance comparison of selected models for Patiala with ACC, MSE and R^2

Models	Training			Validation			Testing		
	ACC	MSE	R^2	ACC	MSE	R^2	ACC	MSE	R^2
ELM 1	86.08	0.1214	0.941	87.01	0.1150	0.948	86.28	0.1189	0.947
ELM 2	97.87	0.0374	0.982	97.53	0.0416	0.981	97.31	0.0404	0.982
ELM 3	80.82	0.1516	0.932	80.66	0.1528	0.981	81.93	0.1434	0.937
MLP 1	96.29	0.0524	0.973	95.58	0.0611	0.973	96.52	0.0517	0.978
MLP 2	96.44	0.0485	0.971	96.07	0.0557	0.975	96.63	0.0483	0.979
MLP 3	95.18	0.0601	0.972	95.12	0.0656	0.973	95.43	0.0577	0.977
SVM 1	97.27	0.0391	0.981	97.22	0.0448	0.982	97.13	0.0427	0.981
SVM 2	98.92	0.0103	0.992	98.58	0.0161	0.995	98.62	0.0133	0.994
SVM 3	97.64	0.0378	0.982	97.51	0.0424	0.983	97.39	0.0411	0.981
En_ELM 1	99.41	0.0108	0.991	99.05	0.0414	0.990	99.13	0.0133	0.996
En_MLP 2	96.14	0.0592	0.974	98.93	0.0337	0.986	99.13	0.0336	0.985
En_SVM 3	**99.72**	**0.0073**	**0.995**	**99.51**	**0.0084**	**0.993**	**99.46**	**0.0087**	**0.994**

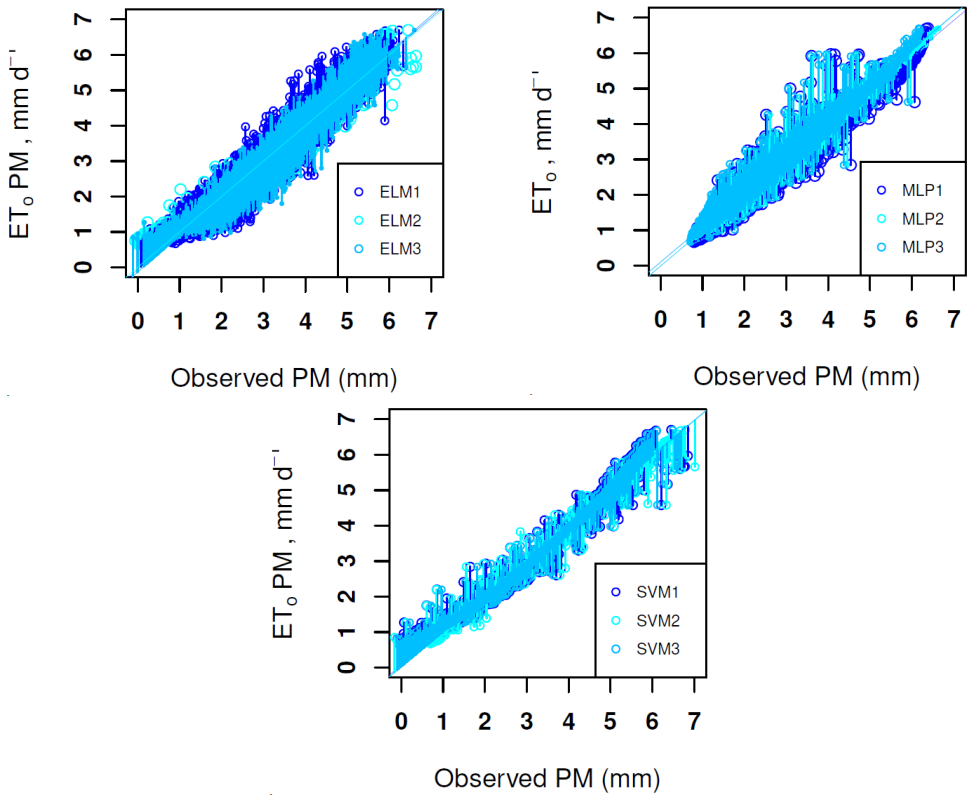

Figure 5.5: Estimated ET$_o$ with ELM, MLP and SVM models using training dataset

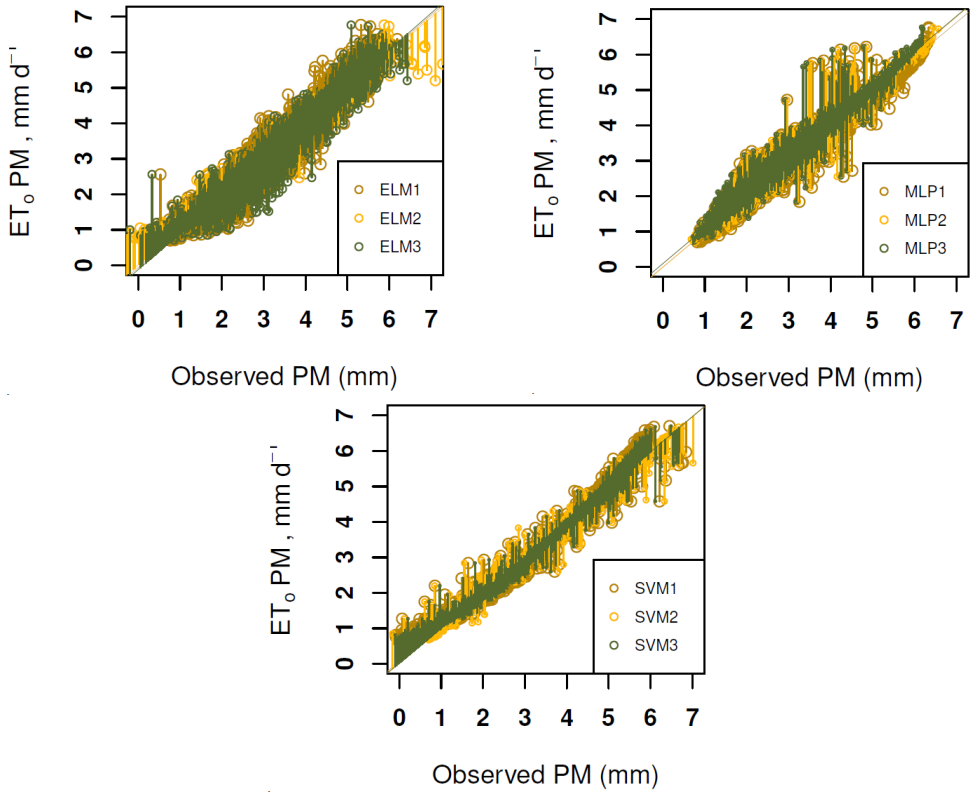

Figure 5.6: Estimated ET$_o$ with ELM, MLP and SVM models using validation dataset

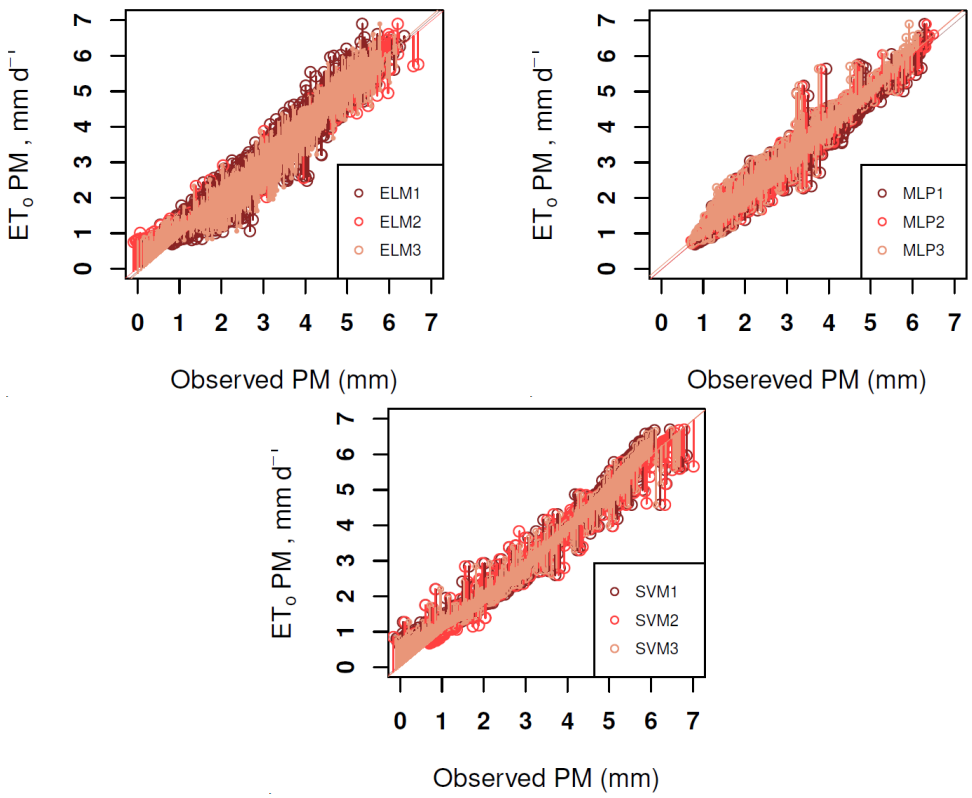

Figure 5.7: Estimated ET_o with ELM, MLP and SVM models using testing dataset

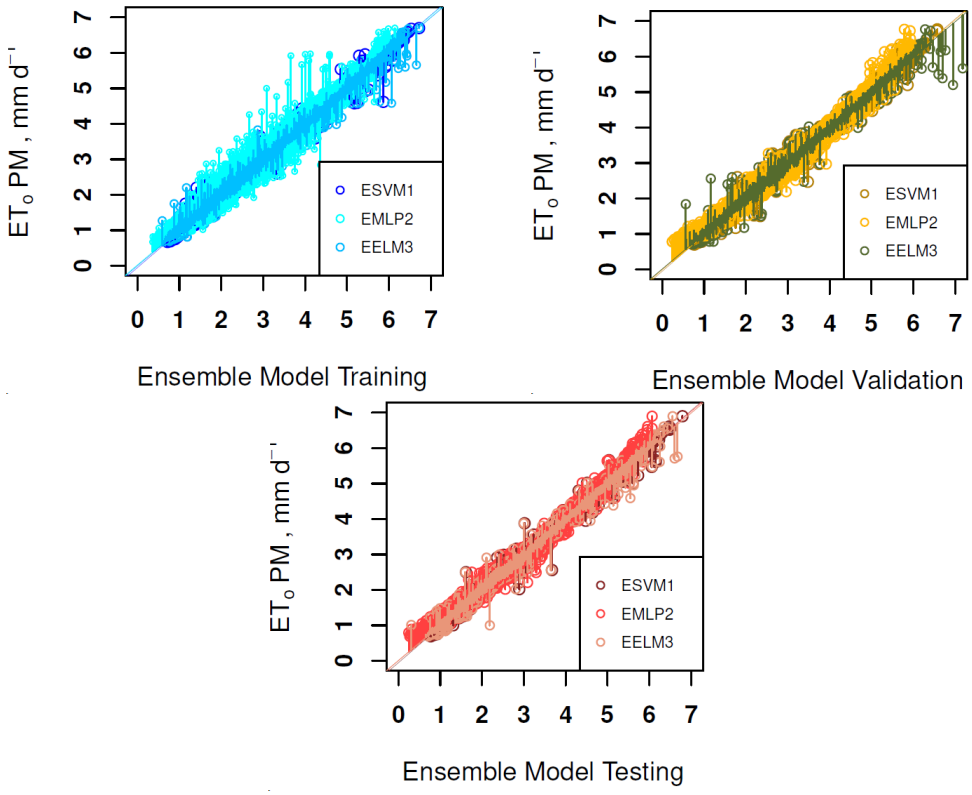

Figure 5.8: Estimated ET_o with proposed ensemble models using training, validation and testing datasets

Table 5.3: Performance comparison of selected models for Patiala with RMSE, NRMSE, NSE and MAE

Models	Training				Validation				Testing			
	RMSE	NRMSE	NSE	MAE	RMSE	NRMSE	NSE	MAE	RMSE	NRMSE	NSE	MAE
ELM 1	0.3485	23.40%	0.95	0.2666	0.3391	22.8%	0.95	0.2625	0.3448	23.0%	0.95	0.2662
ELM 2	0.1936	13.00%	0.98	0.1428	0.2404	13.70%	0.97	0.1451	0.2012	13.40%	0.98	0.1450
ELM 3	0.3894	26.10%	0.93	0.3117	0.3909	26.30%	0.93	0.3118	0.3786	25.2%	0.94	0.3016
MLP 1	0.2289	15.30%	0.98	0.1628	0.2471	16.60%	0.97	0.1698	0.2275	15.10%	0.98	0.1668
MLP 2	0.2203	14.80%	0.98	0.1542	0.2361	15.90%	0.97	0.1586	0.2199	14.60%	0.98	0.1569
MLP 3	0.2451	16.40%	0.97	0.1729	0.2361	15.90%	0.97	0.1586	0.2199	14.60%	0.98	0.1569
SVM 1	0.1979	13.30%	0.98	0.1381	0.2117	14.30%	0.98	0.1411	0.2068	13.80%	0.98	0.1403
SVM 2	0.1015	6.80%	0.98	0.0374	0.1270	8.60%	0.99	0.0416	0.1156	7.70%	0.99	0.0401
SVM 3	0.1945	13.00%	0.98	0.1431	0.2059	13.90%	0.98	0.1457	0.2027	13.50%	0.98	0.1453
En_ELM 1	0.1043	7.00%	0.99	0.0673	0.1184	8.00%	0.99	0.0702	0.1156	7.70%	0.99	0.0708
En_MLP 2	0.2433	16.30%	0.97	0.1808	0.1835	12.40%	0.98	0.1453	0.1835	12.20%	0.99	0.1452
En_SVM 3	0.0085	5.70%	0.99	0.0614	0.0919	6.20%	0.99	0.0631	0.0935	6.20%	0.99	0.0639

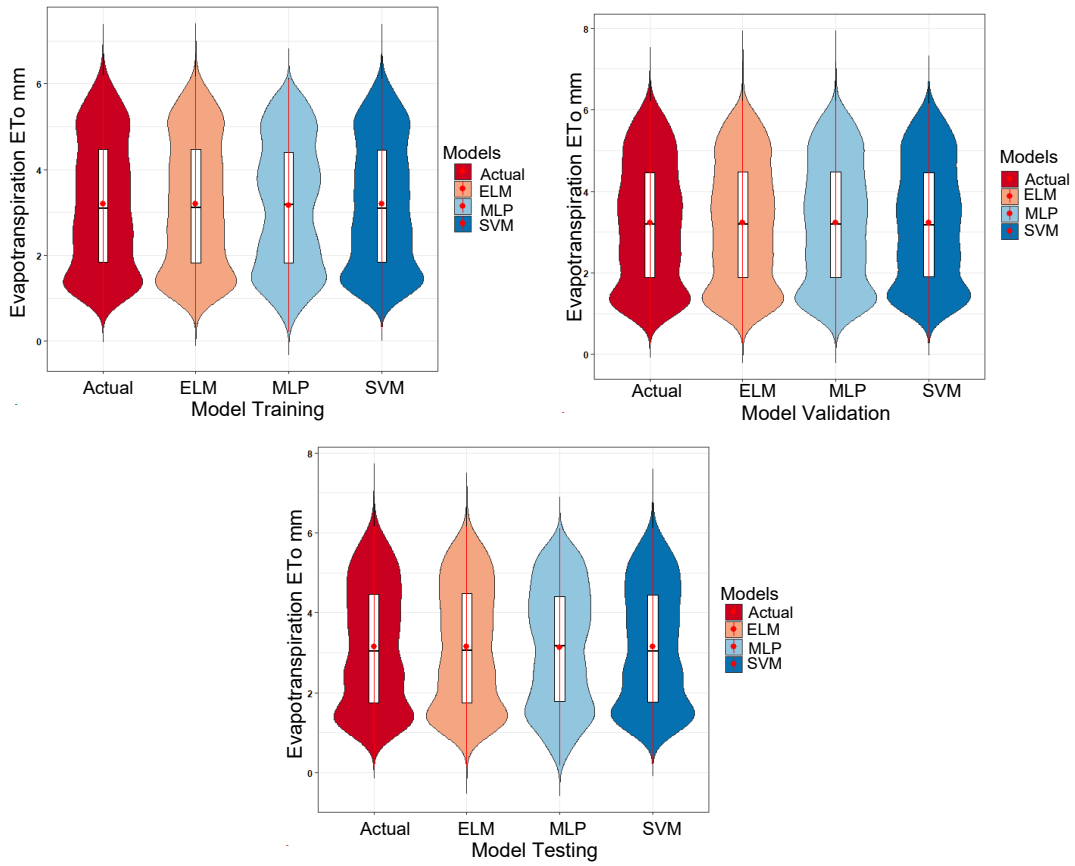

Figure 5.9: Comparison of forecasted ET_o results with ensemble models using training, validation, and testing datasets

In this research work, the performance of the twelve machine learning models of ET_o estimation is given in Fig. 5.9. Out of these twelve models the all ensemble three models generate the best performance respect to accuracy and R^2. The predicted accuracy of an ensemble model is analyzed and it improved the accuracy dramatically.

Fig. 5.10 shows the $Ensemble_{SVM}$ model gives the best accuracy results (MSE= 0.0073, 0.0084 and 0.0087, RMSE= 0.0085, 0.0919 and 0.0935, ACC= 99.7, 99.5 and 99.4, NRMSE= 5.70%, 6.20% and 6.20%, NSE= 0.99, 0.99, and 0.99) for (50%, 30%, and 20%) splitting of training, validation and testing datasets respectively. Table 5.3 also indicated that ELM_3 model provided a slightly better than ELM_1 and ELM_2 model, because both models were shown to be over-fitting in the training dataset. In the case of MLP, all the three models, MLP_1, MLP_2, and MLP_3 presented the better performance in case of the testing as compared to the validation dataset.

Measured daily ET_o performance comparison from 1978-1999 to 2007-2016 for Actual, ELM, MLP and SVM of training, validation and testing datasets of Patiala are demonstrated in Fig. 5.11. The comparison result shows that a $Ensemble_{ELM}$ model achieved the very similar estimation of ET_o values using ($actual_{Ensemble_{ELm}}$ = 22458.5 mm and $estimated_{Ensemble_{ELM}}$ = 22455.7 mm) of training and ($actual_{Ensemble_{Elm}}$

= 8705.6 and estimated$_{Ensemble_{ELM}}$ = 8711.4) of testing, and (actual$_{Ensemble_{MLP}}$ = 13223.9 and estimated$_{Ensemble_{MLP}}$ = 13225.2) of validation datasets respectively. Comparison of model results show that the Ensemble SVM model is more successful in the Patiala station using training, validation and testing scenario. SVM has shown great generalization ability, high training efficiency, and optimal solution. The ensemble-based ELM model presented the high accuracy for the prediction of daily ET_o.

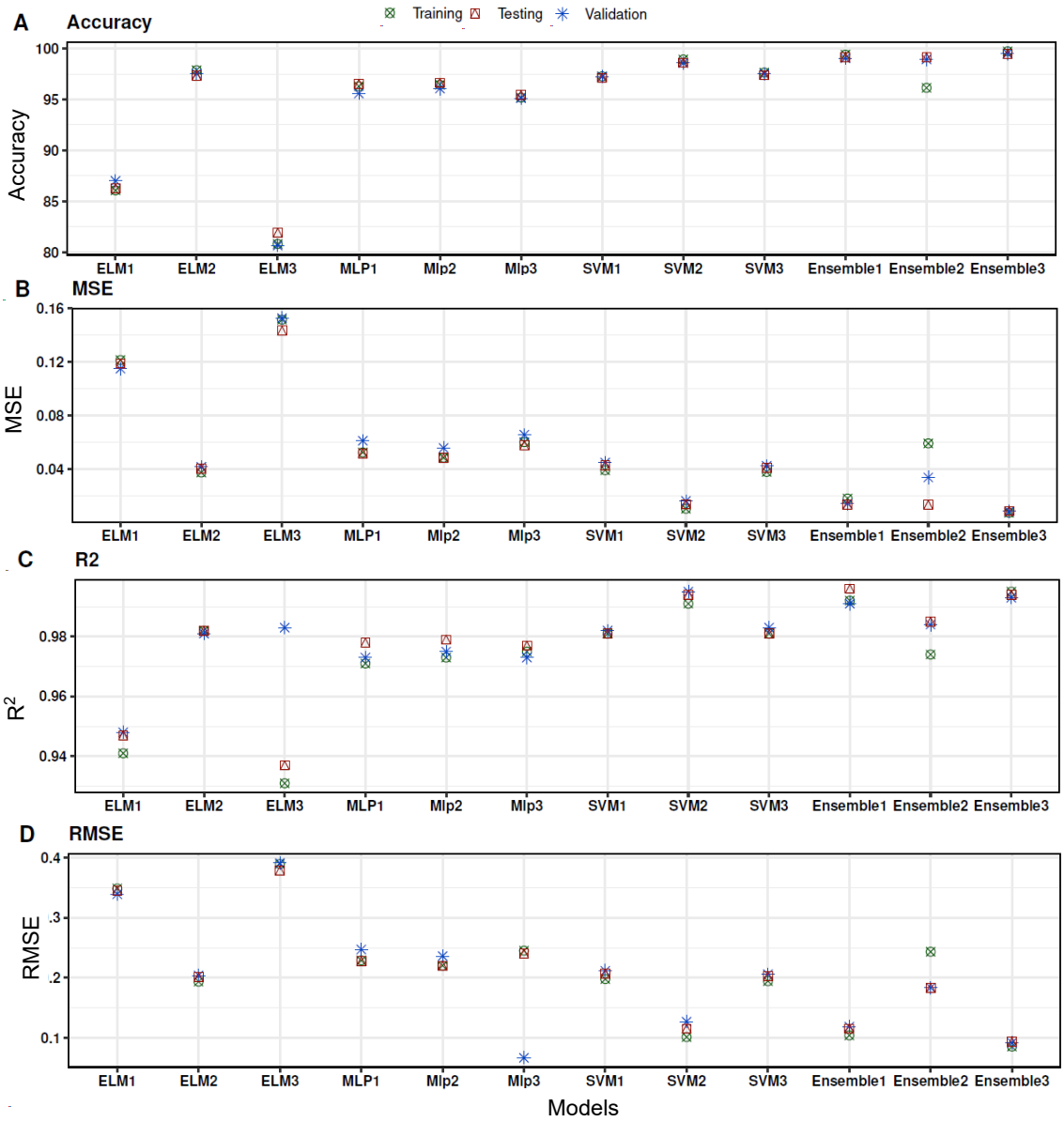

Figure 5.10: Comparison of forecasted ET_o results with ensemble models using ACC, MSE, R^2, and RMSE

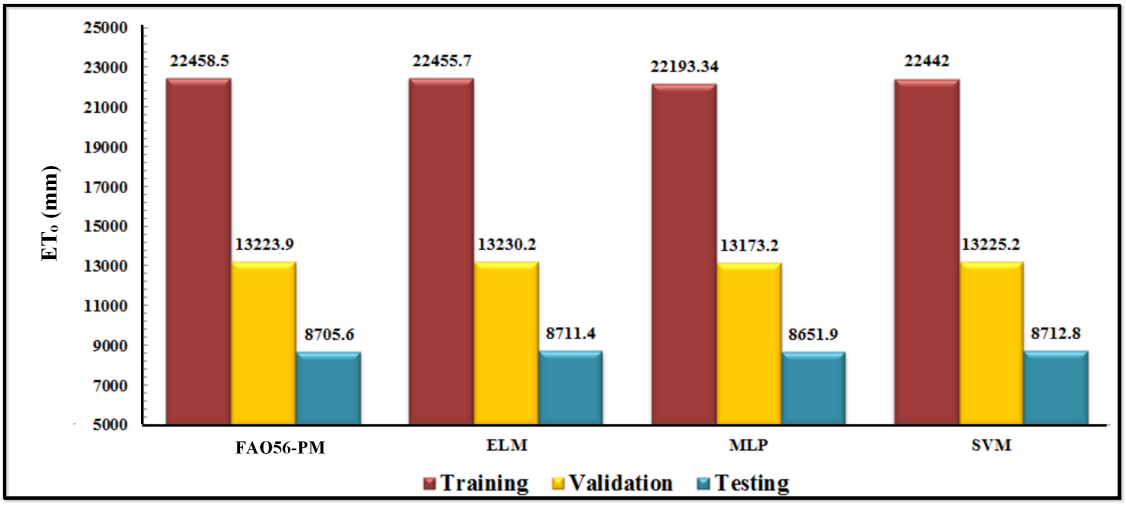

Figure 5.11: Approximation values of observed and estimated ET_o values with ELM, MLP and SVM

5.7 Summary

In this study, we have proposed a multilevel framework to determine the daily ET_o for Patiala station (India). Results indicate that the proposed $Ensemble_{ELM}$ and $Ensemble_{SVM}$ significantly outperforms than other models. Although, the best results are obtained on training, validation and testing datasets without over-fitting using $Ensemble_{MLP}$. Among the nine forecasting models of the first layer performance of SVM_2, SVM_3 and ELM_2 have been comparability better to predict the ET_o.

The following conclusions are drawn from the investigations-

- Multi-level ensembling ET_o forecasting model (MLE-ET_o) has been developed on the basis of the twelve machine learning models such as (ELM_1,.., ELM $_3$), (MLP_1,..,MLP_3), (SVM_1,..,SVM_3), and ($Ensemble_{ELM4}$, $Ensemble_{MLP4}$ and $Ensemble_{SVM4}$).

- The performance of developed models is evaluated over seven statistical error parameters. Through experimental results, it is concluded that classification or regression accuracy can be improved by blending approach from multiple homogeneous models into one using an ensemble approach.

- Comparison of model results show that the Ensemble SVM model is more successful in Patiala station using training, validation and testing scenario. SVM is good at improving generalization, ability and the estimated ET_o generally fit the observed ET_o very well.

- The $Ensemble_{SVM}$ model gives the best accuracy results (MSE= 0.0073, 0.0084 and 0.0087, RMSE= 0.0085, 0.0919 and 0.0935, ACC= 99.7, 99.5 and 99.4,

NRMSE= 5.70%, 6.20% and 6.20%, NSE= 0.99, 0.99, and 0.99) for (50%, 30%, and 20%) splitting of training, validation and testing datasets respectively.

- Among the nine forecasting models of the first layer performance of SVM_2, SVM_3 and ELM_2 have been better comparatively.

Ensemble models showed great applicability in modeling of ET_o, and can be highly recommended for estimating reference crop evapotranspiration (ET_c) for Punjab in future work.

Chapter 6

Crop Evapotranspiration ET_C using FG-RRF

The main objective of this chapter is to design and develop an innovative multilevel model ensembling for accurate estimation of crop coefficient (K_c) and crop evapotranspiration (ET_c) using Fuzzy-Genetic (FG) and Regularization Random Forest (RRF) models.

6.1 Overview

Smart water management has played a significant role in Decision Support System (DSS) to maximize the yield with minimum consumption of water. Accurate estimation of evapotranspiration (ET_o) can provide a scientific basis for developing Irrigation Scheduling. Reference evapotranspiration (ET_o) is the combination of two separate processes whereby water loss from the soil by evaporation and on the other hand water loss from the crop by transpiration.

The main objective of this study is to determine the crop coefficients (K_c) and crop evapotranspiration (ET_c) for Wheat and Maize grown in the Ludhiana district of Punjab, India using ensembling method, such as the Fuzzy-Genetic and regularization random forest models. Wheat and Maize are the most commonly cultivated crops in the region of Punjab, India and have high water consumption. The problems in irrigation sector in India include low irrigation efficiency (30–35%), water-logging, inefficient water supply, and soil salinity. To avoid such negative impact, proper policies and measures are needed to deployed for the quantification of crop growth as well as crop-water balance.

The (K_c) is introduced [36] and several researchers have enhanced the same in the future work [37] [38]. The K_c has the ability to estimate the ET_c. The (K_c) value is sensitive and depends on several aspects such as type of crop, weather parameter, canopy cover density, growth stage, soil moisture and agriculture operations. According to the FAO methodology, the four growing stages of a crop are the initial

stage, crop development stage, mid-season stage and end-season stage. The crop evapotranspiration ET_c is calculated by multiplying a crop coefficient (K_c) with the reference evapotranspiration ET_o. The crop evapotranspiration ET_c method can be expressed by Eq. (1.2).

Two main strategies can be used to improve the use of water in the agriculture sector: (i) Upgrading the irrigation planing and management. (ii) DSS based on artificial and machine learning technology.

The correct estimation of daily ET_c is used to evaluate how much water is required to increase the crop yield, and to decrease the cost. Several research have been carried out in recent years to investigate the (ET_c) using machine learning [119], deep learning [112] and evolutionary algorithms [116]. Wheat is a major cereal crop of India after Rice and it provides more protein than any other cereal crops. India is the region's largest Wheat producer, with 30 million hectares of Wheat area accounting for 107 million metric tons (MMT) of grain [211].

Ludhiana is one of the major Wheat producing district and contributes 12% of total Wheat produced by Punjab [212]. Weather factors such as temperature, rainfall and solar radiation are important for Wheat production. Importance of change in temperature in Wheat growing areas of India has been highlighted [213].

In addition to the improvements in ET_c equations based on reducing the dependency of climatic data, artificial intelligence approaches have been introduced to develop ET_c models in a new pattern. Recently, machine learning models are found to show excellent reliability in Evapotranspiration (ET) estimation and modeling. There are variety of machine learning models based on prediction for reference crop evapotranspiration. However, to accurately separate and estimate the contribution of weather and crop changes on ET variation is still uncertain.

There is a great need to modernize agricultural practices for better water productivity and resource conservation. Any changes in meteorological variables due to climate change will affect evapotranspiration, crop water requirement, and eventually affect water allocation for agriculture and food production [214].

In this chapter, the developed model is implemented on the basis of two factors. Firstly, a crop coefficient (K_c) can be predicted with (Initial, Development, Middle, and Late) stages for Maize and Wheat crops using weather dataset. Secondly, after estimation of (K_c) coefficient value, further, it can be used to predict the (ET_c) value for a particular crop.

The field experiment and crop data for this study has been collected from the PAU, Ludhiana. Moreover, the model has been calibrated for three selected crops of different varieties including Maize (PMH-2) sown on 6^{th} Feb, $Wheat_1$ (PBW-621) sown on 12^{th} Dec and $Wheat_2$ (PBW-502) sown on 25^{th} Oct categorized as timely crops cultivation.

6.2 Fuzzy-Genetic Model

The Fuzzy-Genetic is a machine learning algorithm to develop the prediction model based on Fuzzy Logic (FL). The concept of this model is explained in detail as follows: Fuzzy-Logic (FL), Genetic Algorithm (GA), and Fuzzy-Genetic Algorithm (FGA).

Fuzzy-Logic

"*Fuzzy Logic is a set of mathematical principles for knowledge representation based on degree of membership.*" Fuzzy logic is a soft computing approach which simultaneously manage numerical data and linguistic knowledge. FG is used to get the fuzzy set of values between the crisp set of values. The fuzzy set of values contains the range of possibilities either True or False. Fuzzy control is based on fuzzy sets, fuzzy logic, and fuzzy inference. The four conceptual components of fuzzy control system are defined as follows: Fuzzification phase, Inference engine, Defuzzification phase, and Knowledge base.

Genetic Algorithm

"*Genetic algorithm is a heuristic search and optimization technique to mimic the process of natural evolution.*" GA is used to generate useful solutions (find the maximum or minimum of a function). The principal of GAs have been developed by John Holland and has since been tried on various optimization problems with a high degree of success. Basic concepts of genetic algorithm are as follows:

- Population: Set of possible individual/solutions to the given problem.
- Chromosomes: A set of solution to the given problem.
- Gene: A gene is one element position of chromosome.
- Fitness: A fitness function for optimization.

Genetic algorithm initially starts with a random generation of population, and then, the "selection", "crossover", and "mutation" are preceded until the maximum generation is reached. The basic genetic algorithm involves three types of operators: Selection, Crossover and Mutation.

The Fuzzy-Genetic Algorithm

GAs as an optimization process for finding the optimum parameters of the fuzzy system designed to predict the reference evapotranspiration ET_c. The hybridization

of the GA and fuzzy logic gives an advanced soft computing method called the Genetic-Fuzzy System (GFS), in which a GA is used to evolve a fuzzy system by tuning fuzzy membership functions and learning fuzzy rules.

The genetic algorithm generates a random population of the fuzzy system. At each generation, all the fuzzy systems are tested and their predictions are then compared with the labels and a "performance" is given at each system. The top best system elitism are taken without modification for the next generation. The population is used to generate the population for the next generation using crossover and mutation. Finally, the fuzzy system with the best performance is obtained. The flowchart of Fuzzy-Genetic system is presented in Fig. 6.1.

