EVAPOTRANSPIRATION ESTIMATION USING EMPIRICAL AND MACHINE LEARNING MODELS: A COMPARATIVE ASSESSMENT OF ACCURACY AND TRANSFERABILITY

by

AFNA HAMEEM H (2022-02-007)

ABHIRAM M A (2022-02-010)

SREENANDANA P V (2022-02-013)

PROJECT REPORT

Submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

AGRICULTURAL ENGINEERING

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

DEPARTMENT OF IRRIGATION AND DRAINAGE ENGINEERING

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND FOOD TECHNOLOGY TAVANUR - 679573

MALAPPURAM, KERALA, INDIA

2026

DECLARATION

We hereby declare that this project entitled “**EVAPOTRANSPIRATION ESTIMATION USING EMPIRICAL AND MACHINE LEARNING MODELS: A COMPARATIVE ASSESSMENT OF ACCURACY AND TRANSFERABILITY**” is a Bonafide record of project work done by us during the course of study and that the report has not previously formed the basis for the award to us of any degree, diploma, associateship, fellowship or other similar title of another university or society.

Place: Tavanur

Date: 27/06/2026

AFNA HAMEEM H (2022-02-007)

ABHIRAM M A (2022-02-010)

SREENANDANA P V (2022-02-013)

CERTIFICATE

Certified that the project entitled "**EVAPOTRANSPIRATION ESTIMATION USING EMPIRICAL AND MACHINE LEARNING MODELS: A COMPARATIVE ASSESSMENT OF ACCURACY AND TRANSFERABILITY**" is a record of project work done jointly by Afna Hameem H (2022-02-007), Abhiram M A (2022-02-010), Sreenandana P V (2022-02-013) under my guidance and supervision and that it has not previously formed the basis for the award of any degree, diploma, fellowship, or associateship to them.

Place: Tavanur

Date: 27/06/2026

Dr.Asha Joseph
Professor
Department of Irrigation and Drainage Engineering
Kelappaji College of Agricultural Engineering and Food Technology

ACKNOWLEDGEMENT

First and foremost, we offer our sincere gratitude to God Almighty, whose grace made this endeavor possible and every step of it a success.

*Words would be inadequate to express our deepest gratitude and heartfelt thanks to our project guide, **Dr. Asha Joseph**, Professor, Department of Irrigation and Drainage Engineering, KCAEFT, Tavanur. Her guidance, constructive criticism, and generous assistance at every stage of this project have been invaluable. It was, in fact, her innovative ideas and approachable manner that motivated us to work under her supervision. We consider it our proud privilege to express our heartfelt indebtedness and deepest sense of gratitude to her for laying out the framework of this project work. We hold genuine admiration and regard for her wholehearted support and untiring help.*

*It is our pleasure to place on record our gratitude **to Dr. Jayan P R**, Dean of Faculty and Professor & Head, Department of Farm Machinery and Power Engineering, KCAEFT, Tavanur, for his valuable advice and support extended during this study. We also thank **Dr. Rema K P**, Professor and Head, Department of Irrigation and Drainage Engineering, for the support rendered throughout the course of this work.*

We are grateful to our parents for their constant support and encouragement throughout our period of study, and to our friends and classmates for their assistance during the course of this work.

Finally, we thank all those who, directly or indirectly, contributed to the successful completion of this project.

Afna Hameem H (2022-02-007)
Abhiram M A (2022-02-010)
Sreenandana P V (2022-02-013)

DEDICATED TO

AGRICULTURAL ENGINEERING

PROFESSION

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SYMBOLS AND ABBREVIATION

Sl. No.	Abbreviation or Notation	Description
1	%	Per cent
2	*et al.*	And others
3	°C	Degree Celsius
4	°N, °E	Degrees North, Degrees East
5	Δ	Slope of saturation vapour pressure curve (kPa °C^{-1})
6	γ	Psychrometric constant (kPa °C^{-1})
7	AEZ	Agro-Ecological Zone
8	ANN	Artificial Neural Network
9	CNN	Convolutional Neural Network
10	CNN-LSTM	Hybrid Convolutional Neural Network – Long Short-Term Memory
11	DA	Dalton model
12	DL	Deep Learning
13	DO	Dorji model
14	ea	Actual vapour pressure (kPa)
15	Epan	Pan evaporation (mm day^{-1})
16	es	Saturation vapour pressure (kPa)
17	ET	Evapotranspiration
18	ET_0	Reference evapotranspiration (mm day^{-1})
19	FAO	Food and Agriculture Organization
20	FAO-56 PM	FAO-56 Penman–Monteith model
21	G	Soil heat flux density (MJ m^{-2} day^{-1})
22	h	Hour
23	HS	Hargreaves–Samani model
24	IMD	India Meteorological Department
25	KAU	Kerala Agricultural University
26	KCAEFT	Kelappaji College of Agricultural Engineering and Food Technology
27	km h^{-1}	Kilometre per hour