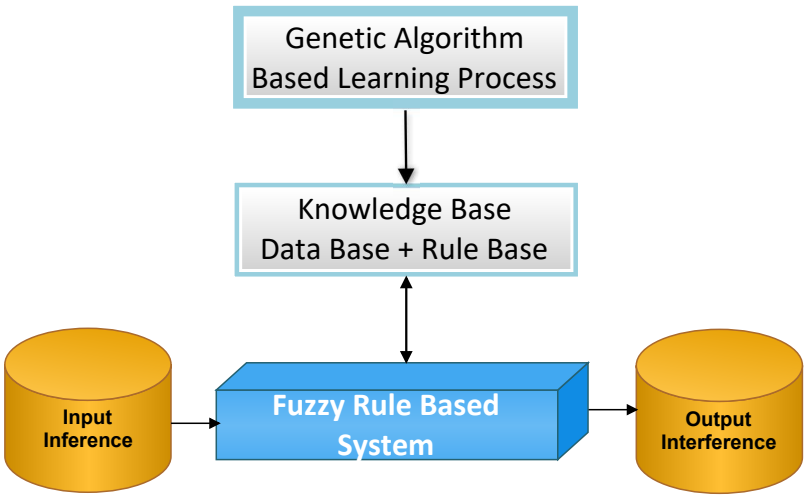

Figure 6.1: Block diagram of fuzzy-genetic system

From the optimization point of view, to find an appropriate fuzzy model is equivalent to code it as a parameter structure and then to find the parameter values that gives us the optimum results for a concrete fitness function. Therefore, the first step in designing a GFS is to decide which parts of the fuzzy system are subjected to optimization by the GA coding them into chromosomes.

6.3 Regularized Random Forest (RRF)

The RRF model has been known by its two primary reasons, firstly RRFs are well-known to provide high accuracies that are competitive with state-of-the-art across a wide range of classification or regression problems. Secondly, RRFs are relatively efficient to train, which is important to our study, since we must train one RRF for each possible subset of features (up to some maximum size). The regularization framework is applied to the random forest and boosted trees, and can be easily applied to other tree models.

Random forest (RF) is a well-established supervised machine learning algorithm for classification and regression problem. There are several techniques to construct the ensemble, for e.g, bagging, and boosting etc. In this experiment the performance of RRF tree is estimated as ensemble model. RRF is a recent advance version of random forest (RF), that applies a regularization framework to random forests that incorporates into the tree growing algorithm [215]. Regularization usually involves the additional penalty to a loss function in order to prevent over-fitting.

6.4 Proposed Model

The proposed model consists of five different stages, i.e. Data collection, Data pre-processing, Modeling, Multilevel modeling, and Estimated output presented in Fig. 6.3.

6.4.1 Data Collection and Pre-processing

The historical weather dataset is collected from IMD and case study of crop coefficient K_c values of three crops such as Maize, $Wheat_1$, $Wheat_2$ from PAU, Ludhiana. Fig. 6.2 shows the map of the study area in Ludhiana. In the Data pre-processing stage, the reference evapotranspiration ET_o and solar radiation R_s parameters are estimated using CROPWAT 8.0 software [216].

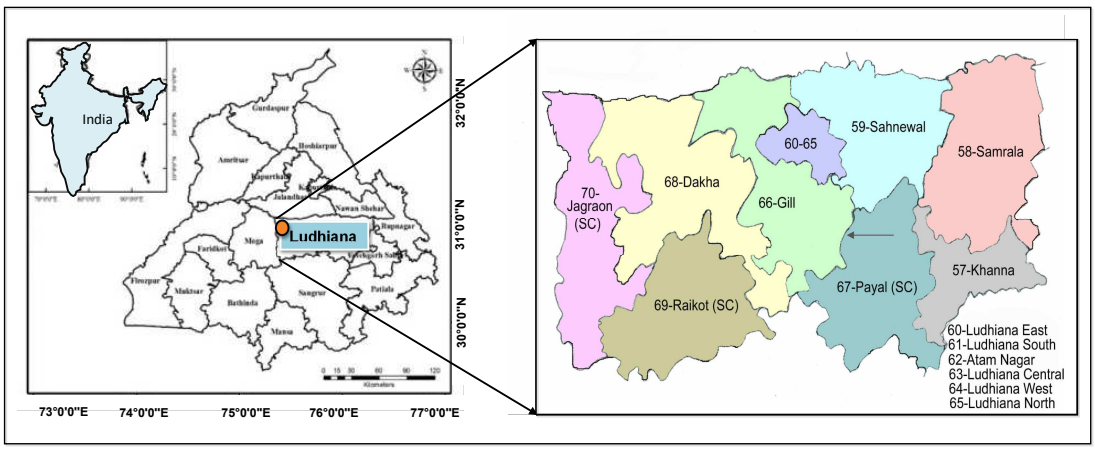

Figure 6.2: Location of study area in Ludhiana

The study is conducted for Ludhiana District which is one of the centrally located District of Punjab (India) with 3706 sq km geographical area. The topography of the study area is a typical representation of an alluvial plain. The climate of the area is sub-tropical steppe, semi-arid with hot dry summers from April to June, hot and humid monsoon period from July to October, winters from November to January and mild climate during February to March. During winter, the average minimum

temperature range between 2-18 (°C) and the average maximum temperature goes up to 16-47 (°C) in summer. The mean relative humidity is lowest during summer and ranges between 30-60%. However, it is highest during monsoon and generally rises up to 90% in July and August. The average annual rainfall varies from 500-650 mm, which is received during the monsoon season from July to September and about 70 to 80 mm rainfall is received during winter months [217].

The crops are chosen to perform the analysis during (Feb 2012 to April 2012) for Maize and two Wheat growing seasons (Dec 2012-April 2013 and Oct 2013-March 2014) at the research farms of Punjab Agricultural University (PAU), Ludhiana.

The weather data has been obtained from the meteorological observatory, PAU, located 200 m from the experimental site. For model training, the daily climatic data is collected from School of Climate Change and Agro-meteorology, Punjab Agriculture University (PAU), Ludhiana. For model testing, weather data is collected from the India Meteorological Department (IMD), Pune (India).

Table 6.1: Statistical parameters of available meteorological variables and ET_o of Ludhiana station

Year	Months	T_{max}	T_{min}	Rh_{max}	Rh_{min}	u_2	I_s	R_s	ET_o
2012	Feb	20.98	8.16	70.71	42.58	4.21	7.43	15.16	1.923
	March	27.93	12.94	59.39	31.71	4.71	7.17	17.32	2.714
	April	33.92	18.58	54.57	30.40	4.70	8.64	24.94	4.735
	Dec	19.43	7.44	91.35	58.81	3.85	5.17	9.70	2.081
2013	Jan	17.04	5.12	94.48	60.61	3.23	5.19	10.36	1.834
	Feb	20.54	9.62	97.11	67.54	3.79	6.38	13.60	2.340
	March	27.63	13.17	94.26	50.45	3.46	9.27	19.89	4.398
	April	31.22	15.66	54.00	27.80	3.82	10.87	23.80	6.728
	Oct	28.80	13.30	75.29	44.14	1.86	8.61	15.91	2.284
	Nov	27.16	10.89	82.23	44.37	1.67	8.83	14.63	1.890
	Dec	20.44	8.27	85.29	59.94	3.10	6.03	11.05	1.204
2014	Jan	16.14	7.10	94.48	72.90	2.77	4.60	9.87	1.123
	Feb	22.75	10.28	88.57	57.14	4.25	7.75	15.66	2.118
	March	24.20	12.11	82.59	55.68	3.41	8.10	18.10	2.765

Table 6.1 presents the data of the year when Wheat and Maize are grown under six weather attributes namely maximum air temperature (T_{max}), minimum air temperature (T_{min}), relative humidity (R_H), wind speed (u_2), solar-radiation (R_s), and sunshine hours (I_s). It is also shows the statistical parameters of meteorological variables at Ludhiana site. It also describes the changing trend of max and min temperature, relative humidity, wind speed, sunshine hours, radiation, and ET_o. Table 6.2 presents seasonal variations of meteorological variables during growing seasons of (2012-2014) of crops Maize, $Wheat_1$ and $Wheat_2$ for the training dataset. To present more detailed climatic condition of the study area, and the seasonal trend of ET_o is shown in Figure 6.3 and Table 6.1.

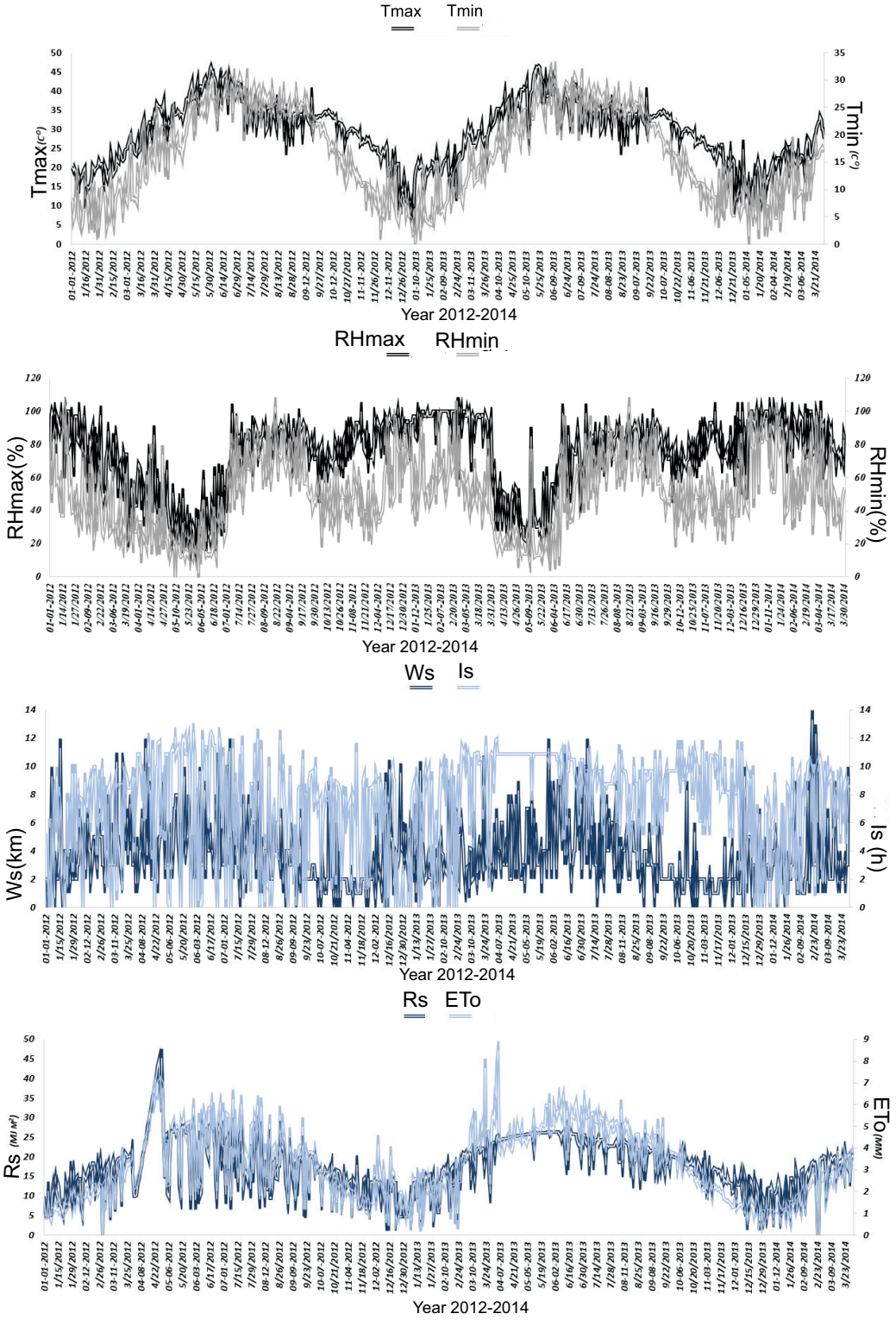

Figure 6.3: Monthly mean variations of daily weather dataset during 2012-2014.

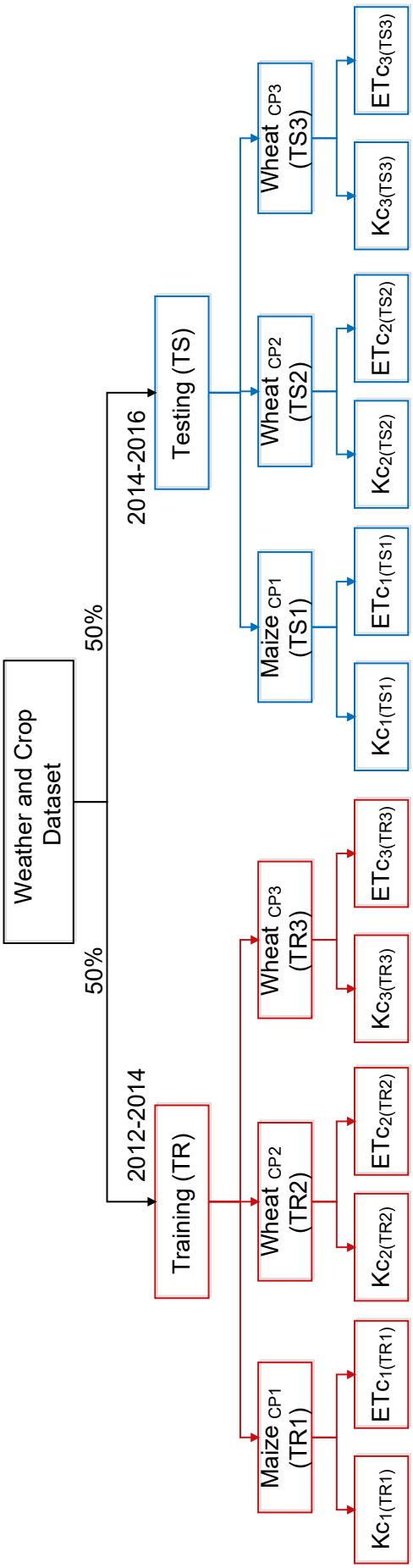

Figure 6.4: Division of dataset

It is observed that in the weather dataset of Maize crop during Feb (2012) to April (2014), the value of ET_o between 0.9 to 7.08 mm day^{-1} value is considered for the training dataset. Similarly, during Feb (2014) to April (2014), ET_o value is between 1.2 to 5.33 mm day^{-1} for testing dataset. For Ludhiana station, the ET_o value 272.3 and 250 mm day^{-1} is considered for training and testing dataset respectively. The weather dataset during Dec (2012) to April (2013), for $Wheat_1$ crop, the ET_o between 0.77 to 7.9 mm day^{-1}, value is considered for the training dataset. Similarly, during Dec (2014) to April (2015) for testing dataset consist the ET_o value between 0.4 to 7.9 mm day^{-1}.

It has been observed that the total amount of water requirement of Ludhiana station ET_o value is found as 356.8 and 308.9 mm day^{-1} using training and testing dataset respectively. However, it is also observed that weather dataset of $wheat_2$ crop during Oct (2015) to March(2016), the ET_o value between 0.9 to 3.01 mm day^{-1} is considered for training dataset. Similarly, during Oct (2015) to March (2016), ET_o from 0.73 to 3.26 mm day^{-1}, value is considered for testing dataset. Ludhiana station for ET_o value 243.4 and 229.3 mm day^{-1} of training and testing dataset respectively.

Fig. 6.4. shows the division of weather and crop dataset training (50%), and testing (50%). $Maize_{CP1}$, $Wheat_{CP2}$, and $Wheat_{CP3}$ the dataset of three crops are considered. These samples are used to predict the two subsets of each sample namely K_c, and ET_c.

6.4.2 Wheat and Maize Crops

The study is conducted for prediction of Maize and Wheat crops co-efficient (K_c) and (ET_c). (K_c) is defined as the ratio of actual crop evapotranspiration (ET_c) to reference evapotranspiration (ET_o). Water requirement of crops is determined by product of (ET_o) and crop coefficient i.e. (K_c). It is useful to determine the water requirement of crops according to their stages and environmental factors.

The K_c values of $Wheat_1$ crop are considered 0.4, 1.15 & 0.4 and for $wheat_2$ are 0.5, 1.36, 1.42 and 0.42 for the initial, mid and last stage of growth respectively. The length of time (days) for initial, development, mid and late seasons for $Wheat_1$ are 29, 55, 14 & 32 days and for $Wheat_2$ 24, 46, 35, 42 days as shown in Table 6.2.

Maize is globally a top ranking cereal in productivity, human food, animal feed and as a source of a large number of industrial products. It is the third most important crop of Kharif season after paddy and cotton in Punjab. The area under Maize in Punjab has declined from 1.65 lakh ha in 2000-01 to 1.27 lakh ha in 2015-16. The K_c values of Maize crop are considered 0.7, 0.85, 1.15 & 1.05 for the initial, mid and end stage of growth respectively.

The length of time (days) for initial, development, mid and late seasons for Maize

Table 6.2: Crop coefficient (Kc) values of different crops at different stages of Ludhiana for the training dataset

Period	Crop	Variety	Initial		Develop		Mid		Late		TD
			K_c	TGD	K_c	TGD	K_c	TGD	K_c	TGD	
06/02/2012-30/04/2012	Maize	PMH-2	0.7	35	0.85	18	1.15	17	1.05	15	85
01/12/2012-05/04/2013	Wheat	PBW621	0.4	29	1.15	55	1.15	14	0.4	32	130
25/10/2013-22/03/2014	Wheat	PBW502	0.5	24	1.36	46	1.42	35	0.42	42	147
Abbreviations represent the following- Crop coefficient: K_c; Total days: TD; Total Growing days: TGD;											

Table 6.3: Crop coefficient (Kc) values of different crops at different stages of Ludhiana for testing dataset

Period	Crop	Variety	Initial		Develop		Mid		Late		TD
			K_c	TGD	K_c	TGD	K_c	TGD	K_c	TGD	
06/02/2014-30/04/2014	Maize	PMH-2	0.7	35	0.85	18	1.15	17	1.05	15	85
01/12/2014-05/04/2015	Wheat	PBW621	0.4	29	1.15	55	1.15	14	0.4	32	130
25/10/2015-22/03/2016	Wheat	PBW502	0.5	24	1.36	46	1.42	35	0.42	42	147
Abbreviations represent the following- Crop coefficient: K_c; Total days: TD; Total Growing days: TGD;											

are 35, 18, 17 and 15 days used in different stage. Detail of selected crops and period of data for the study are shown in Table 6.2 and Table 6.3 using training and testing datasets respectively.

6.4.3 Multi-Ensembling Modeling for K_c and ET_c using FG-RRF

Fuzzy-Genetic model is applied to simulate the K_c and ET_c values using training dataset. Based on the Fuzzy-Genetic model prediction results, both K_c and ET_c prediction probabilities are combined in newly dataset with previous weather dataset. This ensemble dataset is evaluated on the basis of performance metrics to check accuracy of results.

Once all the above stages are successfully completed, the output prediction of each sample is obtained. Then, the obtained results are evaluated by the different performance metrics such as ACC, MSE, RMSE, MAE, NSE, R^2 for each sample.

Performance analysis of the proposed model is depicted in Table 6.4 and Table 6.5.

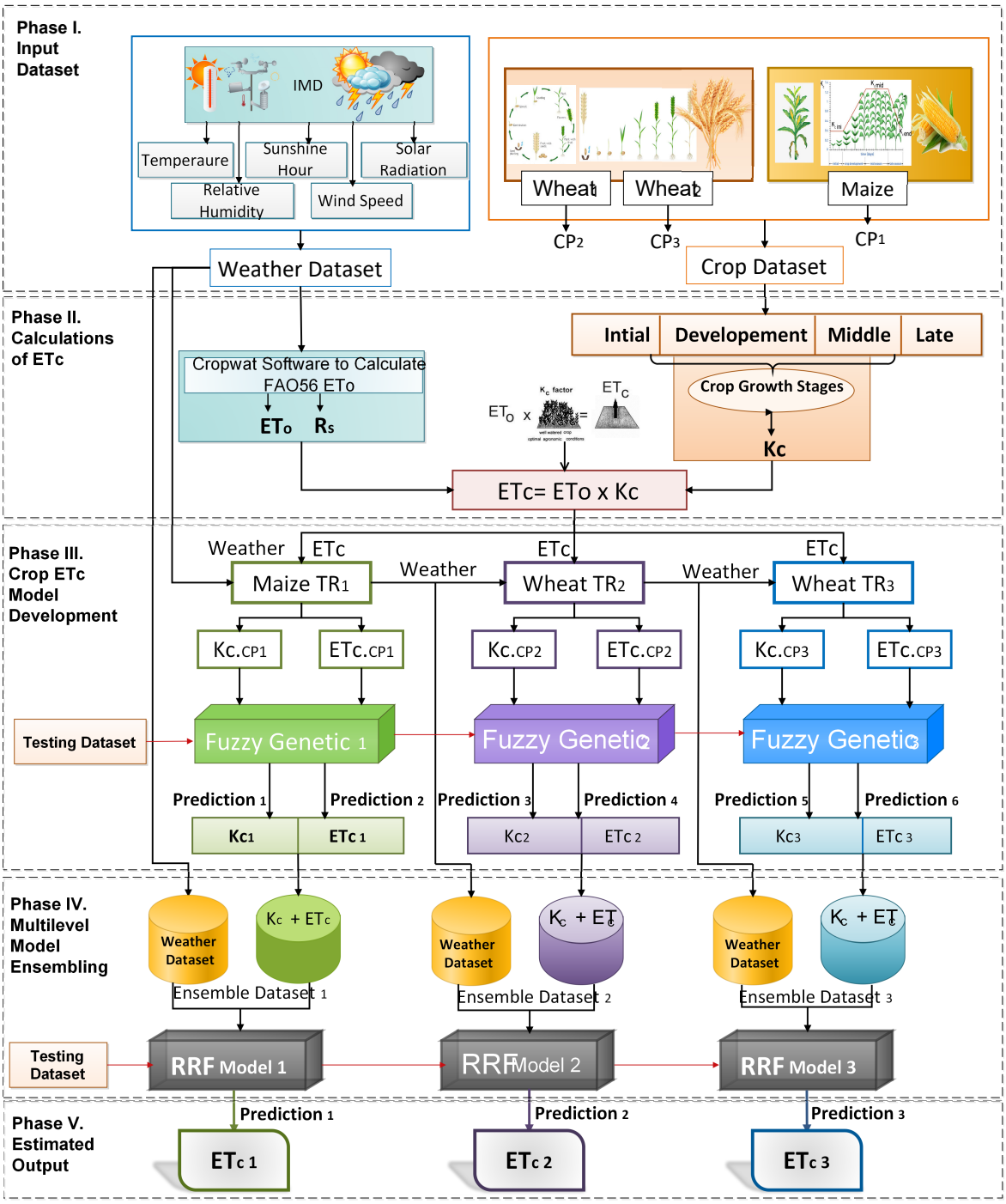

Figure 6.5: Proposed reference crop ET_c model

Then, the ensembling dataset is used to train the RRF model for predicting the ET_c values of each sample of the crop. After getting the best accuracy from training model, the testing dataset is applied to validate the accuracy of model. The Fig 6.5 presents the proposed model for crop ET_c using FG-RRF ensemble model. In this section, the proposed FG-RRF algorithm has demonstrated the process of ensemble modeling by Algorithm 5.1.

Algorithm 5.1: The algorithm steps of proposed model

Input: Weather dataset (T_{max}, T_{min}, Rh_{max}, Rh_{min}, u_2, I_s, Rh_s, ET_o) and Crop dataset ($Maize_{CP1}$), ($Wheat_{CP2}$) and ($Wheat_{CP3}$)

Output: K_c and ET_c values.

Training: ($K_{c1}TR_1$),($K_{c2}TR_2$) and ($K_{c3}TR_3$) for ($Maize_{CP1}$), ($Wheat_{CP2}$) and ($Wheat_{CP3}$) and ($ET_{c1}TR_1$), ($ET_{c2}TR_2$) and ($ET_{c3}TR_3$) for ($Maize_{CP1}TR_1$), ($Wheat_{CP2}TR_2$) and ($Wheat_{CP3}TR_3$).

Testing: ($K_{c1}TS_1$), ($K_{c2}TS_2$) and ($K_{c3}TS_3$) for ($Maize_{CP1}$), ($Wheat_{CP2}$) and ($Wheat_{CP3}$) and ($ET_{c1}TS_1$), ($ET_{c2}TS_2$) and ($ET_{c3}TS_3$) for ($Maize_{CP1}$), ($Wheat_{CP2}$) and ($Wheat_{CP3}$).

Model1: Fuzzy-Genetic$_{M1}$, Fuzzy-Genetic$_{M2}$ and Fuzzy-Genetic$_{M3}$.

Model2: RRF_{M1}, RRF_{M2} and RRF_{M3}.

Ensemble Dataset:
$Ensemble_{ETc1}(TR_1)$,$Ensemble_{ETc2}(TR_2)$,$Ensemble_{ETc3}(TR_3)$ and $Ensemble_{ETc1}(TS_1)$,$Ensemble_{ETc2}(TS_2)$,$Ensemble_{ETc3}(TS_3)$.

Stage1 $\longrightarrow$ **Input Dataset:** Weather dataset and crop dataset

Stage2 $\longrightarrow$ **Data Pre-processing:**

i. **Estimating ET_o, K_c and ET_c:** Estimated ET_o CROPWAT 8.0 from (2012 to 2016) and ET_c calculated by equation (1.2).

ii. **Training and Testing Dataset:** ($Maize_{CP1}$), ($Wheat_{CP2}$) and ($Wheat_{CP3}$).

Stage3 $\longrightarrow$ **Crop ET_c Model Development:**

i. **Model K_c:** Fuzzy-Genetic$_{M1}$, Fuzzy-Genetic$_{M2}$ and Fuzzy-Genetic$_{M3}$ models are trained by ($K_{c1}TR_1$), ($K_{c2}TR_2$) and ($K_{c3}TR_3$).

ii. **Model ET_c:** Fuzzy-Genetic$_{M1}$, Fuzzy-Genetic$_{M2}$ and Fuzzy-Genetic$_{M3}$ models are trained by ($ET_{c1}TR_1$), ($ET_{c2}TR_2$) and ($ET_{c3}TR_3$).

Stage4 $\longrightarrow$ **Multilevel Model Ensembling:**

i **Predictions:** (Fuzzy-Genetic$_{M1,M2,M3}$) models will give prediction probabilities on training dataset of (K_{cPred1}, $_{Pred2}$, $_{Pred3}$)and ($ET_{cPred1,Pred2,Pred3}$) values. Ensemble dataset is combined by predicted (K_{cPred1}, $_{Pred2,Pred3}$) and ($ET_{cPred1,Pred2,Pred3}$).

ii **Ensemble Datasets:** The $Ensemble_{ETc1}(TR_1)$, $Ensemble_{ETc2}(TR_2)$ and $Ensemble_{ETc3}(TR_3)$ datasets are used to train the RRF_{M1}, RRF_{M2} and RRF_{M3} models for predicting the Final $ETc1$, $ETc2$ and $ETc3$. For validation purpose, the testing datasets are used to evaluate this model by $Ensemble_{ETc1}(TS_1)$, $Ensemble_{ETc2}(TS_2)$ and $Ensemble_{ETc3}(TS_3)$

Stage5 $\longrightarrow$ **Estimated Output:** The comparison of the proposed model is shown in results. The proposed fuzzy-genetic and RRF model makes effective decisions using the multilevel ensembling for the each crop sample.

6.5 Simulation Results and Discussion

In this section, the experimental results of FG-RRF (ET_c) model for each sample of crops are presented. Statistical measurements have been used to estimate daily performance of K_c and ET_c model.

The Fuzzy-Genetic and Regularized Random Forest (FG-RRF) proposed model is implemented using R-studio version 3.5.2 to simulate the reference crop coefficient K_c and ET_c. The package "fugeR" implements genetic algorithms to construct an FRBS from numerical data for classification [218]. The package fugeR is designed for training fuzzy systems based on evolutionary algorithms. In this package, there are two main functions which are fugeR.run() for construction of the FRBS model and fugeR.predict() for prediction. For the development of model, we have set maximum number of rules as 5, maximum variables per rule set as 4 and a number of generations considered are 100 to 150, and the population as 200. The elitist parameter was set to be 20% out of every generation. The regularized random forest model is implemented with 'caret' and 'RRF' packages.

6.5.1 ET_c of Maize

Crop coefficient values K_c are taken from PAU Ludhiana [219]. It is observed that the K_c value of Maize crop during (Feb 2012-April 2012) is varying from (0.7, 0.83, 1.15, 1.05) and (0.7, 0.83, 1.15, 1.05) during (Feb 2014-April 2014) for training and testing dataset respectively. Similarly, parameter metrics results have also reported that the Fuzzy-Genetic model provided a slightly better accuracy results in training dataset than testing results for K_c value. However, Table 6.4 shows the Fuzzy-genetic model results (MSE= 0.0003 and 0.0134, RMSE= 0.055 and 0.1160, ACC= 100% and 100%) for splitting of training and testing datasets in ratio (50% and 50%).

During the initial stage, the mean water requirement ET_c varies in 1.27 to 2.15 mm day^{-1} for Maize using Fuzzy-Genetic model. However, there is a slight variation observed with RRF model i.e. 1.22 to 2.05 mm day^{-1}. During the developmental stage, ET_c increases and as the ratio between 2.2 to 2.73 mm day^{-1}, whereas the RRF model varies between 1.99 to 2.85 mm day^{-1}. During the mid season, stage mean water requirement also increases and varies between 2.81 to 5.27 mm day^{-1} whereas across the RRF model it varies between 2.29 to 5.27 mm day^{-1}. During the late-season stage ET_c increases progressively up to end of crop season using Fuzzy-Genetic (5.27 to 7.05) mm day^{-1} and (5.11 to 6.79) mm day^{-1} for RRF model.

Figure 6.6 shows the comparison of observed ET_c and predicted ET_c for Fuzzy-Genetic and RRF models using training and testing datasets. The observed K_c (Blue) and predicted K_c (orange and red) color are also presented in training and testing datasets for Maize crop. It is observed that Fuzzy-Genetic and RRF model

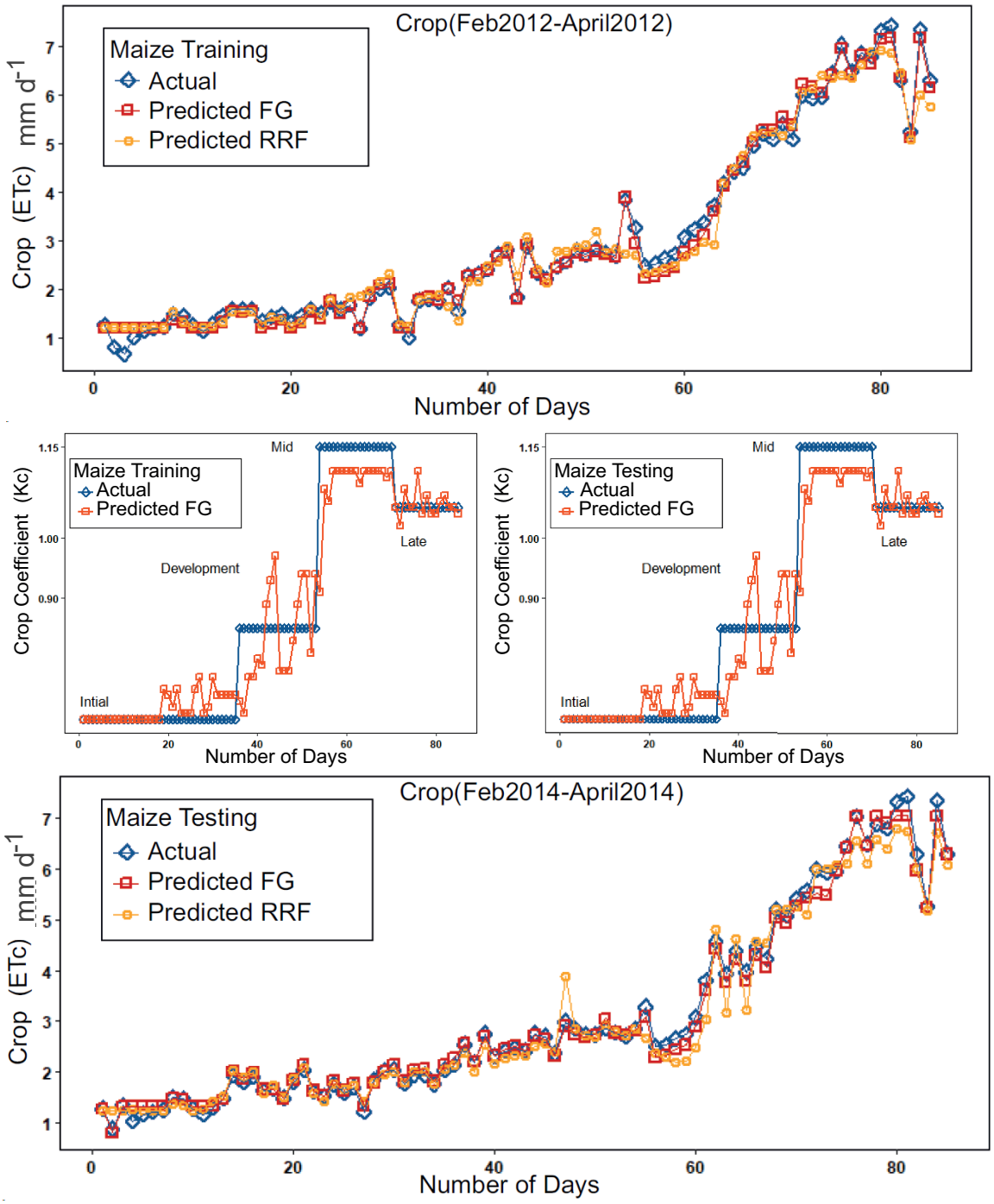

Figure 6.6: Estimated K_c and ET_c values for actual and predicted of maize crop

presented the best performance for testing dataset predicted ET_c in terms of MSE (0.0240 and 0.0711) mm day^{-1} and MSE of (0.0209 and 0.069) mm day^{-1}.

Results of ET_c with RMSE of (0.1530 and 0.2670) mm day^{-1}, MAE of (0.1130 and 0.1580) mm day^{-1}, ACC of (98% and 97%) mm day^{-1} using Fuzzy-Genetic and RRF model for the training dataset. The results show that ET_c with RMSE of (0.1447 and 0.2630) mm day^{-1}, MAE of (0.1048 and 0.1870) mm day^{-1}, ACC of (99% and 98%) mm day^{-1} suggesting a good agreement between Fuzzy-Genetic and

RRF model for the testing dataset.

6.5.2 ET_c of $Wheat_1$

It is observed that the K_c value of $Wheat_1$ crop during (Dec 2012-April 2013) and (Dec 2014-April 2015) is varying from 0.4 (stage I), 1.15 (stage II), 1.15 (stage III), and 0.4 (stage IV) for training and testing datasets respectively.

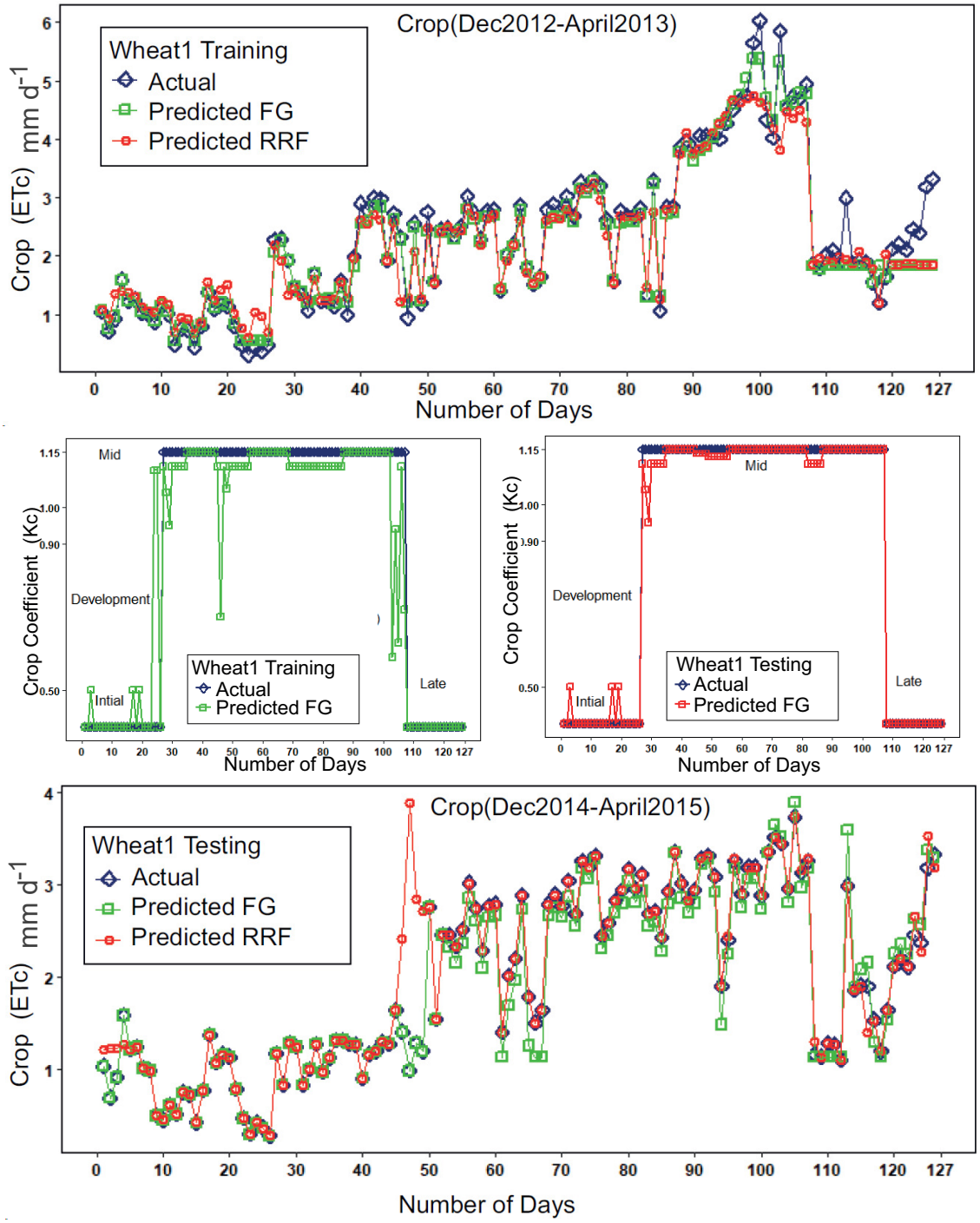

Figure 6.7: Estimated K_c and ET_c values for actual and predicted of $wheat_1$ crop

Table 6.4: Performance comparison of selected models for Patiala with training dataset

	Maize			Wheat1			Wheat2		
Models	Fuzzy-Gentic		RRF	Fuzzy-Gentic		RRF	Fuzzy-Gentic		RRF
Results	ET_c	K_c	ET_c	ET_c	K_c	ET_c	ET_c	K_c	ET_c
MSE	**0.024**	0.003	0.071	0.073	0.064	**0.064**	1.145	0.015	0.267
RMSE	**0.153**	0.055	0.267	0.271	0.253	**0.254**	1.070	0.123	0.516
R2	**0.99**	0.96	0.99	0.98	0.71	**0.98**	0.79	0.97	0.92
MAE	**0.113**	0.040	0.158	0.149	0.151	**0.148**	0.725	0.065	0.355
Acc	**98**	100	97	93	93.1	**97.5**	75	100	85
NSE	**0.993**	0.91	0.98	0.95	0.77	**0.95**	0.97	0.92	0.84

Table 6.5: Performance comparison of selected models for Patiala with testing dataset

	Maize			Wheat1			Wheat2		
Models	Fuzzy-Gentic		RRF	Fuzzy-Gentic		RRF	Fuzzy-Gentic		RRF
Results	ET_c	K_c	ET_c	ET_c	K_c	ET_c	ET_c	K_c	ET_c
MSE	**0.020**	0.0134	0.069	0.124	0.104	**0.156**	**0.012**	0.0257	0.075
RMSE	**0.144**	0.1160	0.263	0.352	0.323	**0.396**	**0.112**	0.1603	0.273
R2	**0.990**	0.990	0.990	0.940	0.830	**0.890**	**0.94**	0.88	0.830
MAE	**0.104**	0.0512	0.187	0.218	0.250	**0.264**	**0.053**	0.0519	0.118
Acc	**99.0**	100.0	98.0	94.0	94.0	**93.0**	99.0	96.0	94.0
NSE	**0.99**	0.59	0.97	0.89	0.75	**0.97**	0.97	0.88	0.68

The comparison of predicted and observed ET_c values using Fuzzy-Genetic and RRF models of training and testing periods are shown in Figure 6.7. Also obtained observed K_c (Blue) and predicted K_c (green and red) color values are presented in training and testing datasets for $Wheat_1$ crop. Comparisons of measured and observed K_c value under Fuzzy-Genetic model, for training periods, have presented best performance than testing dataset using statistical metrics with (MSE= 0.064 and 0.104, RMSE= 0.253 and 0.323, MAE= 0.151 and 0.250).

During the initial stage, mean water requirement ET_c for $Wheat_1$ using Fuzzy-Genetic model is (0.9 to 2.64) mm day^{-1}, however, the stage slight variation is observed with RRF, where it varies between (0.8 and 1.51) mm day^{-1}. During the developmental stage, ET_c increases and it varies between 0.9 to 3.35 mm day^{-1}, whereas across the RRF model it varies between 0.65 to 3.19 mm day^{-1}. During the mid season stage, the mean water requirement also increases and varies between 1.3 to 3.44 mm day^{-1}, whereas across the RRF model it varies between 1.74 to 3.70 mm day^{-1}. During the late-season stage, ET_c increases progressively up to end of crop season using Fuzzy-Genetic (1.11 to 3.55) mm day^{-1} and (1.5 to 1.83) mm day^{-1} for RRF model.

Similar parameter metrics results have also been reported that Fuzzy-Genetic and RRF model provided slightly better accuracy results in training dataset than testing results.

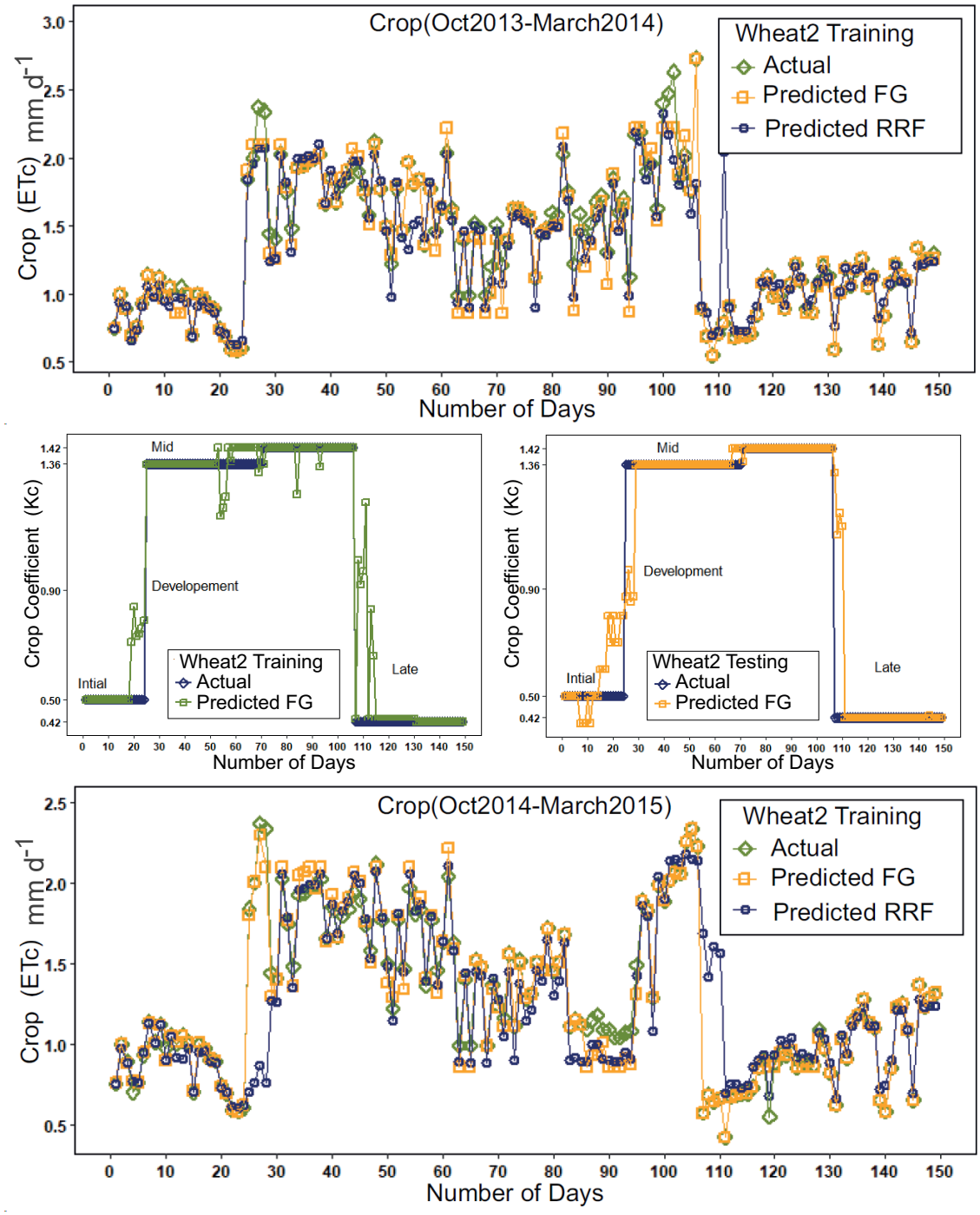

Figure 6.8: Estimated K_c and ET_c values for actual and predicted of $Wheat_2$ crop

However, Table 6.4 shows the Fuzzy-genetic and RRF models with (MSE= 0.073 and 0.0640, RMSE= 0.271 and 0.254, ACC= 93% and 97%, MAE= 0.149 and 0.148) for training dataset. The results show that ET_c with RMSE of (0.352 to 0.396) mm day^{-1}, MSE of (0.124 to 0.156) mm day^{-1}, MAE of (0.218 to 0.264) mm day^{-1}, ACC of (94% and 96%) mm day^{-1} using Fuzzy-Genetic and RRF model during testing dataset. Comparison of observed and predicted K_c and ET_c values for $Wheat_1$ crop using the proposed model for training and testing are presented in Fig. 6.7.

6.5.3 ET_c of $Wheat_2$

It is observed that the K_c value of $Wheat_2$ crop during (Oct 2013-March 2014) and (Oct 2015-March 2016) is varying from 0.5 (stage I), 1.36 (stage II), 1.42 (stage III), and 0.42 (stage IV) for training and testing datasets respectively.

Similar parameter metrics results have also been reported that Fuzzy-Genetic model provided slightly better accuracy results in training dataset than testing results. However, Fig. 6.4 and Fig 6.5 shows the Fuzzy-Genetic model results (MSE= 0.0150 and 0.0720, RMSE= 0.1230 and 0.2680, ACC= 100% and 97%, MAE= 0.065 and 0.1580) for (50% and 50%) splitting of training and testing datasets.

The scatter plots of the observed and predicted ET_c using Fuzzy-Genetic and RRF models of training and testing periods presented in Figure 6.8. The observed K_c (Blue) and predicted K_c (Green and Orange) color values are presented over the training and testing period of $Wheat_2$ crop.

Figure 6.9 and 6.10 indicate the statistical summary of how well patterns match each other in terms of their correlation, their root mean square difference and the ratio of their variances using the Taylor diagram. The colors indicate training and testing dataset results of Maize, $Wheat_1$ and $Wheat_2$ crops observed and predicted values of ET_c and K_c.

Figure 6.11 shows the comparison results of K_c and ET_c using Fuzzy-Genetic and RRF model by a statistical performance with MSE, RMSE, R^2, MAE and ACC.

During the initial stage, mean water requirement ET_c for $Wheat_2$ using Fuzzy-Genetic model is (0.5 to 1.12) mm day^{-1}, however the stage slight variation is observed with RRF, where it varies between (0.6 to 1.12) mm day^{-1}. During the developmental stage, ET_c increases and it varies between 1.36 to 2.24 mm day^{-1}, whereas across the RRF model it varies between 0.6 to 2.11 mm day^{-1}. During the mid season stage, mean water requirement also increases and varies between 1.28 to 2.34 mm day^{-1} whereas across RRF model it varies between 0.8 to 2.18 mm day^{-1}. During the late-season stage ET_c increases progressively up to end of crop season using Fuzzy-Genetic (0.55 to 1.37) mm day^{-1} and (0.6 to 1.23) mm day^{-1} for RRF model. Figure 6.12 shows by the range of ET_c at different stages. The proposed model ET_c is found to be in a low range (between 0 and 2 mm day^{-1}) indicated by blue colour, medium range (between 2 and 4 mm day^{-1}) indicated by red colour, high range (between 4 and 6 mm day^{-1}) indicated by green colour and very high range (greater than 6 mm day^{-1}) indicated by purple colour for training and testing. Moreover, it indicates that the crop evapotranspiration ET_c is divided in to four categories such as (low, medium, high and very-high) for further analysis over training and testing periods.

The results show that ET_c with RMSE of (1.070 to 0.516) mm day^{-1}, MSE of

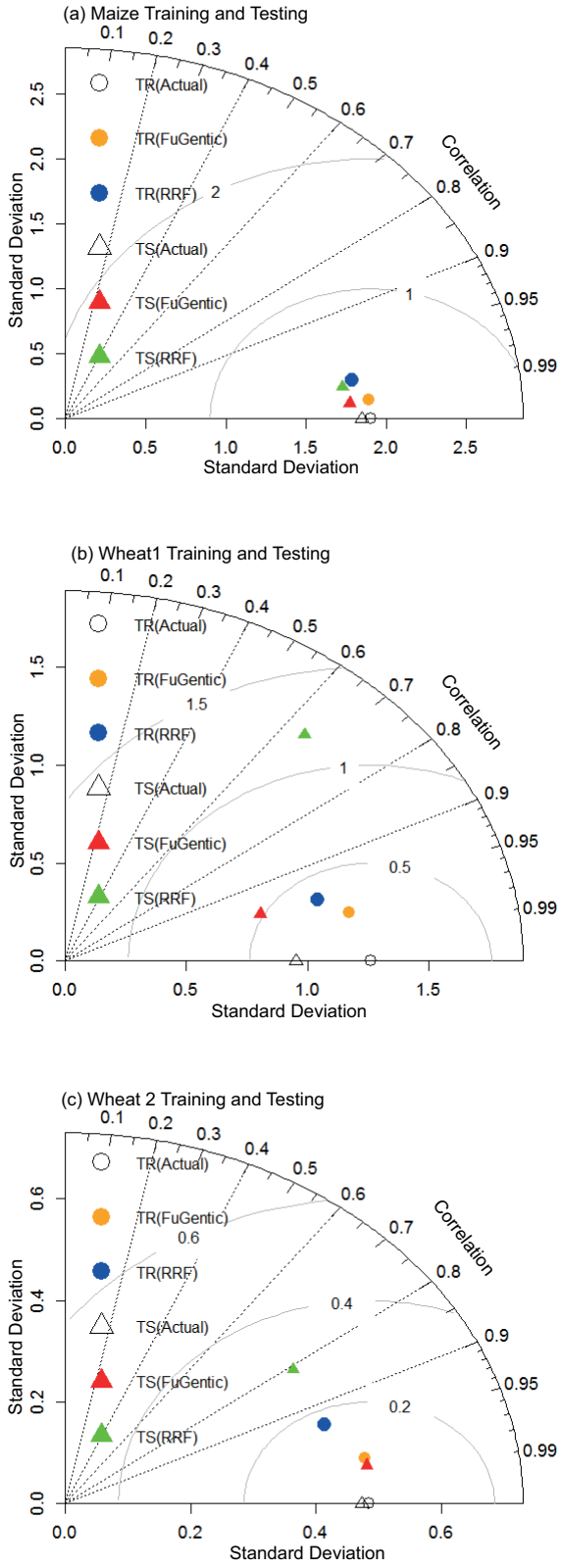

Figure 6.9: Taylor diagram representation of predicted and actual ET_c

(1.145 to 0.2670) mm day^{-1}, MAE of (0.7250 to 0.3550) mm day^{-1}, ACC of (75% and 85%) mm day^{-1} using Fuzzy-Genetic and RRF model for the training dataset.

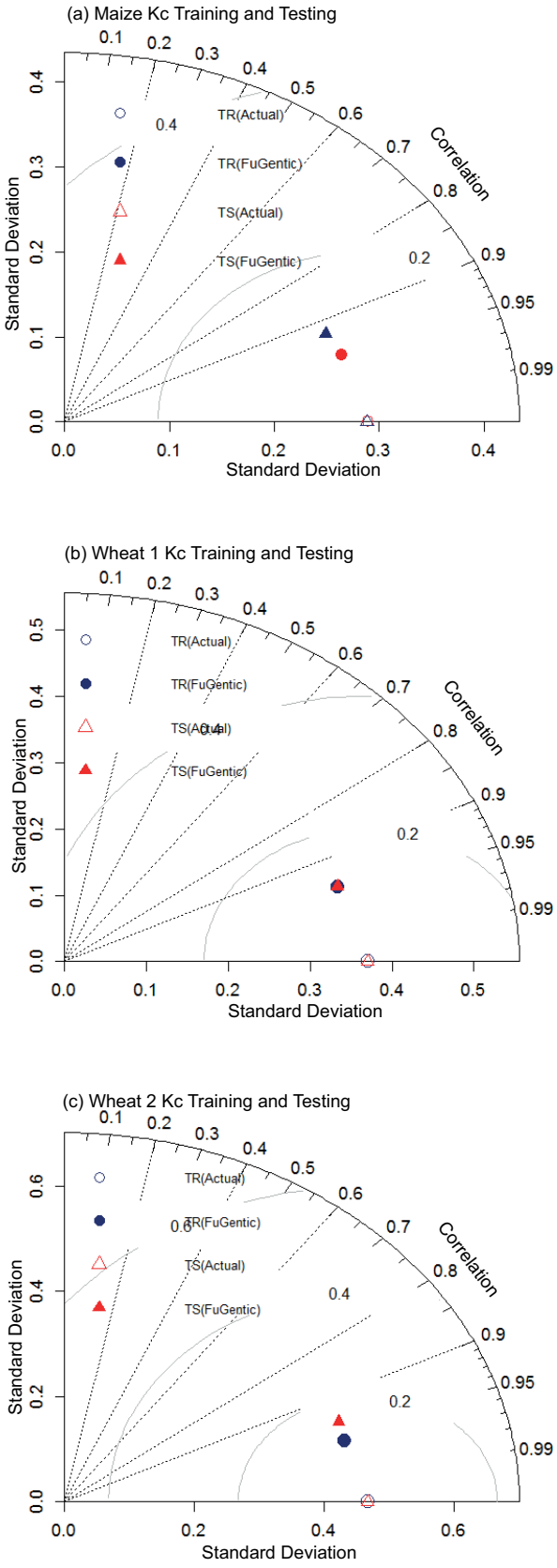

Figure 6.10: Taylor diagram representation of predicted and actual K_c

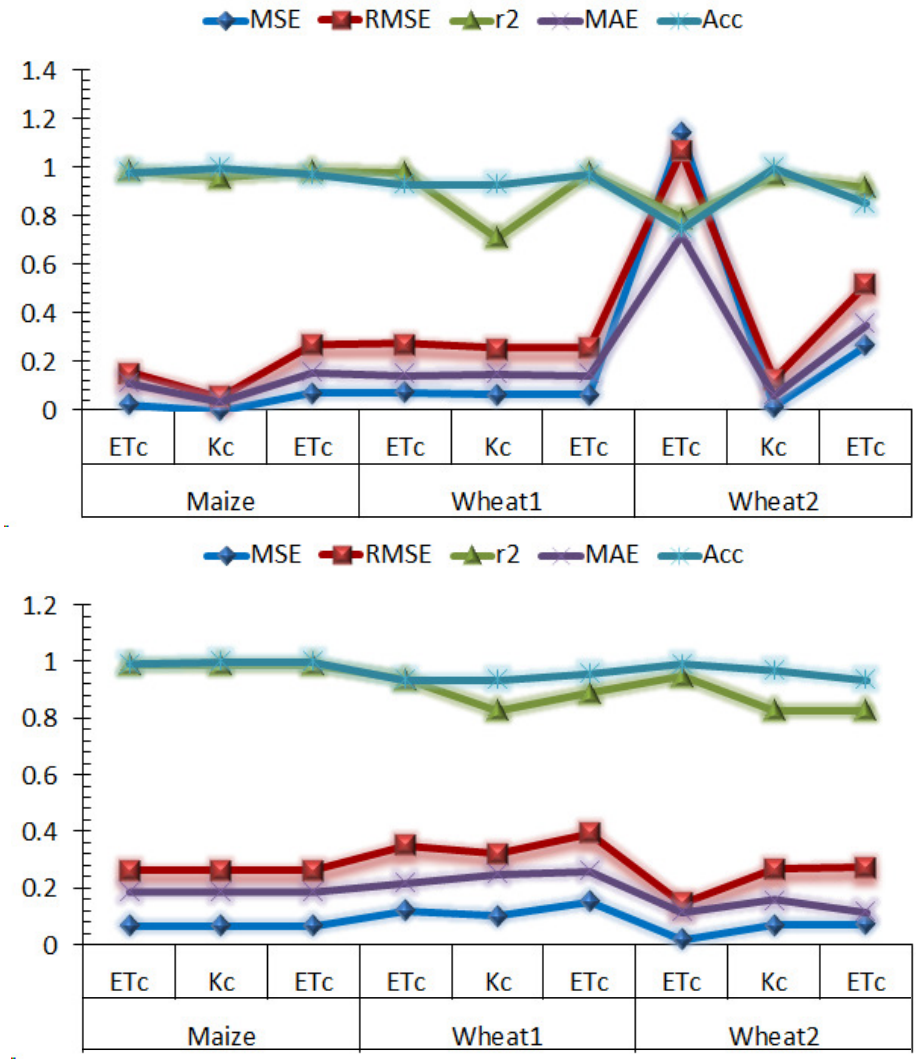

Figure 6.11: The comparison results of proposed model using training and testing datasets

The ET_c results with RMSE of (0.150 to 0.273) mm day^{-1}, MSE of (0.022 to 0.075) mm day^{-1}, MAE of (0.119 to 0.118) mm day^{-1}, ACC of (99% and 94%) mm day^{-1} using Fuzzy-Genetic and RRF model for testing dataset. Fig. 6.8 presented the comparison of the predicted and observed K_c and ET_c values for $Wheat_2$ crop using proposed model for training and testing periods.

6.5.4 Comparison of Actual, Proposed and SVM model of K_c and ET_c

The (FG-RRF) proposed model presented the Fuzzy-Genetic (FG), and Regularized Random Forest (RRF) for estimating value for K_c and ET_c using meteorological and crop coefficient parameters including maximum air temperature (T_{max}), minimum air temperature (T_{min}), relative humidity (R_H), wind speed (u_2), solar-

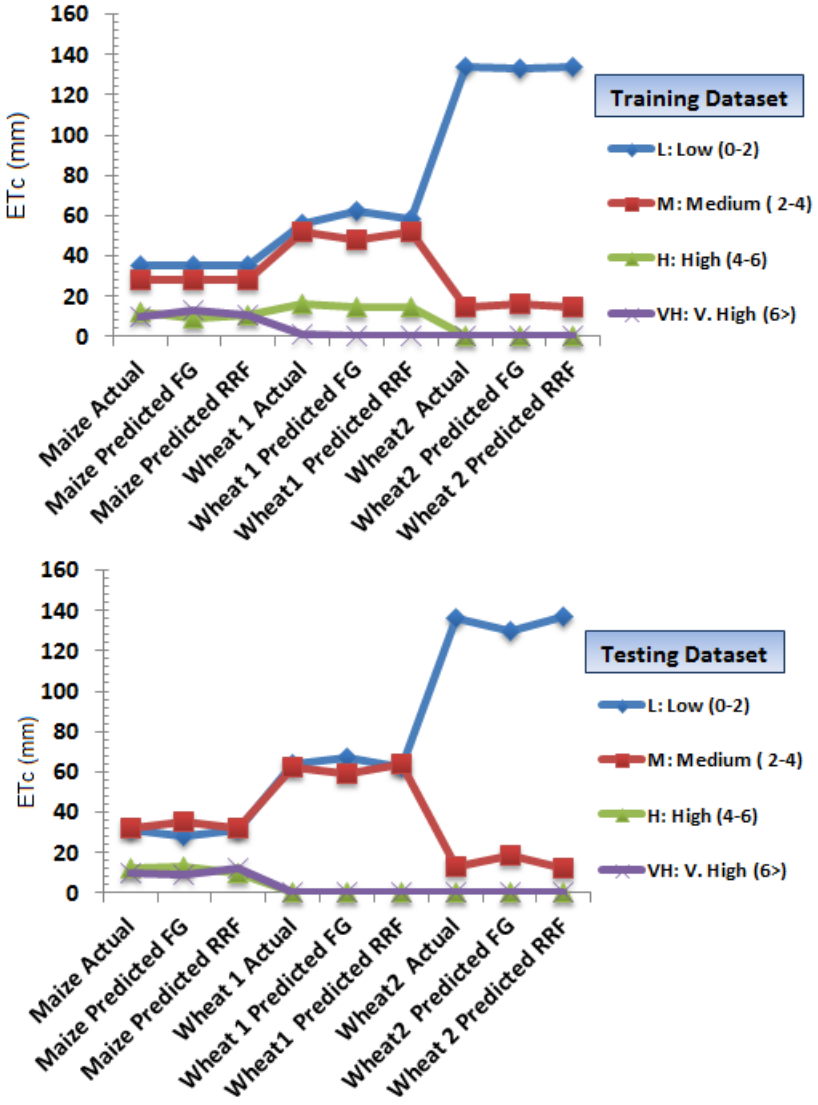

Figure 6.12: Forecasted actual and proposed model ET_c values in range of low, medium, high and very high

radiation (R_s), and sunshine hours (I_s). Shrestha et al. [220] estimated the K_c and ET_c for pepper and watermelon crops using different irrigation and climatic conditions (including air temperature, wind speed, relative humidity, solar radiation, rainfall, soil moisture, irrigation frequency, and water table depth (lysimeter) using Support Vector Machine (SVM) model (Sigma and C).