28	kPa	Kilopascal
29	kPa $°C^{-1}$	Kilopascal per degree Celsius
30	LSTM	Long Short-Term Memory
31	m	Metre
32	m s^{-1}	Metre per second
33	MA	Makkink model
34	MAE	Mean Absolute Error (mm day^{-1})
35	ME	Meyer model
36	MJ m^{-2} day^{-1}	Megajoule per square metre per day
37	ML	Machine Learning
38	mm	Millimetre
39	mm day^{-1}	Millimetre per day
40	MSE	Mean Squared Error
41	MSL	Mean Sea Level
42	n	Sunshine duration (hours day^{-1})
43	P	Rainfall (mm day^{-1})
44	PT	Priestley–Taylor model
45	R^2	Coefficient of Determination
46	Ra	Extraterrestrial radiation (MJ m^{-2} day^{-1})
47	RARS	Regional Agricultural Research Station
48	ReLU	Rectified Linear Unit
49	RF	Random Forest
50	RH	Relative Humidity (%)
51	RMSE	Root Mean Square Error (mm day^{-1})
52	Rn	Net radiation at the crop surface (MJ m^{-2} day^{-1})
53	Rs	Solar radiation derived from sunshine hours (MJ m^{-2} day^{-1})
54	T	Transferability Score (= External R^2 / Internal R^2)
55	Tmean	Mean daily air temperature (°C)
56	Tmax	Maximum air temperature (°C)

57	Tmin	Minimum air temperature (°C)
58	u_2	Wind speed at 2 m height ($m\ s^{-1}$)
59	XGBoost	eXtreme Gradient Boosting

INTRODUCTION

CHAPTER I

INTRODUCTION

Water scarcity has emerged as one of the most pressing global challenges of the twenty-first century, threatening agricultural productivity, food security, and sustainable water resource management. Increasing population growth, rapid urbanisation, industrial expansion, and climate change have intensified pressure on freshwater resources worldwide. Agriculture is the largest consumer of freshwater, accounting for nearly 70% of global freshwater withdrawals, highlighting the urgent need for efficient water management strategies in irrigated agriculture (UNESCO, 2024). Improving irrigation efficiency is therefore essential for conserving water resources while sustaining agricultural production under changing climatic conditions.

Accurate estimation of crop water requirements forms the foundation of efficient irrigation planning and water resource management. Reference evapotranspiration (ET_o), defined as the evapotranspiration rate from a hypothetical reference surface under well-watered conditions, is a key parameter for estimating crop water requirements, irrigation scheduling, hydrological modelling, drought assessment, and water allocation planning (Allen *et al*., 1998). Reliable ET_o estimation is particularly important in regions experiencing increasing water stress, where optimising irrigation water use can substantially enhance agricultural water productivity.

Direct measurement of ET_o through weighing lysimeters is considered the most accurate method for quantifying evapotranspiration. However, lysimeter installations are expensive, require intensive maintenance and technical expertise, and are available only at a limited number of research locations (Allen *et al*., 2011). Consequently, indirect estimation methods have become widely adopted for operational applications. Among these, the FAO-56 Penman–Monteith (PM) method is internationally recognized as the standard approach for ET_o estimation because of its strong physical basis and consistent performance across diverse climatic conditions (Allen *et al*., 1998). The FAO-56 PM equation is generally regarded as the closest practical alternative to lysimeter-derived ET_o and is commonly used as the benchmark for evaluating other estimation methods.