In case of Maize crop, comparison of results of proposed and SVM showed slightly variation in the last stage of crop coefficient (K_c) during training period. It is observed that proposed model provide efficient results than SVM model during training and testing period. It is also observed that SVM model provide correct estimation of ET_c using training dataset. However, proposed model provided superior results than SVM model during testing period. In case of Wheat crop, SVM model showed slightly better performance during training period, but proposed model performed

Table 6.6: Actual, the FG-RRF-predicted, the SVM model values of Maize crop K_c and ET_c with training dataset

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
1	0.7	0.7	0.79998	1.22	1.236774611	1.369801487
2	0.7	0.7	0.79991	1.22	1.220258065	0.920378307
3	0.7	0.7	0.79992	1.22	1.222051282	0.780370336
4	0.7	0.7	0.79992	1.22	1.220277457	1.120366274
5	0.7	0.7	0.79973	1.22	1.230544444	1.259709419
6	0.7	0.7	0.7998	1.22	1.230488351	1.299718444
7	0.7	0.7	0.79986	1.22	1.226061694	1.340374269
8	0.7	0.7	0.80012	1.38	1.558059965	1.599556032
9	0.7	0.7	0.80033	1.34	1.330555856	1.569768558
10	0.7	0.7	0.79992	1.22	1.244519553	1.369868402
11	0.7	0.7	0.7999	1.22	1.227135417	1.250259317
12	0.7	0.7	0.7999	1.22	1.2542775	1.389801487
13	0.7	0.7	0.80007	1.34	1.341614973	1.569734242
14	0.7	0.7	0.80033	1.56	1.551723423	1.710213283
15	0.7	0.7	0.80036	1.54	1.54673975	1.699912441
16	0.7	0.7	0.79993	1.56	1.544184783	1.71030778
17	0.7	0.7	0.79993	1.22	1.270549261	1.460376486
18	0.7	0.7	0.80051	1.29	1.446233512	1.540169347
19	0.7	0.75	0.80003	1.38	1.403200197	1.59990809
20	0.7	0.74	0.80005	1.22	1.264626102	1.440376664
21	0.7	0.72	0.80048	1.32	1.348311287	1.560322832
22	0.7	0.75	0.80008	1.54	1.599483425	1.699728744
23	0.7	0.71	0.79992	1.42	1.463872549	1.619771296
24	0.7	0.71	0.80007	1.76	1.799569106	1.849728225
25	0.7	0.71	0.79992	1.53	1.609648571	1.689721199
26	0.7	0.75	0.80007	1.67	1.858238095	1.790019656
27	0.7	0.77	0.8	1.22	1.886293651	1.309848927
28	0.7	0.71	0.79988	1.88	1.996182927	1.930374741
29	0.7	0.72	0.79993	2.1	2.185054974	2.109870394
30	0.7	0.77	0.79986	2.14	2.330810284	2.139754283
31	0.7	0.74	0.8	1.22	1.278981378	1.36996098
32	0.7	0.74	0.79965	1.22	1.264438739	1.119867394
33	0.7	0.74	0.80025	1.8	1.79941989	1.88035845
34	0.7	0.74	0.80036	1.85	1.85329619	1.909707639
35	0.7	0.74	0.80055	1.78	1.895243455	1.859721091
36	0.85	0.73	0.89975	2.02	1.658930851	2.139717696
37	0.85	0.71	0.89994	1.78	1.373728621	1.660383251
38	0.85	0.77	0.8999	2.29	2.17300289	2.399466945
39	0.85	0.77	0.8974	2.32	2.168454286	2.420035594
40	0.85	0.8	0.89986	2.41	2.495790323	2.520383999
41	0.85	0.79	0.89455	2.68	2.579266484	2.819756209
42	0.85	0.89	0.90015	2.76	2.890909761	2.91014692
43	0.85	0.93	0.89994	1.8	2.270076993	1.949707601
44	0.85	0.97	0.89995	2.95	3.081348231	2.969707601
45	0.85	0.78	0.89994	2.34	2.408937385	2.450144121
46	0.85	0.78	0.89994	2.22	2.146442029	2.329867394
47	0.85	0.78	0.89994	2.46	2.788445355	2.570223601
48	0.85	0.83	0.89994	2.56	2.796302326	2.679771305
49	0.85	0.89	0.90078	2.75	2.879955263	2.900326006
50	0.85	0.94	0.90527	2.69	2.940166667	2.829746491
51	0.85	0.94	0.89995	2.79	3.186690476	2.949868402
52	0.85	0.81	0.90028	2.73	2.764468692	2.869708411
53	0.85	0.94	0.90943	2.67	2.839679775	2.799726375
54	1.15	0.91	1.0499	3.9	2.744571695	3.749756217
55	1.15	1.08	1.05011	2.96	2.704408602	3.180229433
56	1.15	1.06	1.05055	2.24	2.323521505	2.600366825
57	1.15	1.11	1.05007	2.27	2.372231429	2.629855467
58	1.15	1.11	1.05005	2.38	2.42872093	2.750372343
59	1.15	1.11	1.04972	2.46	2.49331016	2.840376219
60	1.15	1.11	1.05011	2.76	2.681986111	3.077781924
61	1.15	1.11	1.04971	2.91	2.791843407	3.1403837
62	1.15	1.11	1.04963	3.13	2.973822581	3.289734242
63	1.15	1.09	1.05007	3.61	2.93200448	3.640274195
64	1.15	1.11	1.04963	4.14	4.189373737	4.099777234
65	1.15	1.11	1.04992	4.45	4.500049828	4.33028064
66	1.15	1.11	1.0499	4.62	4.756782486	4.410280556
67	1.15	1.11	1.04973	5.04	5.162990854	4.860365931
68	1.15	1.11	1.04987	5.28	5.227797178	5.099713125
69	1.15	1.1	1.04992	5.28	5.238751961	4.980360019
70	1.15	1.11	1.04992	5.55	5.163555877	5.310016619
71	1.05	1.05	0.94991	5.38	5.391581395	4.979752085
72	1.05	1.02	0.94993	6.23	6.048996933	5.890259081
73	1.05	1.08	0.94991	6.18	6.126134752	5.819717715
74	1.05	1.05	0.94991	6.05	6.398749031	5.849714167
75	1.05	1.05	0.94961	6.4	6.358708333	6.349870394
76	1.05	1.11	0.94993	6.96	6.411491892	6.940259089
77	1.05	1.04	0.95001	6.44	6.360964618	6.389708596
78	1.05	1.07	0.95007	6.81	6.621955095	6.769777239
79	1.05	1.04	0.95011	6.65	6.898885475	6.690140973
80	1.05	1.04	0.94993	7.15	6.909337814	7.093788416
81	1.05	1.06	0.94992	7.18	6.862989848	7.077498724
82	1.05	1.07	0.95043	6.37	6.447392505	6.199753388
83	1.05	1.05	0.9501	5.13	5.087273529	5.150144106
84	1.05	1.05	0.95011	7.17	5.998854592	7.084368126
85	1.05	1.04	0.9498	6.15	5.748074951	6.200345968

Table 6.7: Actual, the FG-RRF-predicted, the SVM model values of Maize crop K_c and ET_c with testing dataset testing dataset

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
1	0.7	0.7	0.899923594	1.27	1.22	3.07723441
2	0.7	0.7	0.899796651	0.87	1.23	3.074914799
3	0.7	0.7	0.899938191	1.34	1.24	3.077485138
4	0.7	0.7	0.888897966	1.02	1.27	2.881397577
5	0.7	0.7	0.895589907	1.16	1.23	3.010581453
6	0.7	0.7	0.899911371	1.2	1.22	3.077044956
7	0.7	0.7	0.89924346	1.24	1.23	3.064810586
8	0.7	0.7	0.899938815	1.5	1.36	3.077495091
9	0.7	0.7	0.897426883	1.47	1.34	3.034042452
10	0.7	0.7	0.899443129	1.27	1.24	3.068657376
11	0.7	0.7	0.899530207	1.15	1.25	3.070483745
12	0.7	0.73	0.899894166	1.29	1.43	3.076719433
13	0.7	0.72	0.899603734	1.47	1.52	3.071167293
14	0.7	0.78	0.899938922	1.93	2.03	3.077497534
15	0.7	0.74	0.89993841	1.8	1.89	3.07748791
16	0.7	0.74	0.899927627	1.89	2.01	3.077300091
17	0.7	0.71	0.899164482	1.63	1.59	3.062653131
18	0.7	0.71	0.874424316	1.65	1.74	2.685366831
19	0.7	0.75	0.899938974	1.48	1.51	3.07749854
20	0.7	0.76	0.899938983	1.79	1.88	3.077498724
21	0.7	0.75	0.899558233	2.04	2.08	3.071979673
22	0.7	0.73	0.895869892	1.6	1.58	3.025617545
23	0.7	0.72	0.899678123	1.52	1.43	3.073115807
24	0.7	0.72	0.899815724	1.75	1.77	3.075393163
25	0.7	0.73	0.899906848	1.59	1.66	3.077002992
26	0.7	0.72	0.899938983	1.69	1.74	3.077498724
27	0.7	0.73	0.89763882	1.21	1.42	3.039593639
28	0.7	0.71	0.899769907	1.83	1.79	3.074839919
29	0.7	0.71	0.897196578	2.01	1.95	3.03310702
30	0.7	0.71	0.899929877	2.04	2.01	3.077324848
31	0.7	0.71	0.899938983	1.76	1.78	3.077498724
32	0.7	0.71	0.885213384	1.94	2	2.870002557
33	0.7	0.72	0.899755933	1.94	1.99	3.074003866
34	0.7	0.71	0.899714995	1.75	1.77	3.073515096
35	0.7	0.78	0.899231075	2.04	2.05	3.066264047
36	0.7	0.81	0.898918158	2.14	2.14	3.062782037
37	0.85	0.79	0.899855312	2.54	2.38	3.076362597
38	0.85	0.71	0.898358376	2.18	1.99	3.056611997
39	0.85	0.85	0.899348573	2.75	2.53	3.068804854
40	0.85	0.85	0.899938963	2.31	2.16	3.077498322
41	0.85	0.85	0.897391826	2.41	2.26	3.03587182
42	0.85	0.85	0.896255236	2.48	2.31	3.028839215
43	0.85	0.85	0.899541857	2.41	2.31	3.071224131
44	0.85	0.85	0.899923473	2.75	2.5	3.077252714
45	0.85	0.85	0.899937242	2.7	2.55	3.077469718
46	0.85	0.85	0.899826665	2.36	2.41	3.075741449
47	0.85	0.85	0.899936634	2.99	3.88	3.077463195
48	0.85	0.85	0.899936231	2.85	2.85	3.077446308
49	0.85	0.85	0.899938817	2.74	2.71	3.07669394
50	0.85	0.85	0.899938852	2.73	2.68	3.077495784
51	0.85	0.85	0.899936457	2.85	2.94	3.077463554
52	0.85	0.85	0.899937042	2.77	2.8	3.077341712
53	0.85	0.85	0.899945652	2.7	2.72	3.076143645
54	0.85	0.85	0.899938719	2.85	2.85	3.077493986
55	1.15	0.8	0.89993865	3.28	2.66	3.077493197
56	1.15	0.84	0.89993898	2.5	2.37	3.077498682
57	1.15	0.78	0.897155738	2.53	2.29	3.049339259
58	1.15	0.71	0.899938975	2.65	2.19	3.077498523
59	1.15	0.71	0.899938933	2.74	2.21	3.077497632
60	1.15	0.8	0.89920005	3.08	2.48	3.068241937
61	1.15	0.78	0.899938983	3.8	3.03	3.077498726
62	1.15	1.15	0.899938983	4.58	4.82	3.077498724
63	1.15	1.15	0.899936183	3.94	3.16	3.077445472
64	1.15	1.15	0.899815315	4.38	4.63	3.075933127
65	1.15	1.15	0.900091626	3.97	3.21	3.079081088
66	1.15	1.15	0.907493984	4.47	4.56	3.153957919
67	1.15	1.15	0.899938826	4.23	4.54	3.077496491
68	1.15	1.15	0.90003461	5.2	5.21	3.078968055
69	1.15	1.15	0.899939175	5.08	5.2	3.077494332
70	1.15	1.15	0.900816591	5.41	5.27	3.090462929
71	1.15	1.11	0.899938983	5.57	5.11	3.077498724
72	1.05	1.06	0.899939214	5.99	6.01	3.077482718
73	1.05	0.9	0.900080333	5.92	5.99	3.076587264
74	1.05	1.05	0.899940778	5.95	6.09	3.077643962
75	1.05	1.11	0.899938993	6.45	6.1	3.07749921
76	1.05	1.11	0.899939228	7.04	6.55	3.077520456
77	1.05	1.11	0.90074053	6.49	6.11	3.08162261
78	1.05	1.05	0.899982469	6.87	6.59	3.077580061
79	1.05	1.05	0.89993932	6.79	6.4	3.077498233
80	1.05	1.05	0.899938996	7.32	6.79	3.077497418
81	1.05	1.05	0.899938994	7.43	6.74	3.077499253
82	1.05	1.05	0.899939383	6.3	6.01	3.077516911
83	1.05	1.05	0.901730217	5.25	5.19	3.094726772
84	1.05	1.05	0.900287625	7.35	6.72	3.077987353
85	1.05	1.05	0.899944076	6.3	6.07	3.077735059

Table 6.8: Actual, the FG-RRF-predicted, the SVM model values of Wheat crop K_c and ET_c with training dataset

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
1	0.4	0.4	0.59	1.04	1.1504	1.1396
2	0.4	0.4	0.65	0.7	0.8029	0.7996
3	0.4	0.5	0.64	0.92	1.1061	1.0199
4	0.4	0.4	0.64	1.59	1.433	1.6898
5	0.4	0.4	0.5	1.22	1.2119	1.3202
6	0.4	0.4	0.5	1.25	1.2075	1.35
7	0.4	0.4	0.65	1.02	1.0402	1.12
8	0.4	0.4	0.53	0.99	1.0211	1.0898
9	0.4	0.4	0.52	0.86	0.9216	0.9596
10	0.4	0.4	0.65	1.14	1.1639	1.2399
11	0.4	0.4	0.65	1.01	1.0783	1.1104
12	0.4	0.4	0.65	0.48	0.5264	0.58
13	0.4	0.4	0.65	0.76	0.8163	0.8599
14	0.4	0.4	0.65	0.73	0.7887	0.83
15	0.4	0.4	0.65	0.43	0.4863	0.5305
16	0.4	0.4	0.65	0.79	0.8136	0.8899
17	0.4	0.5	0.65	1.38	1.4683	1.4801
18	0.4	0.4	0.65	1.09	1.1084	1.1897
19	0.4	0.5	0.64	1.17	1.2818	1.2703
20	0.4	0.4	0.65	1.14	1.3747	1.2398
21	0.4	0.4	0.89	0.79	0.8571	0.8896
22	0.4	0.4	0.65	0.48	0.57	0.58
23	0.4	0.4	0.65	0.31	0.4143	0.4104
24	0.4	1.1	0.65	0.43	0.6152	0.5299
25	0.4	1.1	0.65	0.37	0.5222	0.4702
26	0.4	0.4	0.66	0.48	0.5005	0.5796
27	1.15	1.11	1.05	2.28	2.2075	2.2692
28	1.15	1.04	1.05	2.29	2.1684	2.269
29	1.15	0.95	1.05	1.92	1.6555	2.0199
30	1.15	1.11	1.05	1.46	1.397	1.5602
31	1.15	1.11	1.05	1.36	1.4003	1.4599
32	1.15	1.11	1.05	1.06	1.0929	1.1599
33	1.15	1.11	1.05	1.69	1.5554	1.7903
34	1.15	1.11	1.05	1.22	1.2095	1.3203
35	1.15	1.15	1.05	1.23	1.2103	1.33
36	1.15	1.15	1.05	1.13	1.197	1.23
37	1.15	1.15	1.05	1.58	1.4996	1.68
38	1.15	1.15	1.05	1	1.0806	1.1004
39	1.15	1.15	1.05	1.99	1.8713	2.09
40	1.15	1.15	1.05	2.9	2.8147	2.7999
41	1.15	1.15	1.05	2.79	2.7324	2.6903
42	1.15	1.15	1.05	3	2.901	2.9
43	1.15	1.15	1.05	2.98	2.8154	2.8797
44	1.15	1.15	1.05	1.93	1.8688	2.0299
45	1.15	1.11	1.05	2.73	2.6827	2.63
46	1.15	0.7	1.05	2.32	1.8825	2.2199
47	1.15	1.11	1.05	0.94	1.0627	1.0403
48	1.15	1.05	1.05	2.56	2.3833	2.4602
49	1.15	1.11	1.05	1.19	1.1447	1.2899
50	1.15	1.11	1.05	2.76	2.6057	2.6599
51	1.15	1.11	1.05	1.56	1.2179	1.6605
52	1.15	1.11	1.05	2.46	2.3331	2.36
53	1.15	1.11	1.05	2.47	2.3874	2.5683
54	1.15	1.11	1.05	2.33	2.2778	2.2521
55	1.15	1.11	1.05	2.51	2.3484	2.4097
56	1.15	1.15	1.05	3.02	2.8356	2.9206
57	1.15	1.15	1.05	2.74	2.6562	2.6401
58	1.15	1.15	1.05	2.29	2.1818	2.2625
59	1.15	1.15	1.05	2.77	2.6714	2.6706
60	1.15	1.15	1.05	2.8	2.7141	2.6999
61	1.15	1.15	1.05	1.41	1.2046	1.5104
62	1.15	1.15	1.05	2.02	1.7898	2.1204
63	1.15	1.15	1.05	2.2	1.9489	2.1802
64	1.15	1.15	1.05	2.87	2.7071	2.7699
65	1.15	1.15	1.05	1.8	1.4279	1.8997
66	1.15	1.15	1.05	1.51	1.1812	1.6099
67	1.15	1.15	1.05	1.64	1.2607	1.7399
68	1.15	1.15	1.05	2.79	2.7043	2.6898
69	1.15	1.11	1.05	2.89	2.7578	2.7897
70	1.15	1.11	1.05	2.77	2.6803	2.6701
71	1.15	1.11	1.05	3.03	2.8687	2.9301
72	1.15	1.11	1.05	2.68	2.6207	2.5798
73	1.15	1.11	1.05	3.26	3.1377	3.1599
74	1.15	1.11	1.05	3.18	3.0963	3.0801
75	1.15	1.11	1.05	3.31	3.2023	3.2104
76	1.15	1.11	1.05	3.2	3.0387	3.0996
77	1.15	1.11	1.05	2.61	2.4769	2.5103
78	1.15	1.11	1.05	1.57	1.2177	1.6698
79	1.15	1.11	1.05	2.78	2.7241	2.6803
80	1.15	1.11	1.05	2.69	2.6863	2.5902
81	1.15	1.11	1.05	2.68	2.6862	2.5798
82	1.15	1.11	1.05	2.82	2.8146	2.7202
83	1.15	1.11	1.05	1.33	1.2726	1.4298
84	1.15	1.11	1.05	3.3	3.0345	3.1999
85	1.15	1.11	1.05	1.06	1.0737	1.16
86	1.15	1.11	1.05	2.83	2.8523	2.7298
87	1.15	1.15	1.05	2.84	2.8622	2.7398
88	1.15	1.15	1.05	3.86	3.9185	3.7603
89	1.15	1.15	1.05	3.97	4.003	3.8696
90	1.15	1.15	1.05	3.85	3.8573	3.7498

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
91	1.15	1.15	1.05	4.06	4.0149	3.9602
92	1.15	1.15	1.05	4.07	4.0248	3.9701
93	1.15	1.15	1.05	4.04	4.0703	3.94
94	1.15	1.15	1.05	4	4.0718	3.8997
95	1.15	1.15	1.05	4.27	4.2394	4.2477
96	1.15	1.15	1.05	4.51	4.5005	4.4101
97	1.15	1.15	1.05	4.67	4.6548	4.5702
98	1.15	1.15	1.05	4.74	4.7261	4.6396
99	1.15	1.15	1.05	5.65	5.2144	5.5502
100	1.15	1.15	1.05	6.02	5.1319	5.9199
101	1.15	1.15	1.05	4.33	4.3336	4.2299
102	1.15	1.15	1.05	4.02	4.0385	3.92
103	1.15	0.59	1.05	5.85	4.8624	5.75
104	1.15	0.94	1.05	4.53	4.5131	4.43
105	1.15	0.63	1.05	4.73	4.4899	4.6301
106	1.15	1.11	1.05	4.7	4.5602	4.6002
107	1.15	0.72	1.05	4.94	4.6849	4.84
108	0.41	0.4	0.54	1.87	1.9035	1.9704
109	0.41	0.4	0.55	1.78	1.8581	1.8801
110	0.41	0.4	0.51	2.03	2.0837	1.9298
111	0.41	0.4	0.52	2.09	2.1045	2.1032
112	0.41	0.4	0.65	1.87	1.9557	1.9699
113	0.41	0.4	0.65	2.99	2.912	2.8904
114	0.41	0.4	0.63	1.87	1.938	1.9702
115	0.41	0.4	0.56	1.91	2.043	2.0097
116	0.41	0.41	0.72	1.9	1.9062	1.9997
117	0.41	0.41	0.65	1.55	1.6315	1.6501
118	0.41	0.41	0.65	1.2	1.1895	1.2997
119	0.41	0.4	0.65	1.64	1.8605	1.7402
120	0.41	0.4	0.63	2.12	2.1099	2.2199
121	0.41	0.4	0.63	2.2	2.252	2.2626
122	0.41	0.4	0.64	2.11	2.2668	2.2099
123	0.41	0.4	0.64	2.46	2.4609	2.36
124	0.41	0.4	0.65	2.39	2.4575	2.2904
125	0.41	0.4	0.65	3.18	2.957	3.0799
126	0.41	0.41	0.65	3.32	3.1056	3.2199

Table 6.9: Actual, the FG-RRF-predicted, the SVM model values of Wheat crop K_c and ET_c with testing dataset

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
1	0.4	0.4	0.9	1.04	1.5856	2.2674
2	0.4	0.4	0.9	0.7	0.8661	2.2692
3	0.4	0.5	0.9	0.92	1.0053	2.2643
4	0.4	0.4	0.9	1.59	1.4455	2.2692
5	0.4	0.4	0.9	1.22	1.3134	2.2692
6	0.4	0.4	0.9	1.25	1.3192	2.264
7	0.4	0.4	0.9	1.02	1.6901	2.2692
8	0.4	0.4	0.89	0.99	1.0709	2.2228
9	0.4	0.4	0.9	0.51	0.7823	2.2693
10	0.4	0.4	0.9	0.46	0.6515	2.2641
11	0.4	0.4	0.9	0.62	1.6343	2.2692
12	0.4	0.4	0.91	0.52	0.7752	2.2834
13	0.4	0.4	0.9	0.76	0.934	2.2672
14	0.4	0.4	0.9	0.73	0.8858	2.2692
15	0.4	0.4	0.9	0.43	0.6883	2.2689
16	0.4	0.4	0.9	0.78	0.8773	2.2706
17	0.4	0.5	0.9	1.38	1.5053	2.2554
18	0.4	0.4	0.9	1.08	1.2205	2.2692
19	0.4	0.5	0.9	1.17	1.2157	2.2692
20	0.4	0.4	0.9	1.14	1.1917	2.2692
21	0.4	0.4	0.85	0.79	0.8894	1.9793
22	0.4	0.4	0.9	0.48	0.6482	2.2678
23	0.4	0.4	0.9	0.31	0.6549	2.2673
24	0.4	1.1	0.9	0.43	0.6232	2.2686
25	0.4	1.1	0.9	0.37	0.6171	2.2337
26	0.4	0.4	0.86	0.29	0.6352	1.7798
27	1.15	1.11	0.9	1.18	1.3746	2.2697
28	1.15	1.04	0.9	0.84	1.2704	2.2265
29	1.15	0.95	0.9	1.29	1.3163	2.1904
30	1.15	1.11	0.9	1.25	1.2249	2.2639
31	1.15	1.11	0.84	0.84	1.2508	1.4874
32	1.15	1.11	0.94	1.01	1.237	1.8022
33	1.15	1.11	0.9	1.28	1.2701	2.2481
34	1.15	1.11	1.04	0.98	1.2279	1.3502
35	1.15	1.15	0.91	1.13	1.1979	2.1325
36	1.15	1.15	0.9	1.32	1.3892	2.2692
37	1.15	1.15	0.93	1.32	1.3909	2.0925
38	1.15	1.15	0.9	1.28	1.3535	2.2692
39	1.15	1.15	0.92	1.27	1.3928	2.2234
40	1.15	1.15	0.9	0.91	1.3036	2.2692
41	1.15	1.15	0.91	1.17	1.3482	2.3148
42	1.15	1.15	0.9	1.2	1.3653	2.3304
43	1.15	1.15	0.95	1.3	1.4424	2.549
44	1.15	1.15	0.99	1.28	1.3939	2.0968
45	1.15	1.11	1.03	1.64	1.8914	2.5868
46	1.15	0.7	0.9	1.41	1.573	2.2832
47	1.15	1.11	0.9	0.99	1.2731	2.2692
48	1.15	1.05	0.9	1.29	1.3246	2.2688
49	1.15	1.11	0.92	1.2	1.3172	1.9631
50	1.15	1.11	0.9	2.76	2.4428	2.2685

	Crop Coefficient (K_c)			Crop Evapotranspiration (ET_c)		
Days	K_c	Proposed	SVM	ET_c	Proposed	SVM
51	1.15	1.11	0.9	1.55	1.5501	2.2513
52	1.15	1.11	0.9	2.46	2.3848	2.2692
53	1.15	1.11	0.9	2.46	2.2793	2.146
54	1.15	1.11	0.9	2.33	2.2392	2.2581
55	1.15	1.11	0.9	2.51	2.3229	2.1955
56	1.15	1.15	0.9	3.02	2.5115	2.2416
57	1.15	1.15	0.89	2.74	2.3743	2.2106
58	1.15	1.15	0.92	2.29	2.4503	2.3719
59	1.15	1.15	0.9	2.77	2.6252	2.2957
60	1.15	1.15	0.9	2.79	2.5715	2.2436
61	1.15	1.15	0.9	1.41	1.5533	2.2634
62	1.15	1.15	0.9	2.02	1.9257	2.2692
63	1.15	1.15	0.9	2.2	2.4698	2.2692
64	1.15	1.15	0.9	2.88	2.7355	2.2655
65	1.15	1.15	0.9	1.79	2.3978	2.2696
66	1.15	1.15	0.9	1.51	1.5565	2.2153
67	1.15	1.15	0.88	1.64	1.7097	2.1064
68	1.15	1.15	0.9	2.79	2.495	2.251
69	1.15	1.11	0.88	2.89	2.7075	2.2267
70	1.15	1.11	0.89	2.77	2.7249	2.2249
71	1.15	1.11	0.9	3.04	2.7574	2.2637
72	1.15	1.11	0.9	2.68	2.5208	2.2687
73	1.15	1.11	0.9	3.25	2.8508	2.2622
74	1.15	1.11	0.99	3.19	2.974	2.5133
75	1.15	1.11	0.9	3.31	3.3138	2.3305
76	1.15	1.11	0.9	2.45	2.9035	2.3647
77	1.15	1.11	0.9	2.58	3.0043	2.3078
78	1.15	1.11	0.9	2.83	3.2033	2.2692
79	1.15	1.11	0.9	2.94	3.1502	2.2699
80	1.15	1.11	0.9	3.17	3.4724	2.27
81	1.15	1.11	0.9	2.96	3.2238	2.2692
82	1.15	1.11	0.9	3.11	3.2421	2.2713
83	1.15	1.11	0.9	2.68	2.7441	2.2658
84	1.15	1.11	0.9	2.71	3.2336	2.2712
85	1.15	1.11	0.9	2.43	3.1283	2.2958
86	1.15	1.11	0.9	2.93	3.309	2.2692
87	1.15	1.15	0.9	3.35	3.4247	2.3741
88	1.15	1.15	0.9	3.02	3.1123	2.2691
89	1.15	1.15	0.9	2.83	2.9669	2.2692
90	1.15	1.15	0.9	2.94	2.881	2.2692
91	1.15	1.15	0.9	3.28	3.3539	2.2693
92	1.15	1.15	0.9	3.31	3.2714	2.2692
93	1.15	1.15	0.89	3.09	3.1786	2.2516
94	1.15	1.15	0.9	1.91	1.9136	2.2692
95	1.15	1.15	0.9	2.4	2.5273	2.2625
96	1.15	1.15	0.92	3.25	3.3146	2.535
97	1.15	1.15	0.9	2.9	3.0375	2.2692
98	1.15	1.15	0.9	3.19	3.2942	2.397
99	1.15	1.15	0.9	3.19	3.1895	2.2818
100	1.15	1.15	0.9	2.88	2.7754	2.2692
101	1.15	1.15	0.9	3.35	3.4464	2.2692
102	1.15	1.15	0.9	3.51	3.6221	2.5155
103	1.15	0.59	0.9	3.44	3.4484	2.2692
104	1.15	0.94	0.9	2.96	3.3427	2.2692
105	1.15	0.63	0.9	3.73	4.1361	2.5125
106	1.15	1.11	0.9	3.13	3.2949	2.2701
107	1.15	0.72	0.9	3.25	3.2926	2.2688
108	0.4	0.4	0.9	1.17	1.744	2.2696
109	0.4	0.4	0.9	1.14	1.7134	2.2692
110	0.4	0.4	0.9	1.29	1.8261	2.4085
111	0.4	0.4	0.9	1.27	1.8258	2.5559
112	0.4	0.4	0.9	1.11	1.7643	2.2694
113	0.4	0.4	0.65	2.99	1.9189	2.8904
114	0.4	0.4	0.63	1.86	1.8562	1.9702
115	0.4	0.4	0.56	1.91	1.8838	2.0097
116	0.4	0.4	0.72	1.9	1.8636	1.9997
117	0.4	0.4	0.65	1.54	1.6513	1.6501
118	0.4	0.4	0.65	1.2	1.5754	1.2997
119	0.4	0.4	0.65	1.64	1.7802	1.7402
120	0.4	0.4	0.63	2.12	1.8567	2.2199
121	0.4	0.4	0.63	2.2	1.8489	2.2626
122	0.4	0.4	0.64	2.11	1.856	2.2099
123	0.4	0.4	0.64	2.46	1.8488	2.36
124	0.4	0.4	0.65	2.38	1.8495	2.2904
125	0.4	0.4	0.65	3.18	1.8495	3.0799
126	0.4	0.4	0.65	3.32	1.8614	3.2199

Table 6.10: Performance comparison of FG-RRF (Proposed model) and SVM model using Maize crop

	Training Maize Kc		Testing Maize Kc	
	Proposed	SVM	Proposed	SVM
MAE	0.04031	0.0895	0.0512	0.1689
MSE	0.00298	0.0084	0.01346	0.0332
RMSE	0.0546	0.0918	0.116	0.1824
ACC	100%	100%	100%	80%
R	0.91	0.74	0.59	0.28
	Training Maize ETc		Testing Maize ETc	
MAE	0.1584	0.1052	0.1871	1.433
MSE	0.07129	0.0124	0.0691	2.776
RMSE	0.2670	0.1114	0.2629	1.666
ACC	97.65%	96.47%	98%	24.71%
R	0.98	0.99	0.97	-2.56

Table 6.11: Performance comparison of FG-RRF (Proposed model) and SVM model using Wheat Crop

	Training Wheat Kc		Testing Wheat Kc	
	FG-RRF	SVM	FG-RRF	SVM
MAE	0.15051	0.1440	0.2497	0.3047
MSE	0.0640	0.0257	0.10433	0.1050
RMSE	0.25315	0.1604	0.3230	0.3241
ACC	93.65%	99.21%	94.44%	75.4%
R	0.50	0.79	0.19	0.17
	Training Wheat ETc		Testing Wheat ETc	
MAE	0.1478	0.0956	0.2635	0.8162
MSE	0.0644	0.0094	0.1564	0.8999
RMSE	0.2539	0.0972	0.3955	0.9486
ACC	97%	99%	93%	65.87%
R	0.95	0.99	0.77	0.010

better outcomes during testing period for K_c. It is observed that the proposed model provide accurate outcomes of ET_c during testing period than SVM model.

Table 6.6 and Table 6.7 presents the values of Maize crop (K_c) and (ET_c) using Actual, Proposed model and SVM model for training dataset and testing datasets respectively. Table 6.8 and Table 6.9 presents the values of Wheat crop (K_c) and (ET_c) using Actual, Proposed model and SVM model for training dataset and testing datasets respectively.

Table 6.10 presents the performance comparison of FG-RRF Proposed model and SVM model for Maize crop (K_c) and (ET_c) using training dataset and testing datasets respectively. Table 6.11 presents the performance comparison of FG-RRF Proposed model and SVM model for $Wheat_1$ crop (K_c) and (ET_c) using training dataset and testing datasets respectively.

6.6 Summary

The main objective of this paper is to design and develop an innovative multi-level model ensembling for accurate estimation of crop coefficient (K_c) and reference evapotranspiration (ET_c) using Fuzzy-Genetic (FG) and Regularization Random Forest(RRF) models. The following conclusions are drawn from the investigations.

- The (FG-RRF) proposed ensemble model estimated accurately (K_c) and (ET_c) for Maize crop than other crops. The (K_c) value was correctly predicted using Fuzzy-Genetic approach for Initial, Development, Middle and Last stage of crops. The climatic and crop parameters such as ($Temperature_{Max}$), (RH_{Max}), (RH_{Min}), (K_c) and (ET_o) found better to estimate (ET_c) for crops.
- To evaluate the accuracy of proposed model, (K_c) and (ET_c) predicted results are compared with the SVM model. The testing dataset found slightly better than training dataset using proposed model as compared to SVM model in case of (K_c) and (ET_c) estimations. The highest accuracy was achieved with by the (K_c) with 99% and (ET_c) with 98% for Maize crop using training and testing period.

Chapter 7

Decision Support System for Irrigation Scheduling using AI (DSS-IS)

This study presents a framework of Decision Support System (DSS) based on Artificial Intelligence (AI) for irrigation scheduling. Particle Swarm Optimization with Deep Neural Network (PSO-DNN) and Deep Learning (DL) models are applied to estimate the irrigation scheduling parameters such as (Irrigation Water Requirement, Net Irrigation, Gross Irrigation and Pumping Time) that can help to improve water use efficiency.

7.1 Overview

"Agriculture is at the heart of our world's economy" - Crop life international. Emerging technologies such as Artificial Intelligence (AI) and Machine Learning (ML) make agriculture ubiquitous and helps to develop model-based Decision Support System (DSS). A combination of DSS method with modern technology and irrigation, is the need of the hour to develop smart utilization of water to the crops in different climate conditions.

In modern farming, the watering process has become one of the important and critical processes due to shortage of fresh water in most of the area of the world [221]. Farmland consume more than 70% of total global freshwater, and is much higher (up to 90%) in the developing countries. In order to estimate crop evapotranspiration (ET_c) a significant amount of fresh water is needed for better crop growth and yield. In addition to rain, the proper irrigation water is also required for the growth of crop.

It is necessary to use an efficient irrigation technique that saves irrigation water without compromising yield production. Punjab is an economically advanced and energy-intensive state of India. In North-Western India, Wheat is usually irrigated

at an IW: CPE of 0.9 with a 75 mm depth of irrigation water (conventional irrigation technique, CP), which results in water waste. The quantity and timing of precipitation have a significant impact on water irrigation needs. The annual precipitation is usually less than ten inches in arid areas and irrigation needs to be done to grow crops successfully.

Crops can be cultivated without irrigation, in semi-arid areas that typically receive 15 to 20 inches of annual precipitation but due to droughts, crop yield decreases and it causes crop failure. Short, and dry periods are common in sub-humid areas, which receive between 20 to 30 inches of annual rainfall. Depending on soil water storage capacity and crop root depth, irrigation is required for short periods during the growth stage in these areas.

Irrigation Decision Support System (IDSS) plays an important role by providing the scientific and reasonable suggestion for efficient utilization of water management [222]. In the last two decades, various researchers have observed the success of the Integrated Decision Support System (IDSS) in various domains of agricultural systems. Mateos et al. [223] presented a scheme of Irrigation Management Information System (SIMIS) to enhance the irrigation efficiency based on a simple water balance method. Maza et al. [224] applied the ArcGIS toolbar to estimate accurate ETo and Net Irrigation Requirement (NIR) values for crop water demand. Srivastava et al. [225] developed a hyper-spectral predictive approach to monitor the soil organic carbon in the Indo-Gangetic Plains of Punjab, India.

Irrigation's primary goal is to supply enough water to crops in order to achieve optimal yield and a high-quality harvest. Accurate estimation of quantity and timings of water can be determined by current climatic conditions, the crop stages, and soil water storage capacity.

Moreover, the field water balance for crop evapotranspiration is required to determine the estimation of crop water requirement.

As motivated by Sharda et al., [219] developed a DSS based on Border, Sprinkler and Drip Irrigation systems to the estimation of when and how to determine water irrigation for Maize, Potato, Chilli, and Wheat crops. They have applied some dataset to build the DSS system such as climate dataset, crop dataset with water quality, soil dataset and irrigation system.

In this chapter, the main contribution is to analyze the irrigation scheduling parameter using feature selection based on random forest algorithm and ensemble-based modeling using PSO and DL models. We have shown that the feature subset can improve the accuracy of this model. The proposed model have been applied on case study which is received from PAU, Ludhiana station. We have taken dataset of Border Irrigation and Sprinkler Irrigation parameters to determine the Irrigation Water Requirement (IWR).

Table 7.1: The original case study parameters of Border Irrigation

Date	Day	ETo	Kc	ETc	RF	Net Ir	Net Ir Cu	Gross Ir	Gross Cu	PT
12-12-2012	12	1.210	0.4	1.010	4.3	0	0	0.000	0	0
14-12-2012	14	1.831	0.4	1.010	13.4	0	0	0.000	0	0
15-01-2013	46	1.071	1.15	2.317	0	47.5	4750000	118.750	11875000	19.7
03-02-2013	64	1.563	1.15	1.797	0	41.7	4170000	104.250	10425000	17.3
04-02-2013	65	1.313	1.15	1.510	6.6	0	0	0.000	0	0
05-02-2013	66	1.425	1.15	1.639	3.2	0	0	0.000	0	0
06-02-2013	67	2.427	1.15	2.792	10	0	0	0.000	0	0
16-02-2013	77	1.363	1.15	1.567	8.5	0	0	0.000	0	0
17-02-2013	78	2.419	1.15	2.782	26	0	0	0.000	0	0
23-02-2013	84	0.924	1.15	1.063	14.6	0	0	0.000	0	0
05-03-2013	95	3.713	1.15	4.270	0	34.98	3498000	87.450	8745000	14.5
15-03-2013	105	4.113	1.15	4.730	0	44.54	4454000	111.350	11135000	18.5
31-03-2013	121	5.493	0.4	2.197	0	31.9	3190000	79.750	7975000	13.2

7.2 Material and Study Area

This section discusses about the case study of Ludhiana District which is one of the centrally located District of Punjab (India) with 3706 sq km geographical area. The topography of the study area is a typical representation of an alluvial plain. The climate of the area is sub-tropical steppe, semi-arid with hot dry summers from April to June, hot and humid monsoon period from July to October, winters from November to January and mild temperature during February to March.

The case study have been carried out at the research farms of Punjab Agricultural University (PAU), Ludhiana, India 30(°C) 54 N, 75(°C) 48 E, elevation 247 m above sea level). For the development of DSS-IR model, three datasets are considered (Weather dataset, Crop dataset and Irrigation Methods). Firstly, the daily meteorological data with the following parameters (viz Max/Min Temperature, Relative Humidity, Wind Speed, Sunshine Hours and Rainfall) are collected from IMD, Pune. Then, ET_o is calculated using weather dataset and CROPWAT software. Secondly, crop dataset parameters (i.e. Crop Variety, Crop Sowing/Planting Date) from package of practices for crops and vegetables of Punjab Agricultural University, Ludhiana are taken for estimation of ET_c. Finally, the irrigation system dataset is considered with several parameters (viz Irrigation Frequency, Irrigation Requirement, Net Irrigation, Gross Irrigation, Pumping Time). Table 7.1 shows the original case study parameters of Border Irrigation and Fig 7.2 presents the crop coefficient values of Wheat crop.

Table 7.2: Crop coefficient (Kc) values of Wheat crop at various stages of Ludhiana station

Period	Variety	Initial		Development		Mid		Late		TD
		K_c	TGD	K_c	TGD	K_c	TGD	K_c	TGD	
2013/12/01-2014/04/31	PBW621	0.4	29	1.15	55	1.15	14	0.4	32	130

7.3 Feature Subset using Fuzzyforest

In this section, the fuzzyforest algorithm is applied for the feature subset. The fuzzyforest algorithm is based on random forest and it is designed to reduce and rank the important number of features in regression. These features are chosen based on a feature recursive exclusion function. The random forests algorithm is applied for both classification and regression problems [180].

This model for the use of a variety of methods for partitioning the features into distinct clusters. The fuzzyforest package allows the analyst to input their own clustering of the features [226]. Commonly, such a partition of the features would be

derived by considering the correlation matrix of the features. The R studio software is used to implement this fuzzyforest package and it gives the analyst the option of utilizing the functionality of the package WGCNA to partition covariates into distinct clusters [227].

WGCNA (Weighted Gene Co-expression Network Analysis) is a rigorous framework for detecting correlation networks [228]. Although WGCNA has been used primarily in genetics, it has also been applied successfully in contexts such as brain imaging and cancer biology [229].

WGCNA takes in the matrix of features and uses the correlation structure to partition the features into distinct groups such that the correlation between features of the same group is large and the correlation between features of separate groups is small. WGCNA constructs a network of features, each feature representing one node, via the correlation matrix of features. It determines modules based on this network.

The working process of model is described in Fig 7.1. In the first step of fuzzy forests, the input features are divided into various samples based on high and low co-relation [230]. Then, the fuzzy forests eliminates the features into two steps: Screening step ,and Selection step as shown in Fig 7.1.

In the Screening step, RFE-RF is used on each module to eliminate the least important features within each module. In the Selection step, a final RFE-RF is used on the surviving features with mtry and ntree parameters.

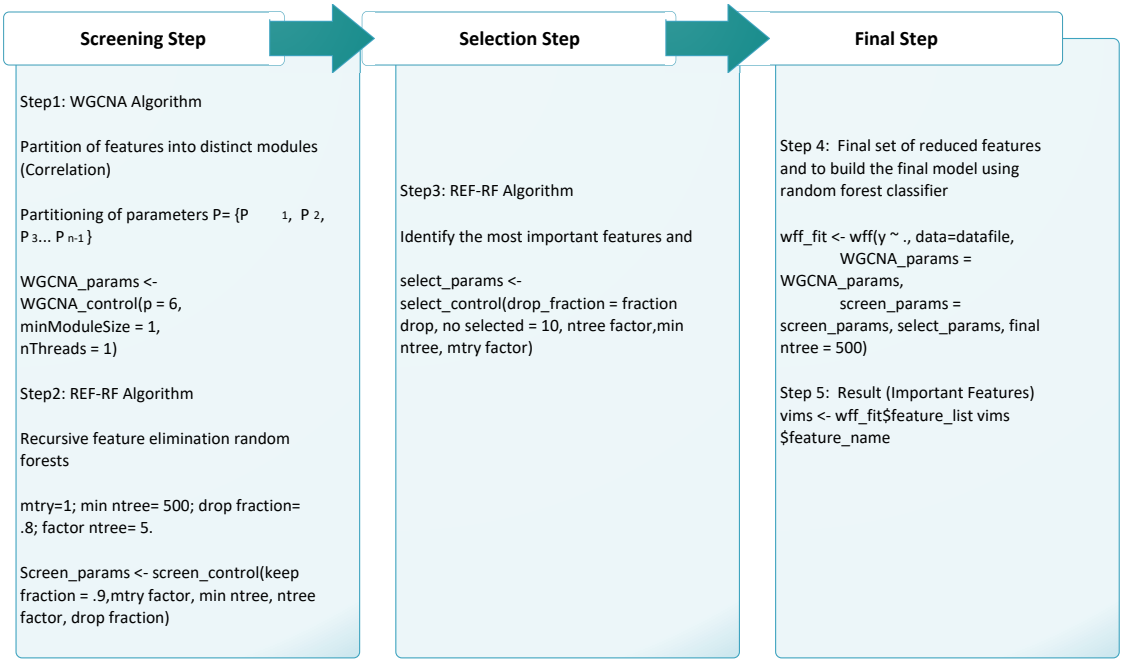

Figure 7.1: Steps of fuzzyforest

7.4 Deep Learning (DL)

"Deep learning (DL) is a subset of machine learning dealing with algorithms inspired by the structure and function of the brain called Artificial Neural Networks ". The soul of DL is its capability to learn multiple levels of abstractions and representations from data [231]. Deep learning requires a huge amount of data to train the model because DL algorithms learn only from data by extracting the features through hierarchical representation. The accuracy of the DL model is dependent on the complexity and size of training data.

In few decade, DL has achieved rapid progress in fields as diverse as computer vision, natural language processing, machine translation language, smart agriculture, health-medical, automatic speech recognition, statistically modeling, and reinforcement learning. In 1943, Warren McCulloch and Walter Pitts developed computer model of the brain using Neural Networks [232]. Ivakhenenko and Lapa applied statistically and supervised models with Polynomial (complicated equations) activation functions, Deep, Feed-Forward, and Multi-Layer Perceptrons [190].

In 1970's, the concept of back propagation is introduced based on errors in Deep Learning models. Zhou and Ji ([233]) proposed a Deep Forest (DF) framework which consists Multi-Layer Deep Learning layers. It combines different ensemble-based methods, including stacking, Random Forests into a structure that is equivalent to a Multi-Layer Neural Network, where each layer consists of RFs instead of neurons.

The efficient modeling of nonlinear functions is performed by Deep Learning. The major benefit of deep hidden layers $x=(x_1, ...x_n)$ where, the input space is high-dimensional. The DL framework considers a uni-variate activation function where, $f_1, \ldots f_n$, applied on each layer (e.g. Cosh, Sin, Sigmoid, Tanh, Heaviside Gate Functions, or Rectified Linear Units) [234]. To determine the suitable parameter for DNN, the model is used with three hidden layers with activation function on each layer including Relu-leak, Eta, and Optimizer.

7.5 Particle Swarm Optimization (PSO) and Deep Neural Network (DNN)

In last few years Auto-ML or Automated Machine Learning has become widely popular to bring the automation in real world. Auto-ML aims to automate methods for model selection, feature extraction and hyper-parameter optimization.

There are two parameters for each machine learning model: Normal parameters are optimized during the training and Hyper-parameters are set manually prior to training by the model designer. The settings of hyper-parameters monitors the behavior of machine learning algorithm. The primary task of Auto-ML is to set

these hyper-parameters automatically to optimize the performance of model [235].

This section presented a new package i.e. Auto-ML for automated machine learning. An Auto-ML package adapts from simple regression to highly tailor-made deep neural networks either with meta heuristic or gradient descent, using custom cost function and automatic hyper-parameters tuning. A mix is motivated by the common tricks on Particle Swarm Optimization (PSO) and Deep Neural Network (DNN).

Particle Swarm Optimization (PSO) is proposed by Kennedy et al. [236]. PSO is one of the population-based meta-heuristic optimization techniques. It is based on the social and individual behavior of biological groups e.g. school of fishes, flocks of birds, and swarms of bees.

Firstly, it prepares a batch of individuals as a population and the state of these individuals is updated with an evolutionary process. After the initialization, an evolution procedure is applied with a definite number of generations, and each particle (individual) finds a possible optimal solution by changing its direction according to its velocity and position of the individual based on best previous experience during each generation. It is easy to implement, and capable to determine the global optimum solution which are closest to the actual solutions and also offers an efficient search and quick procedure. It has been significantly applied in a wide area of fields such as economic problems, control systems, and other real-world applications.

The crucial hyper-parameters for PSO models are (Numiterations, Psopartpopsize) and for DNN model (Hparlayer activation, Mini batch, and Weight decay) parameters used for optimization.

7.6 Model Development for DSS

7.6.1 Estimation of (ET_o)

The Reference evapotranspiration for DSS is same as the Reference evapotranspiration ET_o presented in section 1.2. The estimated results of ET_o, K_c and ET_c are presented in Figure 7.2.

7.6.2 Estimation of (ET_c) and IR Parameters

The crop evapotranspiration for Decision Support System is same as the crop evapotranspiration ET_c presented in section 1.3.

This section presents the procedure for calculating the irrigation requirement, net irrigation in mm and litres, gross irrigation in mm and litres, and pumping discharge time in hours.

The Irrigation Frequency (IF) and ET_o are the most important parameter that

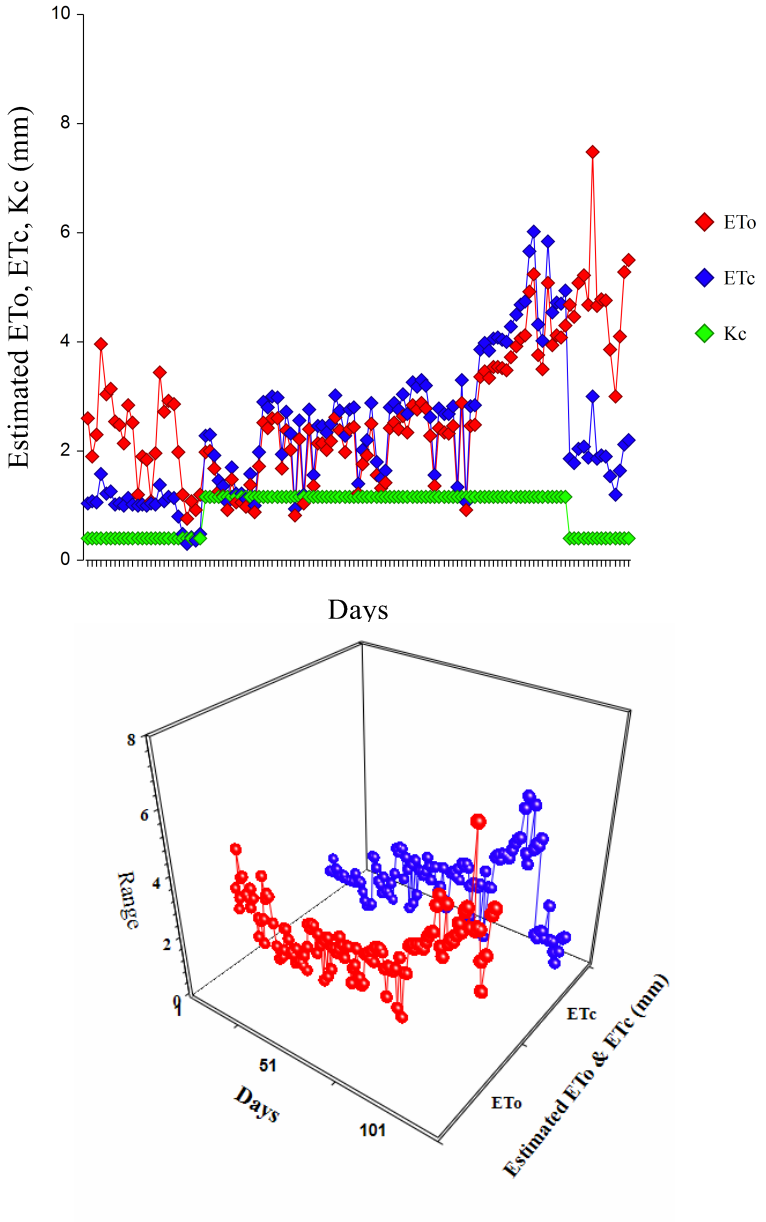

Figure 7.2: Estimation of daily reference evapotranspiration ET_o and crop ET_c

influence the irrigation management system. The need of an irrigation water is defined as the difference between the crop water requirement and effective rainfall. In this study, only effective rainfall factor is considered due to negligible effect of other factors on the requirement of irrigation water, given in Eq. (7.1), and Eq. (7.2):

$$IR_{need} = ET_o = (P_e + G_e + W_b) \tag{7.1}$$

where, W_e signifies water stored in the soil at the beginning of each period (mm), G_e denotes groundwater contribution from water table (mm), P_e is the effective rainfall (mm), IR denotes the Irrigation Requirement (mm), and ET_c is the total

crop evapotranspiration (mm).

$$P_e = P \times \frac{65}{100} \tag{7.2}$$

where, P_e is the effective rainfall (mm), P is the rainfall (mm),

Some rainfall may not be effective due to water loss via surface runoff, deep evaporation or percolation. It depends upon the root zone depth and the soil storage capacity. Generally, in India 50% to 80% of total rainfall is assumed as effective rain. In this study effective rainfall is taken to be 65% of the rainfall. The contribution of groundwater table G_e changes to the E_c due to level of water table below the available water content in the root zone, and soil type. In this study, G_e is taken as zero.

The field Irrigation Efficiency (IE_f) is defined as the volume of removed water from the crop root zone to a given area from the water source. In this study field irrigation efficiencies i.e. 40%, and 65% are taken for Border irrigation and Sprinkler irrigation system respectively as shown by eq (7.3).

$$E_f = \frac{W_{sr}}{W_{ds}} \times 100 \tag{7.3}$$

where,

E_f is the field irrigation system efficiency (%), W_{sr} is the volume of water stored in crop root zone soil, W_{ds} is the volume of water diverted or pumped from the source

The Irrigation Frequency (IF) requirement is calculated by Eq. (7.4):

$$(IF) = \frac{IR_n}{ET_c} \tag{7.4}$$

It is the number of days between two irrigation's consecutively, i= $\frac{d}{ET_c}$, where d denotes the net depth of irrigation application (dose) in mm and ET_c is the daily crop evapotranspiration in mm.

7.6.3 Net Irrigation (Net_{IR})

The Net Irrigation (Net_{IR}) is the requirement of water for the crop growth (mm per day). It is determined by the climate and cropping pattern. The net irrigation depth is identified by examining that how much water is set per irrigation application using local irrigation approach. The net irrigation water requirement for border and sprinkler irrigation system is calculated by Eq. (7.5):

$$\begin{aligned} IR_n &= IR + LR \\ IR_{nv} &= (IR_n * 10^{-3} * W_a * 10^4 * 10^3) \end{aligned} \tag{7.5}$$

where, IR_{nv} is the Net Irrigation Requirement in (litres), W_a is the Net given cropped area (ha), IR_n is the Net Irrigation Requirement in (mm), IR is the Irrigation Requirement (mm), LR is the Leaching Requirement (mm).

The leaching requirement is the additional water required for leaching down the extra salts with irrigation water and makes the EC of soil up to the required level so that it may not harm the crop growth and development [237]. The E_{cw} for Ludhiana station is estimated as 1.5 * E_{cw} [238].

7.6.4 Gross Irrigation (GR_{IR})

The Gross Irrigation (GR_{IR}) requirement is referred as the overall amount of water applied through irrigation. The Net Irrigation requirement plus additional losses and application is the total amount of water required. The (GR_{IR}) requirement is used to evaluate the crop growth losses in different stages, irrigation system and it is determined by Eq. (7.6):

$$\begin{aligned} IR_g &= \frac{IR_n}{E_f} \times 100 \\ IR_{gv} &= \frac{IR_n}{E_f} \times 100 \end{aligned} \tag{7.6}$$

where, IR_n is the Net Irrigation Requirement per day (depth), IR_g is the Gross Irrigation in mm, IR_{gv} is the Gross Irrigation in liters. E_f is the ratio between water that enters and stays in the root zone to meet crop requirements.

7.6.5 Irrigation Pumping (IP_t)

Irrigation Pumping (IP_t) is the maximum number of days that a single irrigation can be applied to an area during the crop's peak consumption period. It is controlled by irrigation system, discharge capacity, equipment and the irrigation interval which is not longer than the irrigation period.

The Pumping Time or irrigation period of border irrigation system per application is determined on the basis of volumetric approach, and the discharge rate i.e. the ratio of volume of container or bucket to time required to fill. At peak demand of irrigation, the minimum flow capacity should matches with the water requirements of the area for any irrigation system.

$$\begin{aligned} IP_t &= \frac{IR_{gv}}{Q} \\ Q &= \frac{V}{T} \times 60 \end{aligned} \tag{7.7}$$

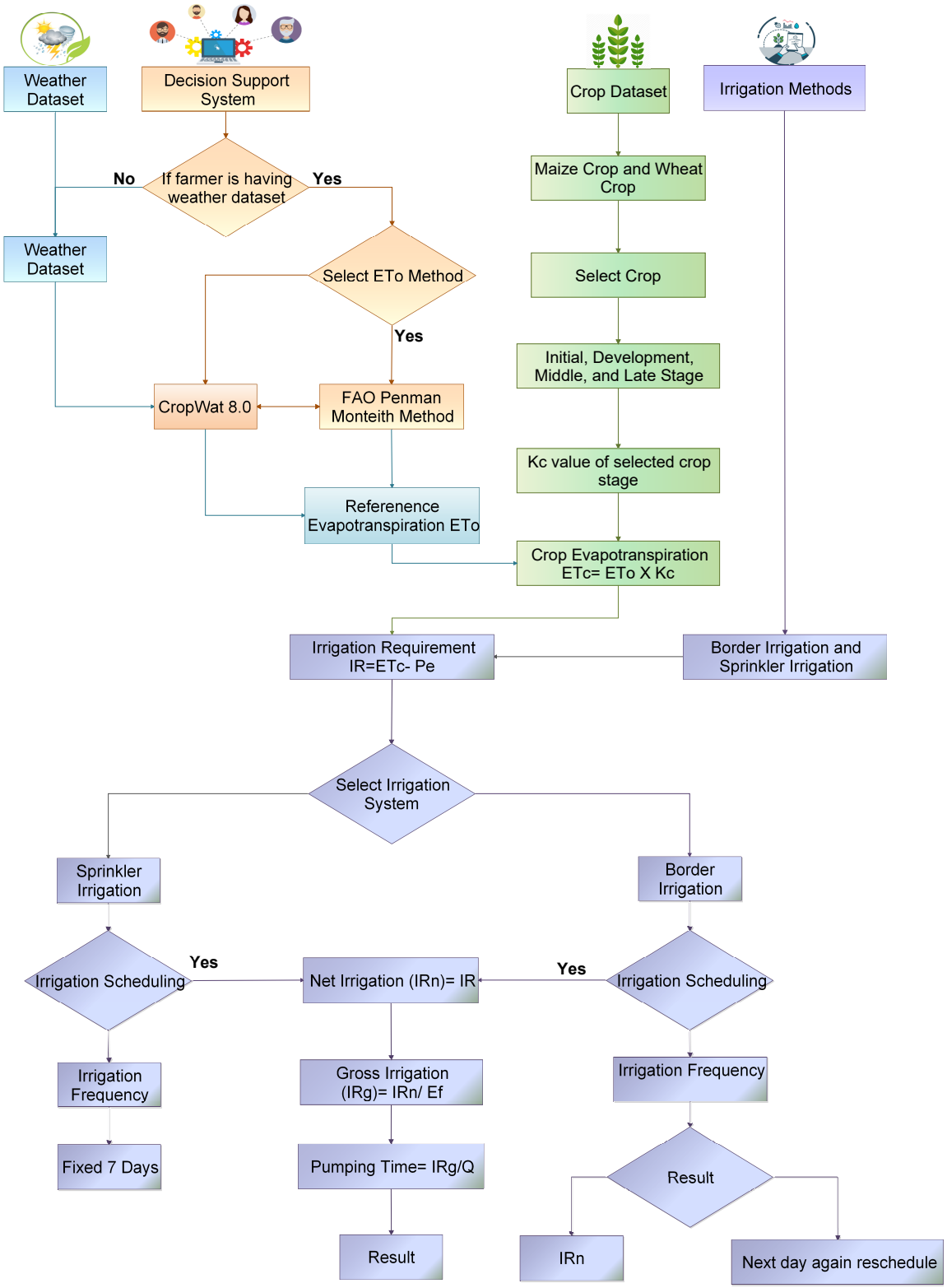

Figure 7.3: Flowchart of decision support system for irrigation scheduling

where, IP_t is referred as irrigation period (hours), Q is a discharge rate of pump (lph), IR_{gv} is a gross irrigation required in term of volume (liters), T is a time required to fill-up container or bucket (minute), V is referred as volume of container or bucket (liters).

In case of border irrigation system, it is determined as when gross irrigation depth is less than equal to previous depth of irrigation. For sprinkler system, the irrigation frequency is taken as 7 days.

7.6.6 Proposed Model

Fig 7.3 presents the flow chart of proposed decision support system for on-farm irrigation scheduling to estimating reference evapotranspiration, crop coefficient, crop evapotranspiration, and irrigation methods.

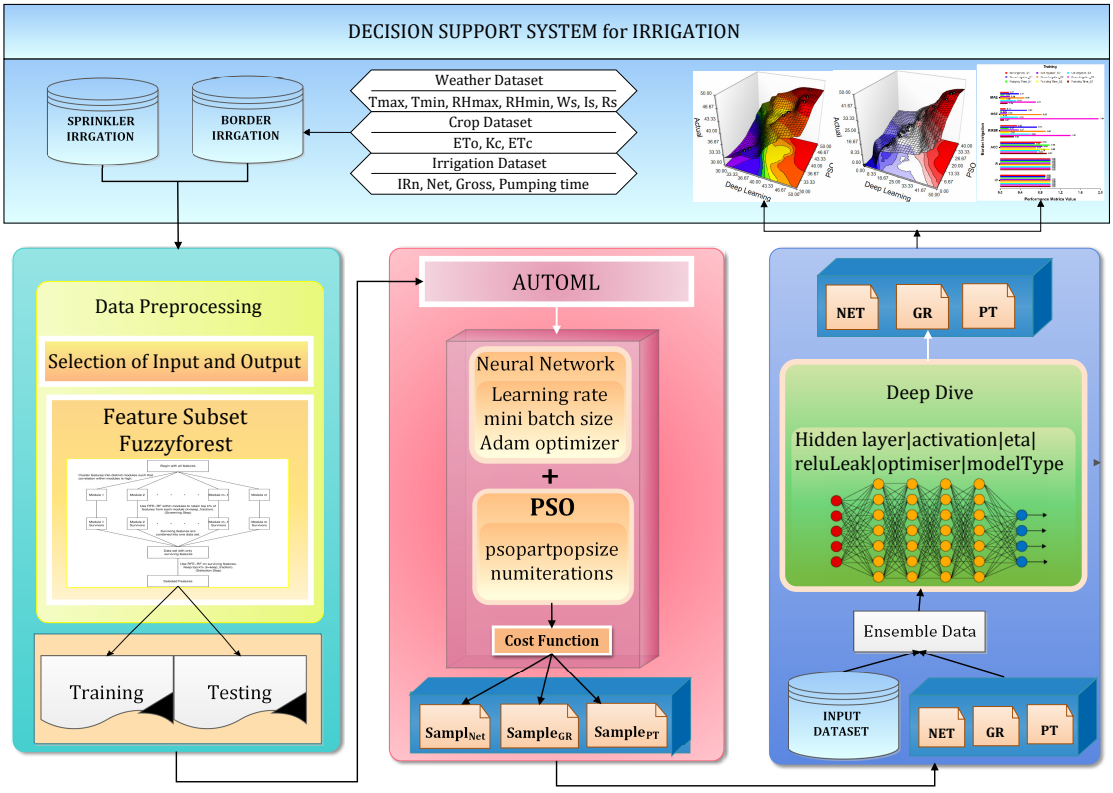

Figure 7.4: Framework of smart DSS-IS

In this process, the first module is about weather dataset selected for this study and FAO-56 Penman Monteith method is used to estimate ET_o for DSS system. After estimating ET_o, crop dataset module is applied to select the crop variety and their stages. Then the crop evapotranspiration is estimated based on ET_o and Kc values of wheat crop. Border and sprinkler irrigation datasets are used to estimate irrigation frequency and their parameters such as irrigation requirement, net irrigation, gross irrigation and irrigation pumping time. These parameters are

considered in this study as case study for further process of ensemble based machine learning modeling.

Fig 7.4 shows the schematic diagram of the proposed ensemble model for DSS-IS using two irrigation methods. It presents the key parts of the proposed DSS-IS model, whose goal is to achieve the accurate forecasting of given target values for particular crop. The proposed model is described in few steps as follows.

The model development is divided into sub-parts: Data collection, Data pre-processing, Data analysis and Ensemble modeling and results.

1. ***Data Collection*:** The data collection is described in previous chapters, where weather data is collected from IMD, Pune and for crop water modeling the case study has been selected from PAU, Ludhiana. The two irrigation scheduling methods are selected for this case study: Border Irrigation and Sprinkler Irrigation.

2. ***Data Pre-processing and Estimation*:** The Reference Evapotranspiration (ET_o) is estimated using FAO-56 Penman Monteith method. The crop coefficient (K_c) is described in Table 7.2, where wheat crop has been selected for case study. The equation (1.2) is applied for the estimation of crop evapotranspiration (ET_c).

   Data Pre-processing covers the data cleaning and normalization where the feature vectors of dataset are normalized to zero mean and unit variance. In Machine Learning, data normalization is a crucial step in the process. It is important to convert all features from vector space to unit space. All input variables are normalized between 0 and 1 before the training of each model.

3. ***Feature Subset*:** The fuzzy forest model is a novel machine learning model that applied for feature subset. This model is used to extract the highly correlated features from the given dataset. Initially, there are 21 features as input and (22, 23, 24) three target values. The 10 features are selected for Border Irrigation and 10 inputs for Sprinkler Irrigation respectively.

4. ***Data Division*:** Border Irrigation and Sprinkler Irrigation datasets are separated individually in training and testing subsets. For 1 year of full crop dataset, it is separated further into two subset of training and testing to assure the generalization ability of model.

5. ***Modeling and Analysis*:** In this step, analysis and ensemble modeling process is applied on training dataset for both irrigation methods.

   Stage I: In this step, each subset of training dataset is used to train the PSO-DNN model using auto-ml package. Then the testing dataset is applied to

validate the model accuracy which generate the individual forecasting results. All the predicted results are evaluated by performance metrics to report their error.

Stage II: The Deep Learning model is selected to aggregate the individual forecasts to obtain the ensemble results. However, to obtain the desired results, the higher accurate predicted values were selected and combined with the previous training dataset to train the deep learning model by using ensemble model. Then, the Deep Learning model (DL-Ensemble) trained with the given ensemble model to predict the three target values such as Net Irrigation, Gross Irrigation and Pumping Time.

6. ***Evaluate Performance*:** The best predicted solution selected in the previous step is used to forecast the three targets on the testing dataset. Further, the error and accuracy of the proposed model is evaluated using performance metrics discussed in result section.

The decision support system is based on the concept of forecasting the water need of crops in order to properly irrigate them. Traditionally, an expert farmer or farming technician make this decision. In this system, agronomist is analyzing information from various sources: Weather Stations (Meteorological data), Crop and Soil characteristics (Type, Variety, Size, etc.) in the crop fields. Based on these parameters two irrigation scheduling methods were applied to calculate the Irrigation Requirement, Net irrigation, Gross Irrigation and Pumping Time. These calculated parameters are selected for prediction purpose by using ensemble based machine learning. To make this decision making process, this information is needed to create the irrigation need for specific crop.

7.7 Results and Discussions

7.7.1 Performance Metrics

In this section, the six statistical indicators have been selected to investigate the performance of models such as using Pearson Correlation (r), Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Mean Square Error (MSE), Accuracy (ACC) and Coefficient of Determination(R^2).

(i) Mean absolute error

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |IR_{actual} - IR_{predict}| \tag{7.8}$$

(ii) Mean square error

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(IR_{actual} - IR_{predict}^{2}) \tag{7.9}$$

(iii) Root mean square error

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(IR_{actual} - IR_{predict})^{2}} \tag{7.10}$$

(iv) Pearson correlation coefficient

$$(r) = \frac{\sum_{1}^{N}(IR_{actual} - \overline{IR_{actual}})(IR_{predict} - IR_{predict})}{\sqrt{\sum_{1}^{N}(IR_{actual} - \overline{IR_{actual}})^{2}\sum_{1}^{N}(IR_{predict} - IR_{predict})^{2}}} \tag{7.11}$$

(v) Coefficient of determination

$$r^{2} = r * r \tag{7.12}$$

(vi) Accuracy

$$\begin{aligned} ACC &= \left(\frac{\sum_{i=1}^{M} IR_{n}}{M} \times 100\right) \\ IR_{n} &= \begin{cases} 1 \; if \; |IR_{actual} - IR_{predict}| < \; error \\ 0 \quad otherwise \end{cases} \end{aligned} \tag{7.13}$$

where, IR is Irrigation Requirements, IR_{actual} is observed/actual values where three target (Net_{IR}, $Gross_{IR}$, and PT_{IR}), M is referred as total no of data vectors, and $IR_{predict}$ as predicted values.

Performance of ensemble model during model development phase: Training period based on six metrics for Border and Sprinkler Irrigation is shown in Table 7.3, and Table 7.4. Similarly, the performance of ensemble model in forecasting of the targets during the testing period based on six metrics for Border and Sprinkler irrigation in Table 7.5 and Table 7.6.

The simulation is carried out to evaluate the performance of the PSO-DNN and DL models of Net_{IR}, $Gross_{IR}$, and PT_{IR} estimation are given in Table 7.3 and 7.4 for Border and Sprinkler Irrigation Systems.

The performance indices, including MAE, MSE, RMSE, R^2, R, and ACC are used for calculating the high-accuracy of deployed models of daily prediction. Here, R^2 and ACC present with largest values and RMSE, MSE, and MAE with lowest values in terms of mean higher model efficiency.

Table 7.3: Performance of DL-ensemble model during model development phase: training period based on six metrics for border irrigation

Models Name	MAE	MSE	RMSE	ACC %	R	R^2
Net irrigation$_{S1}$	0.1715	0.1093	0.3306	82	99	100
Gross Irrigation$_{S1}$	0.3736	0.5338	0.7306	69	99	100
Pumping Time$_{S1}$	0.0823	0.0277	0.1664	99	99	100
Net irrigation$_{S2}$	0.1928	0.1367	0.3697	84	99	100
Gross Irrigation$_{S2}$	0.4782	0.8275	0.9096	71	99	100
Pumping Time$_{S2}$	0.0946	0.0316	0.1777	99	99	100
Net irrigation$_{S3}$	0.3202	0.3593	0.5995	79	99	100
Gross Irrigation$_{S3}$	0.7100	1.9602	1.4001	74	99	100
Pumping Time$_{S3}$	0.1486	0.0741	0.2722	93	99	100

Table 7.4: Performance of DL-ensemble model during model development phase: training period based on six metrics for sprinkler irrigation

Models Name	MAE	MSE	RMSE	ACC%	R	R^2
Net irrigation$_{S1}$	0.2765	0.2163	0.4651	76	99	100
Gross Irrigation$_{S1}$	0.2875	0.2420	0.4920	78	99	100
Pumping Time$_{S1}$	0.3359	0.2677	0.5174	70	99	100
Net irrigation$_{S2}$	0.3112	0.3560	0.5966	76	99	100
Gross Irrigation$_{S2}$	0.5216	0.8535	0.9238	69	99	100
Pumping Time$_{S2}$	0.5230	0.7809	0.8837	69	99	100
Net irrigation$_{S3}$	0.2589	0.2686	0.5183	86	99	100
Gross Irrigation$_{S3}$	0.3571	0.5599	0.7483	82	99	100
Pumping Time$_{S3}$	0.2714	0.4686	0.6845	86	99	100

For feature selection: minimum number of tree as 500, final ntree as 500, ntry factor as 5, drop fraction as 0.8, mtry factor as 1, keep fraction as 0.05, number selected as 10 have been found for best performance using fuzzyforest with R Language.

Table 7.5: Performance of DL-ensemble model in forecasting of the targets during the testing period for border irrigation

Models Name	MAE	MSE	RMSE	ACC%	R	R^2
Net irrigation$_{S1}$	0.1408	0.0901	0.3002	84	99	100
Gross Irrigation$_{S1}$	0.3977	0.5925	0.7698	69	90	100
Pumping Time$_{S1}$	0.0885	0.0341	0.1846	99	99	100
Net irrigation$_{S2}$	0.2115	0.1421	0.3770	76	99	100
Gross Irrigation$_{S2}$	0.5031	0.8105	0.9003	69	90	100
Pumping Time$_{S2}$	0.0923	0.0299	0.1729	99	99	100
Net irrigation$_{S3}$	0.3677	0.4453	0.6673	76	99	100
Gross Irrigation$_{S3}$	0.1230	0.8240	2.8720	76	99	100
Pumping Time$_{S3}$	0.2269	0.2311	0.4808	84	99	100

In the proposed approach, the crucial hyper-parameters PSO (modexec as 'trainwpso', numiterations as 200, psopartpopsize as 220) and DNN (hparlayer activation

Table 7.6: Performance of DL-ensemble model in forecasting of the targets during the testing period for sprinkler irrigation

Models Name	MAE	MSE	RMSE	ACC%	R	R^2
Net irrigation$_{S1}$	0.2924	0.2040	0.4517	69	90	99
Gross Irrigation$_{S1}$	0.3199	0.2838	0.5327	78	99	100
Pumping Time$_{S1}$	0.3289	0.2411	0.4910	70	99	100
Net irrigation$_{S2}$	0.3134	0.3701	0.6084	73	99	100
Gross Irrigation$_{S2}$	0.4492	0.6278	0.7924	69	90	99
Pumping Time$_{S2}$	0.4404	0.6193	0.7870	77	99	100
Net irrigation$_{S3}$	0.1843	0.1487	0.3856	91	99	100
Gross Irrigation$_{S3}$	0.2573	0.3465	0.5886	86	99	100
Pumping Time$_{S3}$	0.1688	0.1861	0.4314	91	99	100

as (relu, linear), mini batch, and weight decay default set) are considered for optimization.

To determine the suitable parameter for DNN, the moderate sized architecture is used with three hidden layers containing (10,8,3) with activation function (Sin, Sin, Sin) on each layer, relu-leak as 0.01, model type as regress, iteration as 1000, eta as 0.01 and optimizer as "Adam". Each experiment is executed till 1000 epochs.

7.7.2 Trends in ET_o and ET_c

The reference evapotranspiration and crop evapotranspiration for trends in the climate meteorological variables are same as the trends of reference evapotranspiration ET_o presented in section 6.4.1. The seasonal ET_o is shown in Figure 6.3 and Table 6.1 to give more climatic condition of the study area in details.

7.7.3 Border Irrigation

According to the best sample results, Initially the Net_{IR} as 200.62 mm/day, $Gross_{IR}$ as 501.55 mm/day, and PT_{IR} as 83.2 hours were considered for training dataset using Border irrigation. It is observed that the $Sample_8$ presented the better results for ($Net_{IR} = 200.34$ and $PT_{IR} = 83.22$) and $Sample_3$ provided accurate results of ($Gross_{IR}$= 500.2). A 3D surface plot representation of Net_{IR}, $Gross_{IR}$, and PT_{IR} estimation of three samples using actual, PSO-DNN, and DL-Ensemble are shown in Fig 7.5, Fig 7.6 and Fig. 7.7 for Border Irrigation.

It is observed that the evaluation parameters based on the $Sample_2$ provides best accuracy in terms of (MAE= 0.1928, MSE=0.1367, RMSE=0.3697, ACC=84.62) for Net_{IR}, the $Sample_3$ provided better results in terms of (MAE=0.7100, MSE= 1.9602, RMSE=1.4001, ACC=74.42) for $Gross_{IR}$, and the $Sample_1$ provided better results in terms of (MAE=0.0823, MSE= 0.0277, RMSE=0.1664, ACC=100) for PT_{IR} during training period respectively.

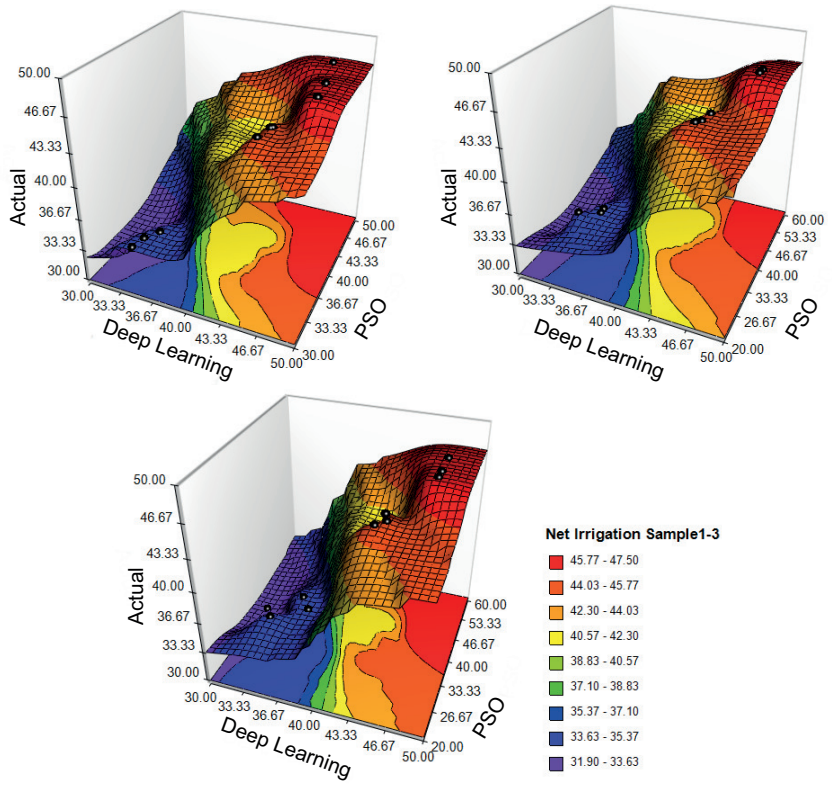

Figure 7.5: Sample$_{1-3}$ of net irrigation for border irrigation

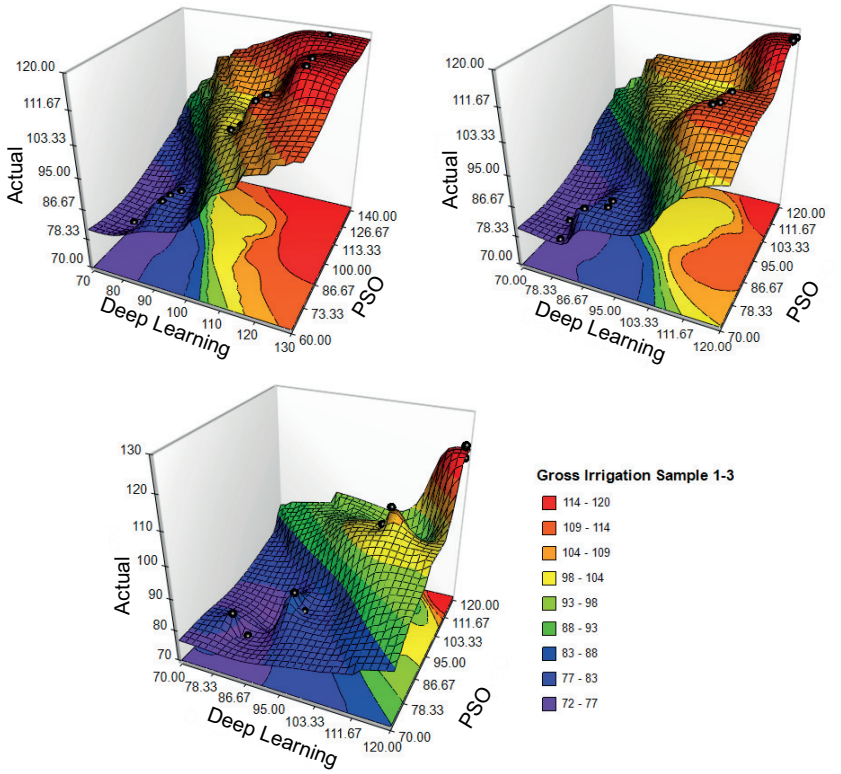

Figure 7.6: Sample$_{1-3}$ for gross irrigation for border irrigation

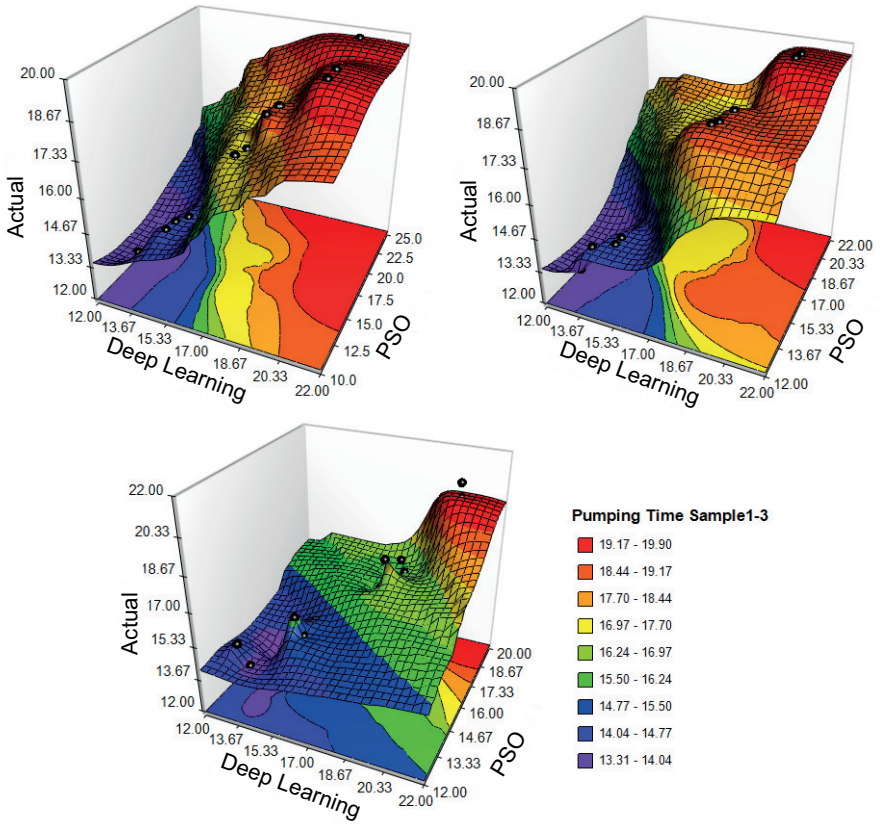

Figure 7.7: Sample$_{1-3}$ for pumping time for border irrigation

Similarly, the other results shows that the proposed model provides best accuracy for Net$_{IR}$ with Sample$_1$ in terms of (MAE=0.1408, MSE=0.0901, RMSE= 0.3002, ACC=84%), for Gross$_{IR}$ with Sample$_3$ in terms of (MAE=0.1230, MSE=0.8240, RMSE= 2.8720, ACC=76%), and for PT$_{IR}$ with Sample$_2$ in terms of (MAE=0.0923, MSE=0.0299, RMSE=0.1729, ACC=99%) during testing period respectively.

Also, it is observed that Sample$_1$ and Sample$_2$ provides highest accuracy as 99% for PT$_{IR}$ during testing period as compared to other Samples.

7.7.4 Sprinkler Irrigation

According to the best sample results, initially the (Net$_{IR}$ as 205.59 mm/day, Gross$_{IR}$ as 315.805 mm/day, and PT$_{IR}$ as 287.175 hour) values were considered for training dataset using Sprinkler Irrigation. It is observed that the Sample$_3$ provides the best results for Net$_{IR}$ = 284.9 mm, and Sample$_3$ provides the accurate results of (Gross$_{IR}$= 315.2, and PT$_{IR}$=287.05).

Similarly, 3D surface plot representation of Net$_{IR}$ Gross$_{IR}$, and PT$_{IR}$ estimation of three samples using actual, PSO-DNN and DL-Ensemble are shown in Fig 7.8, Fig 7.9 and Fig. 7.10 for Sprinkler Irrigation.

It is observed based on evaluation parameters, the Sample$_3$ provides best accu-

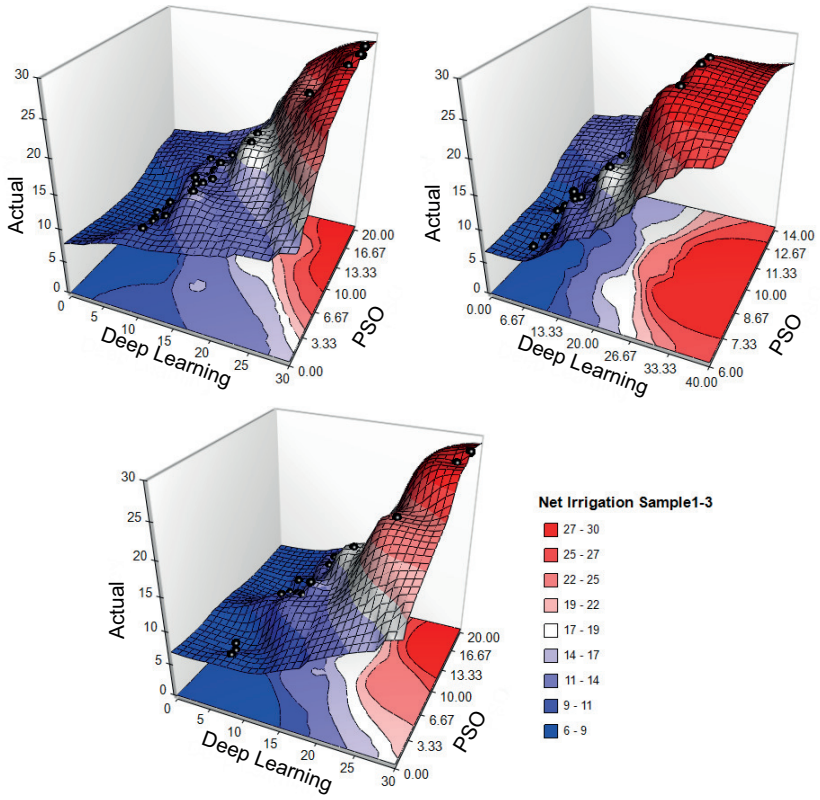

Figure 7.8: Sample$_{1-3}$ for net irrigation of sprinkler irrigation

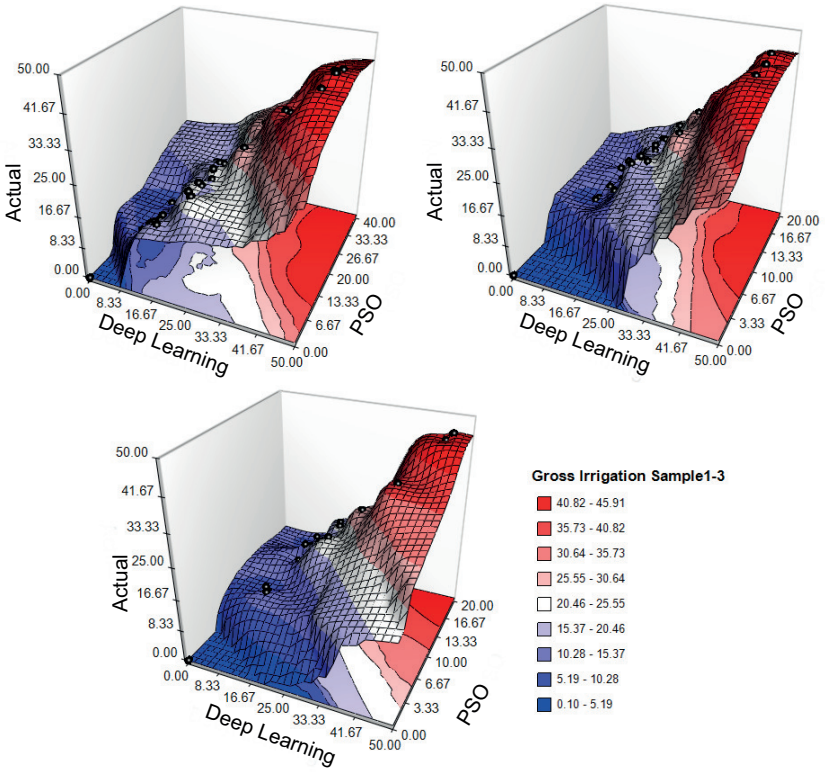

Figure 7.9: Sample$_{1-3}$ for gross irrigation of sprinkler irrigation

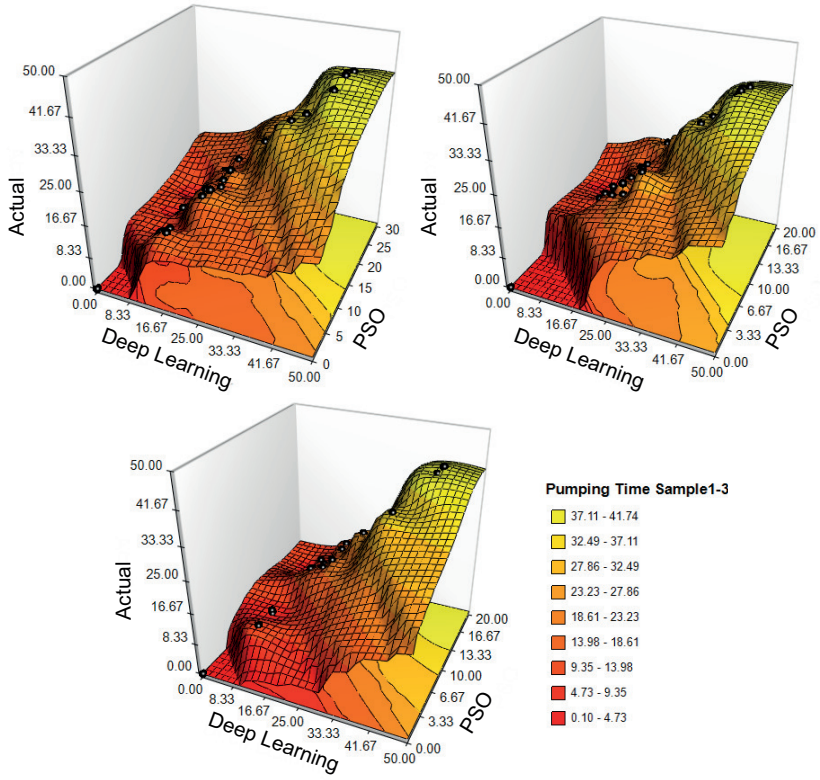

Figure 7.10: Sample$_{1-3}$ for pumping time of sprinkler irrigation

racy in terms of (MAE= 0.2589, MSE= 0.2686, RMSE= 0.5183, ACC= 86.21%) for Net$_{IR}$, provides better results in terms of (MAE= 0.3571, MSE= 0.5599, RMSE= 0.7483, ACC=82.76%) for gross irrigation, and also provides better results in terms of (MAE= 0.2714, MSE= 0.4686, RMSE= 0.6845, ACC= 86%) for pumping time during training period respectively.

Similarly, the other results shows that the proposed model provides best accuracy for Net$_{IR}$ with Sample$_3$ in terms of (MAE=0.1843, MSE=0.1487, RMSE= 0.3856, ACC=91%), for Gross$_{IR}$ with Sample$_3$ in terms of (MAE= 0.2573, MSE=0.3465, RMSE= 0.5886 , ACC= 86%), and for PT$_{IR}$ with Sample$_3$ in terms of (MAE=0.1688, MSE= 0.1861, RMSE= 0.4314, ACC= 91%) during testing period respectively.

Also it is observed that Sample$_3$ provides highest accuracy as 91% for NET$_{IR}$ and PT$_{IR}$ during testing period than other Samples.

7.7.5 Comparison

Initially the case study parameters (Gross$_{IR}$= 365.95 mm, Gross$_{IR}$= 3659505.00 liters, and PT$_{IR}$= 60.99 hours) for Border Irrigation using Wheat crop have been considered and the simulated results were found superior in terms of (Gross$_{IR}$= 731901 liters, and PT$_{IR}$= 12.20 hours) average volume of water required and time of pumping required per irrigation.

These estimation were forecasted based on pre-decided applied depth of irrigation as 45 mm, soil type as sandy loam, net cropped area as 1 hectare, volume of container with 1000 liters and pumping time taken to fill up the tank with 1 minute.

Table 7.7: The estimated results of border irrigation

Border Irrigation Parameters	Case study	Results
Pumping discharge rate	83.591 lph	83.05 lph
Total Irrigation	5	5
Total gross depth of irrigation in mm	501.55 mm	500.52
Total volume of water required for 1 ha	50155000	5005200
Total time of pumping	83.2	83.05
Average volume of water required per irrigation	1003100 l	10,01,040 l
Average time of pumping required per irrigation	16.64 h	16.61 h

Table 7.8: The estimated results of sprinkler irrigation

Sprinkler Irrigation Parameters	Case study	Results
Pumping discharge rate	11000 lph	11000 lph
Total Irrigation	15	14
Total gross depth of irrigation in mm	315.805	315.256
Total volume of water required for 1 ha	3158059892	3152566
Total time of pumping	287.175 h	287.053 h
Average volume of water required per irrigation	225575.64 l	225183.28 l
Average time of pumping required per irrigation	20.5125 h	20.503 h

Table 7.7 shows the names of parameters that are considered to build the decision support system for Border Irrigation. Initially, these parameters are taken to calculate the Border and Sprinkler Irrigation system, based on case study parameters. Then, these calculated parameters are used to train the model using new dataset and forecast the results. The sample results presented the forecasting results using proposed method.

We have applied these parameters to calculate the Border Irrigation parameters to build the new dataset for modeling, the calculated results are Gross_{IR}= 501.55 mm , and PT_{IR}= 83.2 hours. Then these parameters are decided as final case study to carry out the modeling. The sample results of the Border and Sprinkler Irrigation are shown in Table 7.7 and Table 7.8.

Figure 7.11, and Figure 7.12 presents the combo box plot, density plot and dot plot representation of estimated Net_{IR} and Gross_{IR} of Border Irrigation. This figure presents the information about the distribution of values assumed by all observations. Box-plots are particularly useful when comparing several subgroups, where it presents the median, quartiles, symmetry, skewness, and outliers.

In Fig 7.11, the box plot presents the estimated (Net Irrigation) with actual, PSO-DNN, and DL-Ensemble models of three samples. The each sample presents

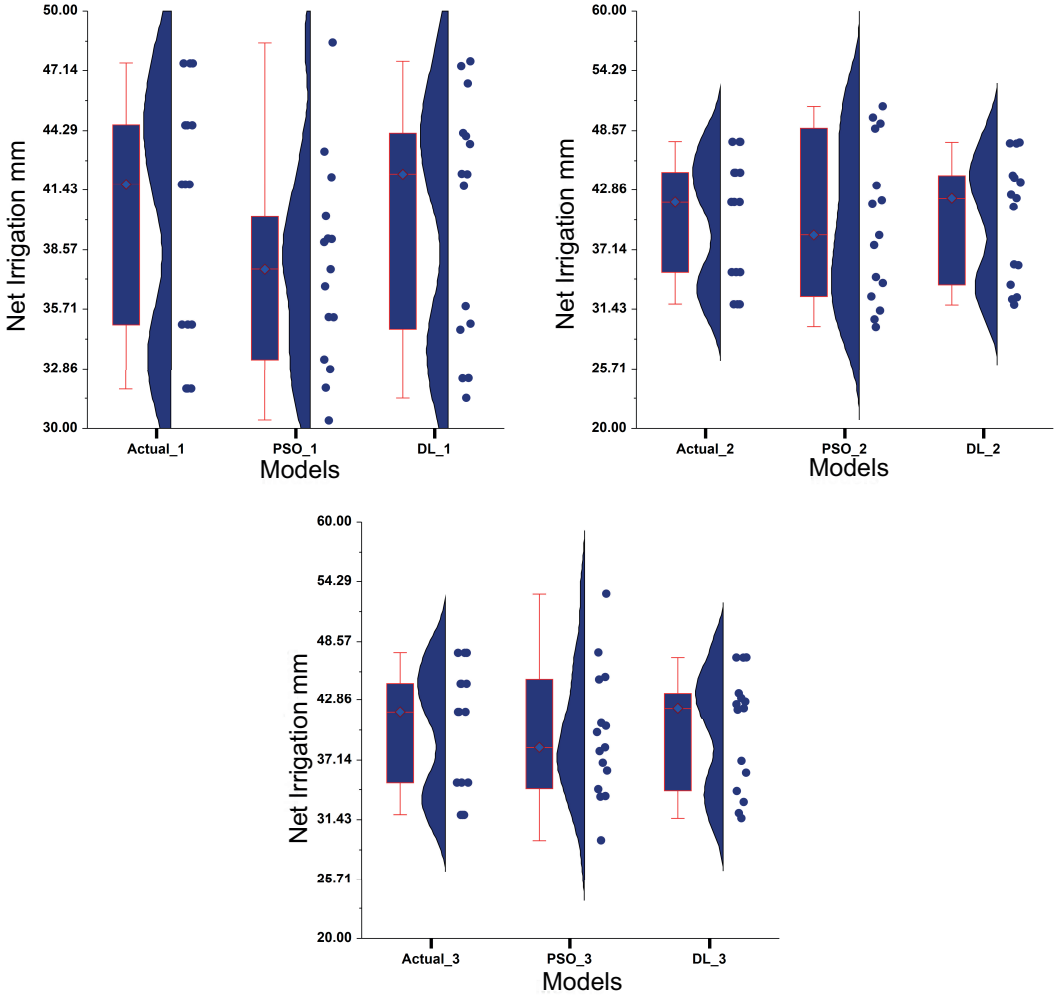

Figure 7.11: Combo box plot, density plot and dot plot for net irrigation

the comparison between Actual, PSO-DNN and DL-Ensemble model, where it is observed that the distribution of each samples accurately predicted using DL model. The median of PSO box plot model illustrates the difference between the other two groups. The median line of a box plot for the PSO model using Sample 1 is outside the box of a comparison box plot (with actual and DL-Ensemble model). However, the actual and DL-Ensemble models provides the negative symmetric results. The smoothened distribution of points along the numeric axis is shown by the Density Plot. It is an estimation of the frequency distribution based on the sample data. The peaks of the density plot are at the area where there is the highest concentration of data points. The figure 7.13 and 7.14 presented the performance metrics of estimated Net_{IR} Gross_{IR}, and PT_{IR} during training and testing results for Border and Sprinkler Irrigation.

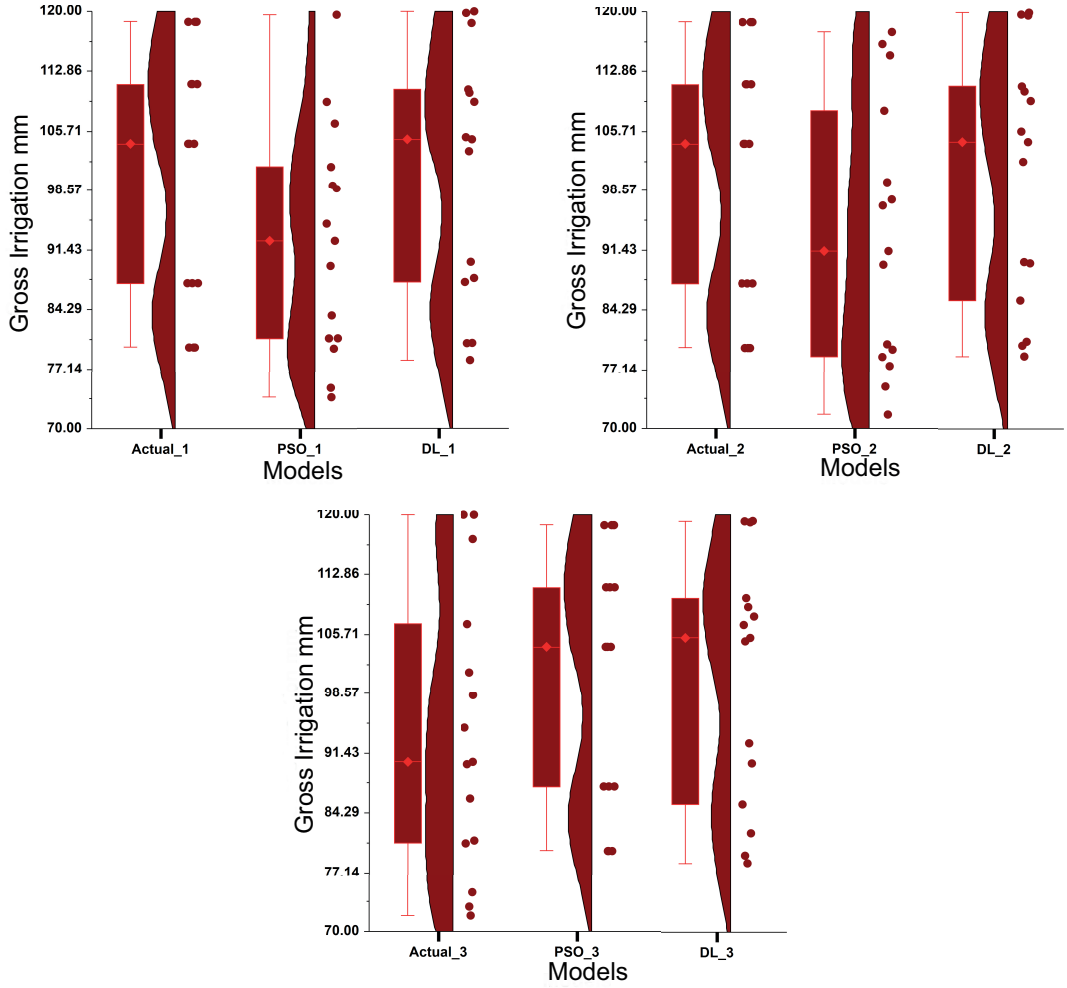

Figure 7.12: Combo box plot, density plot and dot plot for gross irrigation

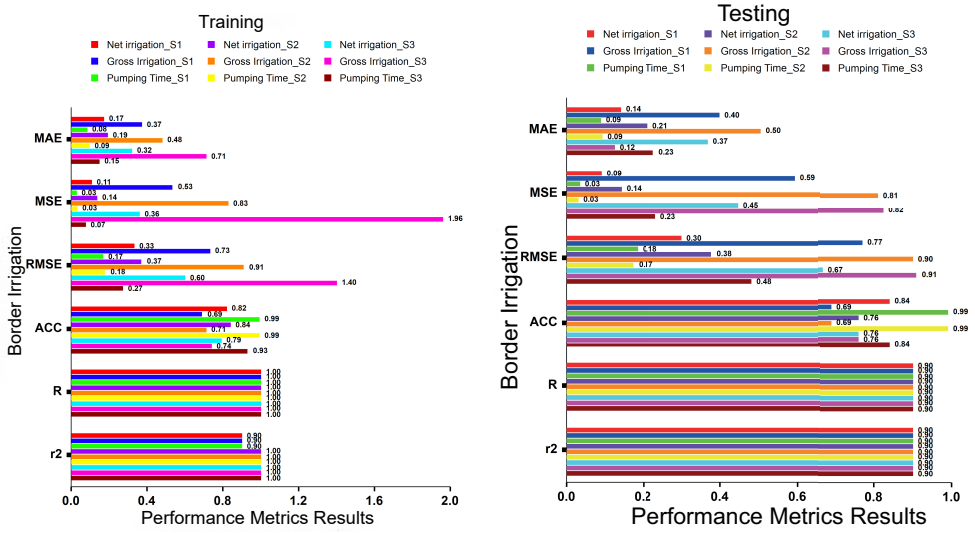

Figure 7.13: Performance metrics presented using box plot for Border Irrigation

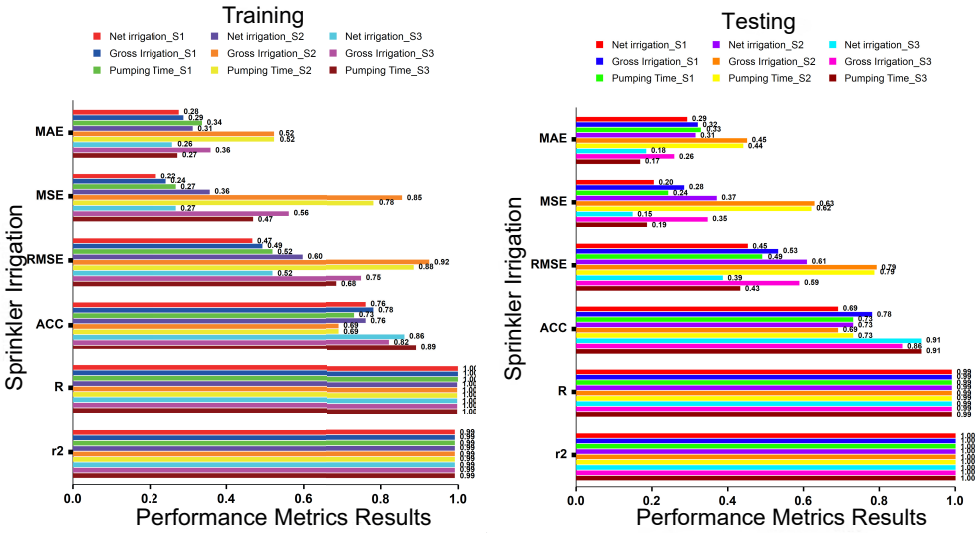

Figure 7.14: Performance metrics presented using box plot for Sprinkler Irrigation

7.8 Conclusion

In this chapter, we have developed DSS for Irrigation Scheduling using Border and Sprinkler Irrigation parameters. The research have been conducted for Wheat crop based on case study of Ludhiana station. The climate dataset is used to estimate the reference evapotranspiration ET_o and crop evapotranspiration ET_c. The Irrigation Water Requirement parameter is calculated using case study based parameters. We have divided these datasets into three samples for feature selection using fuzzy forest. PSO-DNN and DL-Ensemble models are applied to estimate the Net_{IR}, $Gross_{IR}$, and PT_{IR} parameters using training and testing dataset.

- The fuzzy-forest model has been used to select the top 10 features from the input features.
- PSO and DL models are applied to estimate the Net_{IR} $Gross_{IR}$, and PT_{IR} parameters using training and testing dataset. The optimal irrigation results were found as 500 mm and 315 mm for Wheat crop of Border and Sprinkler Irrigation System.
- According to the best sample results, Initially the Net_{IR} as 200.62 mm/day, $Gross_{IR}$ as 501.55 mm/day, and PT_{IR} as 83.2 hours are considered for training dataset using Border Irrigation. It is observed that the $Sample_8$ presented the better results for (Net_{IR} = 200.34 mm, and PT_{IR} = 83.22) and $Sample_3$ provided accurate results of ($Gross_{IR}$= 500.2).
- According to the best sample results, Initially the Net_{IR} as 205.59 mm/day, $Gross_{IR}$ as 315.805 mm/day, and PT_{IR} as 287.175 hours are considered for

training dataset using Sprinkler Irrigation. It is observed that the Sample_3 provided the best results for Net_{IR}= 284.9 mm, and also provided the accurate results of (Gross_{IR}= 315.2, and PT_{IR}= 287.05).

- The newly forecasted case study parameters (Gross_{IR}= 501.55 mm, Gross_{IR}= 50155000 liters, and PT_{IR}= 83.2 hours) for Border Irrigation were considered for modeling and their predicted results were found superior as average volume of water required and time of pumping required per irrigation (Gross_{IR}= 1003100 liters, and PT_{IR}= 16.61 hours).

In Machine Learning, ensemble-based methods are commonly employed to improve the accuracy of regression models' predictions. This study investigated the performance of proposed model, where DL-Ensemble gives accurate results for Border and Sprinkler irrigation system. Our proposed ensemble-based approach investigated that it can generate the accurate ensemble based outcomes in agriculture domain.

Chapter 8

Conclusions and Scope for Further Work

This chapter summarizes the major contributions of the research work and list the scope to extend it. The research work has been carried out to develop the methodology to determine the accurate crop water requirement of Wheat and Maize crops for Hoshiarpur, Patiala and Ludhiana stations. To address this objectives, a Machine Learning and Deep Learning based approaches have been developed to predict the reference evapotranspiration (ET_o) and crop evapotranspiration (ET_c) in Smart Irrigation Scheduling. Section 8.1 presents the the summary of main findings of the research work. Section 8.2 presents the scope of possible extension of the current work.

8.1 Summary of Important Findings

In Chapter-3, the four data-driven models (DL, GBM, GLM, and RF) have been developed under H2O framework for evaluating daily ET_o at Hoshiarpur and Patiala sites in India. The five-fold cross-validation test has been deployed to estimate the performance of considered models. The following conclusions are drawn from the study-

- The combination of six meteorological variables i.e. T_{min}, T_{max}, R_H, u_2, I_s and R_s, been used as input variables. The missing values have been filled appropriately using MissForest for estimation of daily ET_o for Hoshiarpur and Patiala.

- The newly developed DL model showed great capabilities for ET_o estimation and performed much better than other standard existing models. The Deep Learning model performed very well and showed very accurate results in comparison of RF, GBM and GLM.

- The DL model has avoided the over-fitting issue. It has shown higher robustness than conventional approaches.

- The DL model presents high performance for modeling daily ET_o (e.g. NSE= 0.95-0.98, r^2= 0.95-0.99, ACC= 85-95, MSE= 0.0369-0.1215, RMSE= 0.1921-0.2691).

- The performance of deep learning model has been analyzed as 5360 K weights and 5320 K weights model complexity.

In Chapter-4, the Matrix Product State (MPS) is used as quantum classifier to predict the ET_o of Patiala station. The key advantage of classification with MPS quantum circuit is that it can be executed efficiently with small number of qubits. Moreover, the encoding process of classical data can be done using qubits. The following conclusions are drawn from the study-

- In order to analyze the performance of MPS quantum classifier, the dataset is divided into three samples on the basis of pairwise combination of class labels.

- The mapping of classical data into MPS form is beneficial for generating high-order correlations between classes. The bond dimension of MPS manages the parameters of the machine learning model. The larger dimension of bond results in higher accuracy. Even, over-fitting is not observed with an extremely large bond dimension.

- MPS quantum classifier has been trained to avoid both over-fitting and under-fitting. It has dealt with corrected predictors and reduced the variance of the prediction error. It has shown great learning capability for ET_o estimations in Agri dataset.

In Chapter-5, a Multilevel framework is proposed to determine the daily ET_o for Patiala using 38 years of dataset during (1970-1990, 1993-1999, to 2007-2016) from IMD (Pune). The following conclusions are drawn from the investigations-

- Multi-Level Ensembling ET_o (MLE-ET_o) forecasting model has been developed on the basis of the three machine learning models such as ELM, SVM and MLP.

- The performance of developed models is evaluated over seven statistical error parameters. It is observed that classification or regression accuracy can be improved by blending multiple homogeneous models using an ensemble approach.

- Comparison of model results show that the Ensemble SVM model is more accurate. SVM is good at improving generalization ability. The estimated ET_o generally fit the observed ET_o very well.

- The $Ensemble_{SVM}$ model gives the best accuracy results (MSE= 0.0073, 0.0084 and 0.0087, RMSE= 0.0085, 0.0919 and 0.0935, ACC= 99.7, 99.5 and 99.4, NRMSE= 5.70%, 6.20% and 6.20%, NSE= 0.99, 0.99, and 0.99) for (50%, 30%, and 20%) splitting of training, validation and testing datasets respectively.

In Chapter-6, a multi-ensemble based approach is used to predict the crop coefficient (K_c) and crop evapotranspiration (ET_c) for three crops namely Maize, $Wheat_1$ and $Wheat_2$. The proposed approach presents two models namely Fuzzy-Genetic (FG), and Regularized Random Forest (RRF) for estimating value for K_c and ETc of Ludhiana station. The following conclusions are drawn from the investigations.

- The proposed ensemble model estimated accurately (K_c) and (ET_c) for Maize crop.

- The (K_c) value estimated through Fuzzy-Genetic approach helps in correctly predicting Initial, Development, Middle and Last stage of crops.

- The climatic and crop parameters ($Temperature_{Max}$), (RH_{Max}), (RH_{Min}), (K_c) and (ET_o) are found suitable to better estimate (ET_c) for crops.

- To evaluate the accuracy of proposed model, the results from predicted (K_c) and (ET_c) are compared with the SVM model. The testing dataset is found slightly better than training dataset using proposed model as compared to SVM model.

- It has been observed that the proposed model shown the accuracy of 99% (K_c) and 98% (ET_c) for maize crop during training and testing periods.

In Chapter-7, a Decision Support System is developed for Irrigation Scheduling using Border and Sprinkler irrigation system. This research was conducted for wheat crop based on case study of Ludhiana station. The climate dataset was used to estimate the reference evapotranspiration ET_o and crop evapotranspiration ET_c. The following conclusions are made through the study:

- The fuzzy-forest model has been used to select best 10 features from the input data.

- PSO and DL models are applied to estimate the Net_{IR} $Gross_{IR}$, and PT_{IR} parameters. The optimal irrigation results were found as 500 mm and 315 mm for wheat crop of Border and Sprinkler irrigation system.

- According to the best sample results, the Net_{IR} is initialized as 200.62 mm/day, Gross_{IR} is initialized as 501.55 mm/day, and PT_{IR} as 83.2 hour for training dataset using Border irrigation. It is observed that the Sample_8 provided the best results of (Net_{IR} as 200.34 mm, and PT_{IR} as 83.22 h) and Sample_3 provided accurate results of Gross_{IR} as 500.2 mm.

- According to the best sample results, Net_{IR} is set as 205.59 mm/day, Gross_{IR} is initialized as 315.805 mm/day, and PT_{IR} as 287.175 hour for training dataset using Sprinkler irrigation. It is observed that the Sample_3 provided the best results of Net_{IR} as 284.9 mm, and Sample_3 provided the accurate results of (Gross_{IR} as 315.2 mm, and PT_{IR} as 287.05 h).

- The proposed ensemble-based approach is found to predict accurately the outcomes in agriculture domain.

8.2 Future Research Directions

The completion of a study endeavor is the beginning of various possibilities for future work. This dissertation introduced the notion that Machine Learning, Deep Learning, and Data Analytics can be applied in the area of irrigation scheduling. The primary objectives have been successfully completed and the following aspects have been identified for further study effort for smart water irrigation management with Artificial Intelligence technologies.

- To calculate reference evapotranspiration ET_o, based on classification using H2O model framework, and MPS quantum classifier and based on regression, a multilevel ensemble model using machine learning models have been developed for Hoshiarpur, Patiala, and Ludhiana (three stations of Punjab). There is scope to estimate the ET_o and compare with other stations of Punjab, where climate location is varying.

- To calculate the crop evapotranspiration ET_c, an innovative multilevel model ensembling technique has been introduced for the accurate estimation of single crop coefficient K_c and crop evapotranspiration ET_c using Fuzzy-Genetic (FG) and Regularization Random Forest (RRF) models. There is scope to investigate the performance of FG-RRF model using dual crop coefficient approach for improving the estimation of daily crop ET_c.

- The estimation of ET_o and ET_c based on DL and ML have been found efficient. The other efficient approaches to forecast the ET_o and ET_c can be investigated with Reinforcement Learning (RL), Generative Adversarial Network and Variational Autoencoders model (VAE's) models.

- The estimation of irrigation requirement, net irrigation, gross irrigation and pumping time based on Automl and DL-ensemble models have been developed for DSS-Irrigation Scheduling (DSS-IS). The forecasting website based on R-studio and R-shiny to represent the DSS-IS can be investigated. Also, DSS for irrigation scheduling based estimation can be conducted for Wheat and Maize crop using drip irrigation system.

Bibliography

[1] J. Singh, T. Dhaliwal, and D. Grover, "State agricultural profile-punjab," *AERC study*, vol. 30, pp. 12–27, 2012.

[2] R. Kumar and V. Mishra, "Increase in population exposure due to dry and wet extremes in india under a warming climate," *Earth's Future*, vol. 8, no. 12, p. e2020EF001731, 2020.

[3] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and electronics in agriculture*, vol. 147, pp. 70–90, 2018.

[4] M. H. Ali, *Fundamentals of irrigation and on-farm water management: Volume 1*. Springer New York, 2010.

[5] T. Howell, "Irrigation scheduling research and its impact on water use," in *Evapotranspiration and irrigation scheduling, Proceedings of the international conference*. American Society of Agricultural Engineer St. Joseph, MI, 1996, pp. 21–33.

[6] B. Das, "The use of irrigation systems for sustainable fish production in india," *Fisheries in irrigation systems of arid Asia. FAO fisheries Technical Paper*, no. 430, pp. 47–58, 2003.

[7] R. Thompson, M. Gallardo, L. Valdez, and M. Fernández, "Using plant water status to define threshold values for irrigation management of vegetable crops using soil moisture sensors," *Agricultural Water Management*, vol. 88, no. 1-3, pp. 147–158, 2007.

[8] R. G. Allen, L. S. Pereira, D. Raes, M. Smith *et al.*, "Crop evapotranspiration-guidelines for computing crop water requirements-fao irrigation and drainage paper 56," *FAO, Rome*, vol. 300, no. 9, p. D05109, 1998.

[9] M. Ali, "Weather: A driving force in determining irrigation demand," in *Fundamentals of Irrigation and On-farm Water Management: Volume 1*. Springer, 2010, pp. 31–105.

[10] H. F. Blaney and W. D. Criddle, *Determining consumptive use and irrigation water requirements*. US Department of Agriculture, 1962, no. 1275.

[11] R. G. Allen and W. O. Pruitt, "Rational use of the fao blaney-criddle formula," *Journal of Irrigation and Drainage Engineering*, vol. 112, no. 2, pp. 139–155, 1986.

[12] G. H. Hargreaves and Z. A. Samani, "Reference crop evapotranspiration from temperature," *Applied engineering in agriculture*, vol. 1, no. 2, pp. 96–99, 1985.

[13] I. Ladlani, L. Houichi, L. Djemili, S. Heddam, and K. Belouz, "Modeling daily reference evapotranspiration (et0) in the north of algeria using generalized regression

neural networks (grnn) and radial basis function neural networks (rbfnn): a comparative study," *Meteorology and Atmospheric Physics*, vol. 118, no. 3-4, pp. 163–178, 2012.

[14] C. Richardson, "Weather simulation for crop management models," *Transactions of the ASAE*, vol. 28, no. 5, pp. 1602–1606, 1985.

[15] K. Hinrichsen, "The ångström formula with coefficients having a physical meaning," *Solar Energy*, vol. 52, no. 6, pp. 491–495, 1994.

[16] M. Ali, A. Adham, and M. Talukder, "Estimation of solar radiation from climatic parameters," *Bangladesh J Agril Sci*, vol. 32, no. 1, pp. 99–104, 2005.

[17] R. Allen, M. Smith, A. Perrier, and L. S. Pereira, "An update for the definition of reference evapotranspiration," *ICID bulletin*, vol. 43, no. 2, pp. 1–34, 1994.

[18] R. K. Linsley Jr, M. A. Kohler, and J. L. Paulhus, *Hydrology for engineers: 3rd ed*, 1982.

[19] O. G. Sutton, *Micrometeorology: Volume 79*. McGraw-Hill New York, 1953.

[20] J. A. Duffie, W. A. Beckman, and N. Blair, *Solar engineering of thermal processes, photovoltaics and wind*. John Wiley & Sons, 2020.

[21] J. Almorox, V. H. Quej, and P. Martí, "Global performance ranking of temperature-based approaches for evapotranspiration estimation considering köppen climate classes," *Journal of Hydrology*, vol. 528, pp. 514–522, 2015.

[22] P. Droogers and R. G. Allen, "Estimating reference evapotranspiration under inaccurate data conditions," *Irrigation and drainage systems*, vol. 16, no. 1, pp. 33–45, 2002.

[23] C. W. Thornthwaite, "An approach toward a rational classification of climate," *Geographical review*, vol. 38, no. 1, pp. 55–94, 1948.

[24] M. A. Kohler, "Lake and pan evaporation," *Water-Lossa Investigations: Lake Hefner Studies, Technical Report. Geological Survey Professional Paper*, vol. 269, pp. 127–148, 1952.

[25] J. Doorenbos and W. Pruitt, "Crop water requirements," *FAO Irrigation and Drainage Paper 24*, p. 124, 1997.

[26] C. H. B. Priestley and R. Taylor, "On the assessment of surface heat flux and evaporation using large-scale parameters," *Monthly weather review*, vol. 100, no. 2, pp. 81–92, 1972.

[27] M. E. Jensen, R. D. Burman, and R. G. Allen, "Evapotranspiration and irrigation water requirements." Committee on Irrigation Water Requirements, ASCE, 1990.

[28] P. Gavilan, J. Berengena, and R. G. Allen, "Measuring versus estimating net radiation and soil heat flux: Impact on penman–monteith reference et estimates in semiarid regions," *Agricultural Water Management*, vol. 89, no. 3, pp. 275–286, 2007.

[29] H. Tabari, M. E. Grismer, and S. Trajkovic, "Comparative analysis of 31 reference evapotranspiration methods under humid conditions," *Irrigation Science*, vol. 31, no. 2, pp. 107–117, 2013.

[30] S. Maroufpoor, O. Bozorg-Haddad, and E. Maroufpoor, "Reference evapotranspiration estimating based on optimal input combination and hybrid artificial intelligent model: Hybridization of artificial neural network with grey wolf optimizer algorithm," *Journal of Hydrology*, p. 125060, 2020.

[31] B. Bouman, H. Van Keulen, H. Van Laar, and R. Rabbinge, "The 'school of de wit'crop growth simulation models: a pedigree and historical overview," *Agricultural systems*, vol. 52, no. 2-3, pp. 171–198, 1996.

[32] C. Brouwer and M. Heibloem, "Irrigation water management: irrigation water needs," *Training manual*, vol. 3, 1986.

[33] R. ALLEN, L. S. Pereira, D. Raes, and M. Smith, "Chapter 1. introduction to evapotranspiration," *Crop evapotranspiration–Guidelines for computing crop water requirements [online]. Food and Agriculture Organization of the United Nations (FAO). Irrigation and Drainage Paper*, vol. 56, 1998.

[34] R. G. Allen and L. S. Pereira, "Estimating crop coefficients from fraction of ground cover and height," *Irrigation Science*, vol. 28, no. 1, pp. 17–34, 2009.

[35] J. Rolim, A. Navarro, P. Vilar, C. Saraiva, and J. Catalao, "Crop data retrieval using earth observation data to support agricultural water management," *Engenharia Agrícola*, vol. 39, no. 3, pp. 380–390, 2019.

[36] M. E. Jensen, *Water consumption by agricultural plants (Chapter 1).* Academic Press, New York, 1968.

[37] P. W. Doorenbos J, "Guidelines for predicting crop water requirements, irrigation and drainage," *Irrigation and Drainage*, vol. 24, pp. 1–154, 1975.

[38] R. Burman, J. Wright, P. Nixon, and R. Hill, "Irrigation management – water requirements and water balance," 1980.

[39] P. Kingra, S. Hundal, and P. Sharma, "Characterization of crop coefficients for wheat and rice crops in punjab," *J. Agrometeorol*, vol. 6, pp. 58–60, 2004.

[40] L. B. Ferreira and F. F. da Cunha, "New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning," *Agricultural Water Management*, vol. 234, p. 106113, 2020.

[41] A. Phocaides, *Handbook on pressurized irrigation techniques.* Food & Agriculture Org., 2007.

[42] H. Ali, *Practices of Irrigation & On-farm Water Management: Volume 2.* Springer Science & Business Media, 2011, vol. 2.

[43] M. Rinaldi and Z. He, "Decision support systems to manage irrigation in agriculture," in *Advances in agronomy.* Elsevier, 2014, vol. 123, pp. 229–279.

[44] G. Guariso, S. Rinaldi, and R. Soncini-Sessa, "Decision support systems for water management: The lake como case study," *European Journal of Operational Research*, vol. 21, no. 3, pp. 295–306, 1985.

[45] T. A. Russo, N. Devineni, and U. Lall, "Assessment of agricultural water management in punjab, india, using bayesian methods," in *Sustainability of Integrated Water Resources Management.* Springer, 2015, pp. 147–162.

[46] K. G. Liakos, P. Busato, D. Moshou, S. Pearson, and D. Bochtis, "Machine learning in agriculture: A review," *Sensors*, vol. 18, no. 8, p. 2674, 2018.

[47] S. Yan, L. Wu, J. Fan, F. Zhang, Y. Zou, and Y. Wu, "A novel hybrid woa-xgb model for estimating daily reference evapotranspiration using local and external meteorological data: Applications in arid and humid regions of china," *Agricultural Water Management*, vol. 244, p. 106594.

[48] J. Shiri, A. H. Nazemi, A. A. Sadraddini, G. Landeras, O. Kisi, A. F. Fard, and P. Marti, "Comparison of heuristic and empirical approaches for estimating reference evapotranspiration from limited inputs in iran," *Computers and Electronics in Agriculture*, vol. 108, pp. 230–241, 2014.

[49] O. Kişi, A. P. Ali Baba, and J. Shiri, "Generalized neurofuzzy models for estimating daily pan evaporation values from weather data," *Journal of irrigation and drainage engineering*, vol. 138, no. 4, pp. 349–362, 2012.

[50] M. Kumar, N. Raghuwanshi, and R. Singh, "Artificial neural networks approach in evapotranspiration modeling: a review," *Irrigation science*, vol. 29, no. 1, pp. 11–25, 2011.

[51] H. Tabari, O. Kisi, A. Ezani, and P. H. Talaee, "Svm, anfis, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment," *Journal of Hydrology*, vol. 444, pp. 78–89, 2012.

[52] A. H. Zaji and H. Bonakdari, "Performance evaluation of two different neural network and particle swarm optimization methods for prediction of discharge capacity of modified triangular side weirs," *Flow Measurement and Instrumentation*, vol. 40, pp. 149–156, 2014.

[53] S. S. Abdullah, M. A. Malek, N. S. Abdullah, O. Kisi, and K. S. Yap, "Extreme learning machines: a new approach for prediction of reference evapotranspiration," *Journal of Hydrology*, vol. 527, pp. 184–195, 2015.

[54] O. Kisi, "Modeling reference evapotranspiration using three different heuristic regression approaches," *Agricultural Water Management*, vol. 169, pp. 162–172, 2016.

[55] Y. Feng, N. Cui, L. Zhao, X. Hu, and D. Gong, "Comparison of elm, gann, wnn and empirical models for estimating reference evapotranspiration in humid region of southwest china," *Journal of Hydrology*, vol. 536, pp. 376–383, 2016.

[56] A. G. Ivakhnenko and V. G. Lapa, "Cybernetic predicting devices," Purdue Univ Lafayette Ind School of Electrical Engineering, Tech. Rep., 1966.

[57] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, "Deep learning for smart manufacturing: Methods and applications," *Journal of Manufacturing Systems*, vol. 48, pp. 144–156, 2018.

[58] B. Brahma and R. Wadhvani, "Solar irradiance forecasting based on deep learning methodologies and multi-site data," *Symmetry*, vol. 12, no. 11, p. 1830, 2020.

[59] M. Ahsan, M. Kumari, and T. Sharma, "Detection of context-varying rumors on twitter through deep learning," *Int. J. Adv. Sci. Technol*, vol. 128, pp. 45–58, 2019.

[60] G. Yang, Y. Huang, and C. Zhao, "Agri-bigdata: A smart pathway for crop nitrogen inputs," *Artificial Intelligence in Agriculture*, vol. 4, pp. 150–152, 2020.

[61] Y. Yang, Y. Cui, K. Bai, T. Luo, J. Dai, W. Wang, and Y. Luo, "Short-term forecasting of daily reference evapotranspiration using the reduced-set penman-monteith model and public weather forecasts," *Agricultural Water Management*, vol. 211, pp. 70–80, 2019.

[62] M. Gocić, S. Motamedi, S. Shamshirband, D. Petković, S. Ch, R. Hashim, and M. Arif, "Soft computing approaches for forecasting reference evapotranspiration," *Computers and Electronics in Agriculture*, vol. 113, pp. 164–173, 2015.

[63] X. Dou and Y. Yang, "Evapotranspiration estimation using four different machine learning approaches in different terrestrial ecosystems," *Computers and Electronics in Agriculture*, vol. 148, pp. 95–106, 2018.

[64] M. A. Yassin, A. Alazba, and M. A. Mattar, "Artificial neural networks versus gene expression programming for estimating reference evapotranspiration in arid climate," *Agricultural Water Management*, vol. 163, pp. 110–124, 2016.

[65] P. Wable, M. Jha, and S. Gorantiwar, "Assessing suitability of temperature-based reference evapotranspiration methods for semi-arid basin of maharashtra," *Journal of Agrometeorology*, vol. 21, no. 3, pp. 351–356, 2019.

[66] M. T. Hobbins, A. Wood, D. J. McEvoy, J. L. Huntington, C. Morton, M. Anderson, and C. Hain, "The evaporative demand drought index. part i: Linking drought evolution to variations in evaporative demand," *Journal of Hydrometeorology*, vol. 17, no. 6, pp. 1745–1761, 2016.

[67] S. M. Vicente-Serrano, M. Tomas-Burguera, S. Beguería, F. Reig, B. Latorre, M. Peña-Gallardo, M. Y. Luna, A. Morata, and J. C. González-Hidalgo, "A high resolution dataset of drought indices for spain," *Data*, vol. 2, no. 3, p. 22, 2017.

[68] P. Martí, P. González-Altozano, R. López-Urrea, L. A. Mancha, and J. Shiri, "Modeling reference evapotranspiration with calculated targets. assessment and implications," *Agricultural Water Management*, vol. 149, pp. 81–90, 2015.

[69] K. C. Perera, A. W. Western, D. E. Robertson, B. George, and B. Nawarathna, "Ensemble forecasting of short-term system scale irrigation demands using real-time flow data and numerical weather predictions," *Water Resources Research*, vol. 52, no. 6, pp. 4801–4822, 2016.

[70] M. Valipour, M. A. G. Sefidkouhi, M. Raeini *et al.*, "Selecting the best model to estimate potential evapotranspiration with respect to climate change and magnitudes of extreme events," *Agricultural Water Management*, vol. 180, pp. 50–60, 2017.

[71] N. Malamos, P. Barouchas, I. Tsirogiannis, A. Liopa-Tsakalidi, and T. Koromilas, "Estimation of monthly fao penman-monteith evapotranspiration in gis environment, through a geometry independent algorithm," *Agriculture and Agricultural Science Procedia*, vol. 4, pp. 290–299, 2015.

[72] A. Tegos, N. Malamos, A. Efstratiadis, I. Tsoukalas, A. Karanasios, and D. Koutsoyiannis, "Parametric modelling of potential evapotranspiration: A global survey," *Water*, vol. 9, no. 10, p. 795, 2017.

[73] D. L. Ficklin, S. L. Letsinger, H. Gholizadeh, and J. T. Maxwell, "Incorporation of the penman–monteith potential evapotranspiration method into a palmer drought severity index tool," *Computers & Geosciences*, vol. 85, pp. 136–141, 2015.

[74] M. M. Heydari, A. Tajamoli, S. H. Ghoreishi, M. K. Darbe-Esfahani, and H. Gilasi, "Evaluation and calibration of blaney–criddle equation for estimating reference evapotranspiration in semiarid and arid regions," *Environmental Earth Sciences*, vol. 74, no. 5, pp. 4053–4063, 2015.

[75] A. P. Patil and P. C. Deka, "An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs," *Computers and Electronics in Agriculture*, vol. 121, pp. 385–392, 2016.

[76] L. Wu, Y. Peng, J. Fan, Y. Wang, and G. Huang, "A novel kernel extreme learning machine model coupled with k-means clustering and firefly algorithm for estimating monthly reference evapotranspiration in parallel computation," *Agricultural Water Management*, vol. 245, p. 106624, 2021.

[77] J. A. Bellido-Jiménez, J. Estévez, and A. P. García-Marín, "New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of spain," *Agricultural Water Management*, vol. 245, p. 106558, 2021.

[78] T. Wu, W. Zhang, X. Jiao, W. Guo, and Y. A. Hamoud, "Evaluation of stacking and blending ensemble learning methods for estimating daily reference evapotranspiration," *Computers and Electronics in Agriculture*, vol. 184, p. 106039, 2021.

[79] S. Kar, V. K. Purbey, S. Suradhaniwar, L. B. Korbu, J. Kholová, S. S. Durbha, J. Adinarayana, and V. Vadez, "An ensemble machine learning approach for determination of the optimum sampling time for evapotranspiration assessment from high-throughput phenotyping data," *Computers and Electronics in Agriculture*, vol. 182, p. 105992, 2021.

[80] Y. Bai, S. Zhang, N. Bhattarai, K. Mallick, Q. Liu, L. Tang, J. Im, L. Guo, and J. Zhang, "On the use of machine learning based ensemble approaches to improve evapotranspiration estimates from croplands across a wide environmental gradient," *Agricultural and Forest Meteorology*, vol. 298, p. 108308, 2021.

[81] S. Vulova, F. Meier, A. D. Rocha, J. Quanz, H. Nouri, and B. Kleinschmit, "Modeling urban evapotranspiration using remote sensing, flux footprints, and artificial intelligence," *Science of The Total Environment*, p. 147293, 2021.

[82] R. M. Adnan, A. Malik, A. Kumar, K. S. Parmar, and O. Kisi, "Pan evaporation modeling by three different neuro-fuzzy intelligent systems using climatic inputs," *Arabian Journal of Geosciences*, vol. 12, no. 20, p. 606, 2019.

[83] R. M. Adnan, Z. Chen, X. Yuan, O. Kisi, A. El-Shafie, A. Kuriqi, and M. Ikram, "Reference evapotranspiration modeling using new heuristic methods," *Entropy*, vol. 22, no. 5, p. 547, 2020.

[84] R. M. Adnan, Z. Liang, S. Heddam, M. Zounemat-Kermani, O. Kisi, and B. Li, "Least square support vector machine and multivariate adaptive regression splines for streamflow prediction in mountainous basin using hydro-meteorological data as inputs," *Journal of Hydrology*, vol. 586, p. 124371, 2020.

[85] S. Heddam, O. Kisi, A. Sebbar, L. Houichi, and L. Djemili, "New formulation for predicting daily reference evapotranspiration (et 0) in the mediterranean region of algeria country: Optimally pruned extreme learning machine (opelm) versus online sequential extreme learning machine (oselm)," *Water Resources in Algeria-Part I: Assessment of Surface and Groundwater Resources*, pp. 181–199, 2020.

[86] Y. Tikhamarine, A. Malik, D. Souag-Gamane, and O. Kisi, "Artificial intelligence models versus empirical equations for modeling monthly reference evapotranspiration," *Environmental Science and Pollution Research*, vol. 27, pp. 30 001–30 019, 2020.

[87] B. Mohammadi and S. Mehdizadeh, "Modeling daily reference evapotranspiration via a novel approach based on support vector regression coupled with whale optimization algorithm," *Agricultural Water Management*, p. 106145, 2020.

[88] M. A. Mattar, "Using gene expression programming in monthly reference evapotranspiration modeling: A case study in egypt," *Agricultural Water Management*, vol. 198, pp. 28–38, 2018.

[89] H. Tao, L. Diop, A. Bodian, K. Djaman, P. M. Ndiaye, and Z. M. Yaseen, "Reference evapotranspiration prediction using hybridized fuzzy model with firefly algorithm: Regional case study in burkina faso," *Agricultural water management*, vol. 208, pp. 140–151, 2018.

[90] S. Q. Salih, M. F. Allawi, A. A. Yousif, A. M. Armanuos, M. K. Saggi, M. Ali, S. Shahid, N. Al-Ansari, Z. M. Yaseen, and K.-W. Chau, "Viability of the advanced adaptive neuro-fuzzy inference system model on reservoir evaporation process simulation: case study of nasser lake in egypt," *Engineering Applications of Computational Fluid Mechanics*, vol. 13, no. 1, pp. 878–891, 2019.

[91] Z. M. Yaseen, A. M. Al-Juboori, U. Beyaztas, N. Al-Ansari, K.-W. Chau, C. Qi, M. Ali, S. Q. Salih, and S. Shahid, "Prediction of evaporation in arid and semi-arid regions: a comparative study using different machine learning models," *Engineering applications of computational fluid mechanics*, vol. 14, no. 1, pp. 70–89, 2020.

[92] Y. Falamarzi, N. Palizdan, Y. F. Huang, and T. S. Lee, "Estimating evapotranspiration from temperature and wind speed data using artificial and wavelet neural networks (wnns)," *Agricultural Water Management*, vol. 140, pp. 26–36, 2014.

[93] O. Kisi, "Pan evaporation modeling using least square support vector machine, multivariate adaptive regression splines and m5 model tree," *Journal of Hydrology*, vol. 528, pp. 312–320, 2015.

[94] R. Ballesteros, J. F. Ortega, and M. Á. Moreno, "Foreto: new software for reference evapotranspiration forecasting," *Journal of Arid Environments*, vol. 124, pp. 128–141, 2016.

[95] M. K. Goyal, B. Bharti, J. Quilty, J. Adamowski, and A. Pandey, "Modeling of daily pan evaporation in sub tropical climates using ann, ls-svr, fuzzy logic, and anfis," *Expert systems with applications*, vol. 41, no. 11, pp. 5267–5276, 2014.

[96] Y. Chen, W. Yuan, J. Xia, J. B. Fisher, W. Dong, X. Zhang, S. Liang, A. Ye, W. Cai, and J. Feng, "Using bayesian model averaging to estimate terrestrial evapotranspiration in china," *Journal of Hydrology*, vol. 528, pp. 537–549, 2015.

[97] S. Mehdizadeh, J. Behmanesh, and K. Khalili, "Using mars, svm, gep and empirical equations for estimation of monthly mean reference evapotranspiration," *Computers and Electronics in Agriculture*, vol. 139, pp. 103–114, 2017.

[98] A. Malik, A. Kumar, and O. Kisi, "Monthly pan-evaporation estimation in indian central himalayas using different heuristic approaches and climate based models," *Computers and Electronics in Agriculture*, vol. 143, pp. 302–313, 2017.

[99] A. P. Patil and P. C. Deka, "An extreme learning machine approach for modeling evapotranspiration using extrinsic inputs," *Computers and Electronics in Agriculture*, vol. 121, pp. 385–392, 2016.

[100] N. S. Punn, S. K. Sonbhadra, and S. Agarwal, "Covid-19 epidemic analysis using machine learning and deep learning algorithms," *MedRxiv*, 2020.

[101] R. K. Gupta, A. R. Gupta, N. Pathik, R. Pateriya, P. K. Chaurasiya, U. Rakjak, T. N. Verma, A. M. Alosaimi, and M. A. Hussein, "Novel deep neural network technique for detecting environmental effect of covid-19," *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, pp. 1–19, 2021.

[102] S. Agrawal, A. Tiwari, and I. Goel, "Genetically optimized deep neural learning for breast cancer prediction," in *Soft Computing for Problem Solving 2019*. Springer, 2020, pp. 127–139.

[103] S. Tripathy, V. K. Rai, and J. Mathew, "Marpuf: physical unclonable function with improved machine learning attack resistance," *IET Circuits, Devices & Systems*, 2021.

[104] X. Li, D. Chang, Z. Ma, Z.-H. Tan, J.-H. Xue, J. Cao, and J. Guo, "Deep interboost networks for small-sample image classification," *Neurocomputing*, 2020.

[105] N. E. M. Khalifa, M. H. N. Taha, D. E. Ali, A. Slowik, and A. E. Hassanien, "Artificial intelligence technique for gene expression by tumor rna-seq data: a novel optimized deep learning approach," *IEEE Access*, vol. 8, pp. 22 874–22 883, 2020.

[106] G. Ghazaei, A. Alameer, P. Degenaar, G. Morgan, and K. Nazarpour, "Deep learning-based artificial vision for grasp classification in myoelectric hands," *Journal of neural engineering*, vol. 14, no. 3, p. 036025, 2017.

[107] S. Bothe, T. Gärtner, and S. Wrobel, "On-line handwriting recognition with parallelized machine learning algorithms," in *Annual Conference on Artificial Intelligence*. Springer, 2010, pp. 82–90.

[108] S. K. Pandey, R. B. Mishra, and A. K. Tripathi, "Machine learning based methods for software fault prediction: A survey," *Expert Systems with Applications*, p. 114595, 2021.

[109] K. P. Singh, M. Kansal, and K. Deep, "Ga-nr for optimal design of water distribution networks," *International Journal of Operational Research*, vol. 20, no. 3, pp. 241–261, 2014.

[110] X. Han, Z. Wei, B. Zhang, Y. Li, T. Du, and H. Chen, "Crop evapotranspiration prediction by considering dynamic change of crop coefficient and the precipitation effect in back-propagation neural network model," *Journal of Hydrology*, vol. 596, p. 126104, 2021.

[111] R. Mehta and V. Pandey, "Reference evapotranspiration (eto) and crop water requirement (etc) of wheat and maize in gujarat," *Journal of agrometeorology*, vol. 17, no. 1, p. 107, 2015.

[112] A. Elbeltagi, J. Deng, K. Wang, A. Malik, and S. Maroufpoor, "Modeling long-term dynamics of crop evapotranspiration using deep learning in a semi-arid environment," *Agricultural Water Management*, vol. 241, p. 106334, 2020.

[113] T. A. Russo, N. Devineni, and U. Lall, "Assessment of agricultural water management in punjab, india, using bayesian methods," in *Sustainability of Integrated Water Resources Management*. Springer, 2015, pp. 147–162.

[114] D. L. Ehret, B. D. Hill, T. Helmer, and D. R. Edwards, "Neural network modeling of greenhouse tomato yield, growth and water use from automated crop monitoring data," *Computers and Electronics in Agriculture*, vol. 79, no. 1, pp. 82–89, 2011.

[115] S. Maurya and V. K. Jain, "Fuzzy based energy efficient sensor network protocol for precision agriculture," *Computers and Electronics in Agriculture*, vol. 130, pp. 20–37, 2016.

[116] G. Yang, L. Liu, P. Guo, and M. Li, "A flexible decision support system for irrigation scheduling in an irrigation district in china," *Agricultural water management*, vol. 179, pp. 378–389, 2017.

[117] Y. S. Chauhan, G. C. Wright, D. Holzworth, R. C. Rachaputi, and J. O. Payero, "Aquaman: a web-based decision support system for irrigation scheduling in peanuts," *Irrigation Science*, vol. 31, no. 3, pp. 271–283, 2013.

[118] P. Gavilán, N. Ruiz, and D. Lozano, "Daily forecasting of reference and strawberry crop evapotranspiration in greenhouses in a mediterranean climate based on solar radiation estimates," *Agricultural Water Management*, vol. 159, pp. 307–317, 2015.

[119] H. Tabari, C. Martinez, A. Ezani, and P. H. Talaee, "Applicability of support vector machines and adaptive neurofuzzy inference system for modeling potato crop evapotranspiration," *Irrigation Science*, vol. 31, no. 4, pp. 575–588, 2013.

[120] S. S. Yamaç and M. Todorovic, "Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data," *Agricultural Water Management*, vol. 228, p. 105875, 2020.

[121] H. Z. Abyaneh, A. M. Nia, M. B. Varkeshi, S. Marofi, and O. Kisi, "Performance evaluation of ann and anfis models for estimating garlic crop evapotranspiration," *Journal of irrigation and drainage engineering*, vol. 137, no. 5, pp. 280–286, 2011.

[122] G. Rana and N. Katerji, "Measurement and estimation of actual evapotranspiration in the field under mediterranean climate: a review," *European Journal of agronomy*, vol. 13, no. 2-3, pp. 125–153, 2000.

[123] C. Liu, X. Zhang, and Y. Zhang, "Determination of daily evaporation and evapotranspiration of winter wheat and maize by large-scale weighing lysimeter and micro-lysimeter," *Agricultural and Forest Meteorology*, vol. 111, no. 2, pp. 109–120, 2002.

[124] R. López-Urrea, A. Montoro, J. González-Piqueras, P. López-Fuster, and E. Fereres, "Water use of spring wheat to raise water productivity," *Agricultural Water Management*, vol. 96, no. 9, pp. 1305–1310, 2009.

[125] S. Shah and R. Edling, "Daily evapotranspiration prediction from louisiana flooded rice field," *Journal of irrigation and drainage engineering*, vol. 126, no. 1, pp. 8–13, 2000.

[126] K. C. Reddy, "Development of crop coefficient models of castor and maize crops," *European Journal of Agronomy*, vol. 69, pp. 59–62, 2015.

[127] J. Ko, G. Piccinni, T. Marek, and T. Howell, "Determination of growth-stage-specific crop coefficients (kc) of cotton and wheat," *Agricultural Water Management*, vol. 96, no. 12, pp. 1691–1697, 2009.

[128] F. Tong and P. Guo, "Simulation and optimization for crop water allocation based on crop water production functions and climate factor under uncertainty," *Applied Mathematical Modelling*, vol. 37, no. 14-15, pp. 7708–7716, 2013.

[129] X. Mo, R. Guo, S. Liu, Z. Lin, and S. Hu, "Impacts of climate change on crop evapotranspiration with ensemble gcm projections in the north china plain," *Climatic change*, vol. 120, no. 1-2, pp. 299–312, 2013.

[130] J.-P. Lhomme, R. Mougou, and M. Mansour, "Potential impact of climate change on durum wheat cropping in tunisia," *Climatic Change*, vol. 96, no. 4, pp. 549–564, 2009.

[131] Z. Liu, X. Yang, X. Lin, P. Gowda, S. Lv, and J. Wang, "Climate zones determine where substantial increases of maize yields can be attained in northeast china," *Climatic Change*, vol. 149, no. 3-4, pp. 473–487, 2018.

[132] M. Salama, K. M. Yousef, and A. Mostafa, "Simple equation for estimating actual evapotranspiration using heat units for wheat in arid regions," *Journal of Radiation Research and Applied Sciences*, vol. 8, no. 3, pp. 418–427, 2015.

[133] S. S. Anapalli, L. R. Ahuja, P. H. Gowda, L. Ma, G. Marek, S. R. Evett, and T. A. Howell, "Simulation of crop evapotranspiration and crop coefficients with data in weighing lysimeters," *Agricultural Water Management*, vol. 177, pp. 274–283, 2016.

[134] M. O'Grady, D. Langton, F. Salinari, P. Daly, and G. O'Hare, "Service design for climate-smart agriculture," *Information Processing in Agriculture*, 2020.

[135] C. A. Midingoyi, C. Pradal, A. Enders, D. Fumagalli, H. Raynal, M. Donatelli, I. N. Athanasiadis, C. Porter, G. Hoogenboom, D. Holzworth *et al.*, "Crop2ml: An open-source multi-language modeling framework for the exchange and reuse of crop model components," *Environmental Modelling & Software*, p. 105055, 2021.

[136] C. Gutierrez-Ninahuaman and R. Gonzalez-Herrera, "Software to analyze eto. compilation of indirect methods," *Environmental Modelling & Software*, p. 105056, 2021.

[137] H. Navarro-Hellín, J. Martínez-del Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sánchez, "A decision support system for managing irrigation in agriculture," *Computers and Electronics in Agriculture*, vol. 124, pp. 121–131, 2016.

[138] Z. Li and Z. Sun, "Optimized single irrigation can achieve high corn yield and water use efficiency in the corn belt of northeast china," *European Journal of Agronomy*, vol. 75, pp. 12–24, 2016.

[139] S. Brar, S. Mahal, A. Brar, K. Vashist, N. Sharma, and G. Buttar, "Transplanting time and seedling age affect water productivity, rice yield and quality in north-west india," *Agricultural Water Management*, vol. 115, pp. 217–222, 2012.

[140] S. Prihar, K. Khera, K. Sandhu, and B. Sandhu, "Comparison of irrigation schedules based on pan evaporation and growth stages in winter wheat 1," *Agronomy Journal*, vol. 68, no. 4, pp. 650–653, 1976.

[141] J. Timsina, D. Godwin, E. Humphreys, S. Kukal, D. Smith *et al.*, "Evaluation of options for increasing yield and water productivity of wheat in punjab, india using the dssat-csm-ceres-wheat model," *Agricultural Water Management*, vol. 95, no. 9, pp. 1099–1110, 2008.

[142] A. Paraskevopoulos and A. Singels, "Integrating soil water monitoring technology and weather based crop modelling to provide improved decision support for sugarcane irrigation management," *Computers and electronics in agriculture*, vol. 105, pp. 44–53, 2014.

[143] Y. Ma, S. Feng, and X. Song, "Evaluation of optimal irrigation scheduling and groundwater recharge at representative sites in the north china plain with swap model and field experiments," *Computers and Electronics in Agriculture*, vol. 116, pp. 125–136, 2015.

[144] M. Afzal, A. Battilani, D. Solimando, and R. Ragab, "Improving water resources management using different irrigation strategies and water qualities: field and modelling study," *Agricultural Water Management*, vol. 176, pp. 40–54, 2016.

[145] G. Yang, L. Liu, P. Guo, and M. Li, "A flexible decision support system for irrigation scheduling in an irrigation district in china," *Agricultural water management*, vol. 179, pp. 378–389, 2017.

[146] E. Giusti and S. Marsili-Libelli, "A fuzzy decision support system for irrigation and water conservation in agriculture," *Environmental Modelling & Software*, vol. 63, pp. 73–86, 2015.

[147] A. K. L. Sahoo, Bhabagrahi and R. K. Sahu., "Fuzzy multiobjective and linear programming based management models for optimal land-water-crop system planning," *Water resources management*, vol. 20, pp. 931–948, 2006.

[148] M. J. Reddy and D. N. Kumar, "Evolving strategies for crop planning and operation of irrigation reservoir system using multi-objective differential evolution," *Irrigation Science*, vol. 26, no. 2, pp. 177–190, 2008.

[149] J. Adeyemo and F. Otieno, "Differential evolution algorithm for solving multi-objective crop planning model," *Agricultural water management*, vol. 97, no. 6, pp. 848–856, 2010.

[150] W. Ren, Q. Xiang, Y. Yang, H. Cui, and L. Dai, "Implement of fuzzy control for greenhouse irrigation," in *International Conference on Computer and Computing Technologies in Agriculture*. Springer, 2010, pp. 267–274.

[151] S. P. Friedman, G. Communar, and A. Gamliel, "Didas–user-friendly software package for assisting drip irrigation design and scheduling," *Computers and Electronics in Agriculture*, vol. 120, pp. 36–52, 2016.

[152] D. Isern, S. Abelló, and A. Moreno, "Development of a multi-agent system simulation platform for irrigation scheduling with case studies for garden irrigation," *Computers and electronics in agriculture*, vol. 87, pp. 1–13, 2012.

[153] M. García-Vila and E. Fereres, "Combining the simulation crop model aquacrop with an economic model for the optimization of irrigation management at farm level," *European Journal of Agronomy*, vol. 36, no. 1, pp. 21–31, 2012.

[154] T. Jackson, M. A. Hanjra, S. Khan, and M. Hafeez, "Building a climate resilient farm: A risk based approach for understanding water, energy and emissions in irrigated agriculture," *Agricultural Systems*, vol. 104, no. 9, pp. 729–745, 2011.

[155] M. R. Morris, A. Hussain, M. H. Gillies, and N. J. O'Halloran, "Inflow rate and border irrigation performance," *Agricultural Water Management*, vol. 155, pp. 76–86, 2015.

[156] J. Burguete, A. Lacasta, and P. García-Navarro, "Surcos: A software tool to simulate irrigation and fertigation in isolated furrows and furrow networks," *Computers and Electronics in Agriculture*, vol. 103, pp. 91–103, 2014.

[157] L. S. Pereira, J. Gonçalves, B. Dong, Z. Mao, and S. Fang, "Assessing basin irrigation and scheduling strategies for saving irrigation water and controlling salinity in the upper yellow river basin, china," *Agricultural Water Management*, vol. 93, no. 3, pp. 109–122, 2007.

[158] G. H. Schmitz, N. Schütze, and U. Petersohn, "New strategy for optimizing water application under trickle irrigation," *Journal of irrigation and drainage engineering*, vol. 128, no. 5, pp. 287–297, 2002.

[159] J. A. d. J. V. J. M. T. M.-B. Alvarez, Jose Fernando Ortega and E. L. Mata, "Mopeco: an economic optimization model for irrigation water management," *Irrigation Science*, vol. 23, no. 2, pp. 61–75, 2004.

[160] N. S. de Paly, Michael and A. Zell, "Determining crop-production functions using multi-objective evolutionary algorithms," in *IEEE Congress on Evolutionary Computation*. IEEE, 2010, pp. 1–8.

[161] J. L. Y. S. X. Z. Li, Hongjun and Y. Lei, "Web-based irrigation decision support system with limited inputs for farmers," *Agricultural Water Management*, vol. 210, pp. 279–285, 2018.

[162] M. K. Rowshon, N. S. Dlamini, M. A. Mojid, M. Adib, M. S. M. Amin, and S. H. Lai, "Modeling climate-smart decision support system (csdss) for analyzing water demand of a large-scale rice irrigation scheme," *Agricultural water management*, vol. 216, pp. 138–152, 2019.

[163] E. Antonopoulou, S. Karetsos, M. Maliappis, and A. Sideridis, "Web and mobile technologies in a prototype dss for major field crops," *Computers and Electronics in Agriculture*, vol. 70, no. 2, pp. 292–301, 2010.

[164] C. Li, R. Dutta, C. Kloppers, C. D'Este, A. Morshed, A. Almeida, A. Das, and J. Aryal, "Mobile application based sustainable irrigation water usage decision support system: An intelligent sensor cloud approach," in *SENSORS, 2013 IEEE*. IEEE, 2013, pp. 1–4.

[165] A. Bonfante, E. Monaco, P. Manna, R. De Mascellis, A. Basile, M. Buonanno, G. Cantilena, A. Esposito, A. Tedeschi, C. De Michele *et al.*, "Lcis dss—an irrigation supporting system for water use efficiency improvement in precision agriculture: A maize case study," *Agricultural Systems*, vol. 176, p. 102646, 2019.

[166] R. Ragab, "Integrated management tool for water, crop, soil and n-fertilizers: the saltmed model," *Irrigation and drainage*, vol. 64, no. 1, pp. 1–12, 2015.

[167] A. Bandyopadhyay, A. Bhadra, R. Swarnakar, N. Raghuwanshi, and R. Singh, "Estimation of reference evapotranspiration using a user-friendly decision support system: Dss_et," *Agricultural and forest meteorology*, vol. 154, pp. 19–29, 2012.

[168] M. Olberz, K. Kahlen, and J. Zinkernagel, "Assessing the impact of reference evapotranspiration models on decision support systems for irrigation," *Horticulturae*, vol. 4, no. 4, p. 49, 2018.

[169] R. Ballesteros, J. F. Ortega, and M. Á. Moreno, "Foreto: new software for reference evapotranspiration forecasting," *Journal of Arid Environments*, vol. 124, pp. 128–141, 2016.

[170] M. Gocic and S. Trajkovic, "Software for estimating reference evapotranspiration using limited weather data," *Computers and Electronics in Agriculture*, vol. 71, no. 2, pp. 158–162, 2010.

[171] C. O. Stöckle, M. Donatelli, and R. Nelson, "Cropsyst, a cropping systems simulation model," *European journal of agronomy*, vol. 18, no. 3-4, pp. 289–307, 2003.

[172] N. Brisson, B. Mary, D. Ripoche, M. H. Jeuffroy, F. Ruget, B. Nicoullaud, P. Gate, F. Devienne-Barret, R. Antonioletti, C. Durr *et al.*, "Stics: a generic model for the simulation of crops and their water and nitrogen balances. i. theory and parameterization applied to wheat and corn," 1998.

[173] J. Williams, C. Jones, and P. T. Dyke, "A modeling approach to determining the relationship between erosion and soil productivity," *Transactions of the ASAE*, vol. 27, no. 1, pp. 129–0144, 1984.

[174] J. Jones, G. Tsuji, G. Hoogenboom, L. Hunt, P. Thornton, P. Wilkens, D. Imamura, W. Bowen, and U. Singh, "Decision support system for agrotechnology transfer: Dssat v3," in *Understanding options for agricultural production*. Springer, 1998, pp. 157–177.

[175] C. Giménez, M. Gallardo, C. Martínez-Gaitán, C. Stöckle, R. Thompson, and M. Granados, "Vegsyst, a simulation model of daily crop growth, nitrogen uptake and evapotranspiration for pepper crops for use in an on-farm decision support system," *Irrigation Science*, vol. 31, no. 3, pp. 465–477, 2013.

[176] J. Ritchie, U. Singh, D. Godwin, and W. Bowen, "Cereal growth, development and yield," in *Understanding options for agricultural production*. Springer, 1998, pp. 79–98.

[177] A. Boretti and L. Rosa, "Reassessing the projections of the world water development report," *NPJ Clean Water*, vol. 2, no. 1, pp. 1–6, 2019.

[178] ICAR, "State-specific Technological Interventions for Higher Agricultural Growth," Indian Council of Agricultural Research, New Delhi, India, Tech. Rep., 2007.

[179] Anonymous, *Package of practices for crops of Punjab - Rabi*. Additional Director of Communication for Punjab Agricultural University, Ludhiana, 2012.

[180] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[181] M. Belgiu and L. Drăguţ, "Random forest in remote sensing: A review of applications and future directions," *ISPRS journal of photogrammetry and remote sensing*, vol. 114, pp. 24–31, 2016.

[182] J. H. Friedman, "Greedy function approximation: a gradient boosting machine," *Annals of statistics*, pp. 1189–1232, 2001.

[183] L. Breiman, "Using adaptive bagging to debias regressions," Technical Report 547, Statistics Dept. UCB, Tech. Rep., 1999.

[184] M. Malohlava and A. Candel, "Gradient boosting machine with h2o," *H20. ai*, 2017.

[185] R. E. Chandler and H. S. Wheater, "Analysis of rainfall variability using generalized linear models: a case study from the west of ireland," *Water Resources Research*, vol. 38, no. 10, 2002.

[186] J. A. Nelder and R. J. Baker, *Generalized linear models*. Wiley Online Library, 1972.

[187] C. Yang, R. Chandler, V. Isham, and H. Wheater, "Spatial-temporal rainfall simulation using generalized linear models," *Water Resources Research*, vol. 41, no. 11, 2005.

[188] P. McCullagh, "Generalized linear models," *European Journal of Operational Research*, vol. 16, no. 3, pp. 285–292, 1984.

[189] T. Nykodym, T. Kraljevic, N. Hussami, A. Rao, and A. Wang, "Generalized linear modeling with h2o," *Published by H2O. ai, Inc*, 2016.

[190] A. G. Ivakhnenko and V. G. Lapa, "Cybernetic predicting devices," Purdue Univ Lafayette Ind School of Electrical Engineering, Tech. Rep., 1966.

[191] A. Candel, V. Parmar, E. LeDell, and A. Arora, "Deep learning with h2o," *H2O. ai Inc*, 2016.

[192] Z. Sahri, R. Yusof, and J. Watada, "Finnim: Iterative imputation of missing values in dissolved gas analysis dataset," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 4, pp. 2093–2102, 2014.

[193] D. J. Stekhoven and P. Bühlmann, "Missforest—non-parametric missing value imputation for mixed-type data," *Bioinformatics*, vol. 28, no. 1, pp. 112–118, 2011.

[194] R. Core, "Team," *R: A language and environment for statistical computing*, 2015.

[195] A. S. Bhatia and A. Kumar, "Quantifying matrix product state," *Quantum Information Processing*, vol. 17, no. 3, p. 41, 2018.

[196] E. Miles Stoudenmire and D. J. Schwab, "Supervised learning with quantum-inspired tensor networks," *arXiv preprint arXiv:1605.05775*, 2016.

[197] W. Huggins, P. Patel, K. B. Whaley, and E. M. Stoudenmire, "Towards quantum machine learning with tensor networks," *arXiv preprint arXiv:1803.11537*, 2018.

[198] E. Grant, M. Benedetti, S. Cao, A. Hallam, J. Lockhart, V. Stojevic, A. G. Green, and S. Severini, "Hierarchical quantum classifiers," *arXiv preprint arXiv:1804.03680*, 2018.

[199] C. Huntingford, E. S. Jeffers, M. B. Bonsall, H. M. Christensen, T. Lees, and H. Yang, "Machine learning and artificial intelligence to aid climate change research and preparedness," *Environmental Research Letters*, vol. 14, no. 12, p. 124007, 2019.

[200] J. Gautam, A. Chakrabarti, S. Agarwal, A. Singh, S. Gupta, and J. Singh, "Monitoring and forecasting water consumption and detecting leakage using an iot system," *Water Supply*, vol. 20, no. 3, pp. 1103–1113, 2020.

[201] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: a new learning scheme of feedforward neural networks," in *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, vol. 2. IEEE, 2004, pp. 985–990.

[202] Y. Cai, X. Liu, Y. Zhang, and Z. Cai, "Hierarchical ensemble of extreme learning machine," *Pattern Recognition Letters*, vol. 116, pp. 101–106, 2018.

[203] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[204] V. N. Vapnik, "An overview of statistical learning theory," *IEEE transactions on neural networks*, vol. 10, no. 5, pp. 988–999, 1999.

[205] A. A. Aburomman and M. B. I. Reaz, "A novel svm-knn-pso ensemble method for intrusion detection system," *Applied Soft Computing*, vol. 38, pp. 360–372, 2016.

[206] X. Zhang and Q. Song, "A multi-label learning based kernel automatic recommendation method for support vector machine," *PloS one*, vol. 10, no. 4, p. e0120455, 2015.

[207] A. Anand, G. Pugalenthi, and P. Suganthan, "Predicting protein structural class by svm with class-wise optimized features and decision probabilities," *Journal of theoretical biology*, vol. 253, no. 2, pp. 375–380, 2008.

[208] C. Cortes and V. Vapnik, "Support-vector networks," *Machine learning*, vol. 20, no. 3, pp. 273–297, 1995.

[209] T. Asefa, M. Kemblowski, M. McKee, and A. Khalil, "Multi-time scale stream flow predictions: the support vector machines approach," *Journal of hydrology*, vol. 318, no. 1-4, pp. 7–16, 2006.

[210] J. S. Racine, "Rstudio: A platform-independent ide for r and sweave," *Journal of Applied Econometrics*, vol. 27, no. 1, pp. 167–172, 2012.

[211] USDA, *GAIN Report:Global Agricultural Information Network*. Foreign Agricultural Service, Department of Agriculture, Washington, DC, 2020.

[212] K. Gill, S. Sandhu, S. Mishra *et al.*, "Pre-harvest wheat yield prediction using ceres-wheat model for ludhiana district, punjab, india," *Journal of Agrometeorology*, vol. 20, no. 4, p. 319, 2018.

[213] S. Sandhu, P. Tripathi, S. Patel, R. Prasad, N. Solanki, R. Kumar, C. Singh, A. Dubey, V. Rao *et al.*, "Effect of intra-seasonal temperature on wheat at different locations of india: A study using ceres-wheat model," *Journal of Agrometeorology*, vol. 18, no. 2, p. 222, 2016.

[214] X. Zhang, S. Chen, H. Sun, L. Shao, and Y. Wang, "Changes in evapotranspiration over irrigated winter wheat and maize in north china plain over three decades," *Agricultural Water Management*, vol. 98, no. 6, pp. 1097–1104, 2011.

[215] H. Deng, "Guided random forest in the rrf package," *arXiv preprint:1306.0237*, 2013.

[216] G. Muñoz and J. Grieser, "Climwat 2.0 for cropwat," *Water Resources, Development and Management Service*, pp. 1–5, 2006.

[217] Y. Khadatare, M. Kaushal *et al.*, "Modeling of water table behaviour in district ludhiana-a case study," *Journal of Agricultural Engineering*, vol. 43, no. 4, pp. 36–42, 2006.

[218] A. Bujard and L. Rcpp, "Package 'fuger'." [Online]. Available: https://CRAN.R-project.org/package=fugeR

[219] S. R. Patel, A. and M. Siag, "Development of decision support system for on-farm irrigation water management," *International Journal of Pure Applied Bio-science*, vol. 5, no. 3, pp. 749–763, 2017.

[220] N. Shrestha and S. Shukla, "Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment," *Agricultural and Forest Meteorology*, vol. 200, pp. 172–184, 2015.

[221] L. Burton, N. Dave, R. Fernandez, K. Jayachandran, and S. Bhansali, "Smart gardening iot soil sheets for real-time nutrient analysis," *Journal of The Electrochemical Society*, vol. 165, no. 8, p. B3157, 2018.

[222] W. Wang, Y. Cui, Y. Luo, Z. Li, and J. Tan, "Web-based decision support system for canal irrigation management," *Computers and Electronics in Agriculture*, vol. 161, pp. 312–321, 2019.

[223] L. Mateos, I. López-Cortijo, and J. A. Sagardoy, "Simis: the fao decision support system for irrigation scheme management," *Agricultural Water Management*, vol. 56, no. 3, pp. 193–206, 2002.

[224] M. Maza, A. Bandyopadhyay, and A. Bhadra, "Development of gis toolbar to estimate reference evapotranspiration and net irrigation requirement on raster based approach," *Agricultural Engineering International: CIGR Journal*, vol. 22, no. 3, pp. 27–42, 2020.

[225] R. Srivastava, D. Sarkar, S. S. Mukhopadhayay, A. Sood, M. Singh, R. A. Nasre, and S. A. Dhale, "Development of hyperspectral model for rapid monitoring of soil organic carbon under precision farming in the indo-gangetic plains of punjab, india," *Journal of the Indian Society of Remote Sensing*, vol. 43, no. 4, pp. 751–759, 2015.

[226] D. Conn, T. Ngun, G. Li, and C. M. Ramirez, "Fuzzy forests: Extending random forest feature selection for correlated, high-dimensional data," *Journal of Statistical Software*, vol. 91, no. 1, pp. 1–25, 2019.

[227] P. Langfelder and S. Horvath, "Wgcna: an r package for weighted correlation network analysis," *BMC bioinformatics*, vol. 9, no. 1, pp. 1–13, 2008.

[228] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical applications in genetics and molecular biology*, vol. 4, no. 1, 2005.

[229] P. Langfelder and S. Horvath, "Wgcna: an r package for weighted correlation network analysis," *BMC bioinformatics*, vol. 9, no. 1, pp. 1–13, 2008.

[230] B. Zhang and S. Horvath, "A general framework for weighted gene co-expression network analysis," *Statistical applications in genetics and molecular biology*, vol. 4, no. 1, 2005.

[231] Z.-H. Zhou and J. Feng, "Deep forest," *National Science Review*, vol. 6, no. 1, pp. 74–86, 2019.

[232] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The bulletin of mathematical biophysics*, vol. 5, no. 4, pp. 115–133, 1943.

[233] L. V. Utkin, A. A. Meldo, and A. V. Konstantinov, "Deep forest as a framework for a new class of machine-learning models," *National Science Review*, vol. 6, no. 2, pp. 186–187, 2019.

[234] N. G. Polson and V. O. Sokolov, "Deep learning for short-term traffic flow prediction," *Transportation Research Part C: Emerging Technologies*, vol. 79, pp. 1–17, 2017.

[235] F. Hutter, H. Hoos, and K. Leyton-Brown, "An efficient approach for assessing hyperparameter importance," in *International conference on machine learning*. PMLR, 2014, pp. 754–762.

[236] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proceedings of ICNN'95-international conference on neural networks*, vol. 4. IEEE, 1995, pp. 1942–1948.

[237] M. S. Pal, *Recent Advances in Irrigation Water Management*. Kalyani Publishers, Ludhiana, Punjab, India, 2012.

[238] R. S. Ayers, D. W. Westcot *et al.*, *Water quality for agriculture*. Food and Agriculture Organization of the United Nations, Rome, 1985, vol. 29.