Despite its accuracy, the FAO-56 PM method requires several meteorological variables, including air temperature, solar radiation, relative humidity, and wind speed. In many developing regions and data-scarce environments, complete and reliable meteorological datasets are often unavailable, limiting the application of this method (Sun *et al*., 2024). To address this challenge, several empirical ET_0 estimation methods have been developed, including the Hargreaves–Samani, Blaney–Criddle, Priestley–Taylor, and Jensen–Haise equations. These methods require fewer meteorological inputs and are easier to implement; however, their predictive performance is often highly dependent on local climatic conditions and geographical characteristics (Tabari *et al*., 2013). As a result, empirical models frequently exhibit location-specific behaviour and may require calibration before application in new environments.

The challenge of ET_0 estimation is particularly significant in India due to its vast geographical extent and remarkable agro-climatic diversity. The country encompasses a wide range of climatic zones, from humid tropical regions to semi-arid and arid environments, resulting in substantial spatial variability in evapotranspiration processes. Consequently, the accuracy of empirical ET_0 estimation methods varies considerably across regions, necessitating local validation and performance assessment. In Kerala, contrasting agro-climatic conditions exist within relatively short geographical distances, making it an ideal region for evaluating the robustness and adaptability of ET_0 estimation models.

Recent advances in artificial intelligence and machine learning (ML) have provided new opportunities for ET_0 estimation, particularly in regions where meteorological data availability is limited. Unlike traditional empirical equations, machine learning algorithms can effectively capture complex nonlinear relationships between climatic variables and evapotranspiration processes without relying on explicit physical assumptions (Li *et al*., 2025). Various machine learning techniques, including Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), Extreme Gradient Boosting (XGBoost), and other ensemble learning methods, have demonstrated promising performance in ET_0 estimation using limited climatic inputs (Wu *et al*., 2024; Khosravi *et al*., 2024). Numerous studies have reported that ML models can achieve prediction accuracies comparable to, and in some cases exceeding, those of conventional empirical methods and even the FAO-56 PM model under specific conditions (Ahmadi *et al*., 2023).

Although machine learning approaches have shown considerable potential for ET_0 estimation, an important challenge remains regarding their transferability and generalization capability. Most ML models are developed and validated using datasets from a single location and may experience significant reductions in predictive performance when applied to regions with different climatic and environmental characteristics (Li *et al*., 2025). Understanding whether a model trained in one agro-climatic region can be successfully applied to another is crucial for its operational deployment and broader applicability. However, studies examining the spatial transferability of machine learning-based ET_0 models remain relatively limited, particularly in tropical regions characterized by diverse climatic conditions.

Against this background, the present study entitled "Evapotranspiration Estimation Using Empirical and Machine Learning Models: A Comparative Assessment of Accuracy and Transferability" evaluated the performance of selected empirical and machine learning models for estimating reference evapotranspiration at Pattambi and Vellayani, two contrasting agro-climatic locations in Kerala, India. The study further investigates the comparative accuracy of empirical and machine learning approaches and assesses the spatial transferability of the developed machine learning models between the two locations. The findings are expected to provide valuable insights into the reliability, robustness, and generalisation potential of machine learning techniques for ET_0 estimation under diverse agro-climatic conditions, thereby supporting improved irrigation management and sustainable water resource planning. The specific objectives of the study are as follows.

1. To estimate reference evapotranspiration (ET_0) at Pattambi and Vellayani using selected empirical methods and machine learning (ML) models.

2. To evaluate and compare the performance of empirical and machine learning models for ET_0 estimation under varying climatic conditions.

3. To investigate the spatial transferability of the developed machine learning models by assessing their applicability and predictive reliability across different locations.

REVIEW OF LITERATURE

CHAPTER II

REVIEW OF LITERATURE

The accurate estimation of reference evapotranspiration (ET_0) is fundamental to agricultural water management, irrigation scheduling, and sustainable use of water resources. Researchers worldwide have explored a wide spectrum of approaches for ET_0 estimation, from physically-based methods such as the FAO-56 Penman–Monteith equation to advanced machine learning models including Random Forest, CNN, LSTM, and hybrid architectures. Studies relevant to the objectives of the present investigation are reviewed below, classified under appropriate sub-headings arranged in chronological order within each section.

2.1 ESTIMATION OF REFERENCE EVAPOTRANSPIRATION BY FAO 56 PENMAN-MONTEITH METHOD

The FAO-56 Penman–Monteith (FAO-56 PM) method is universally recognised as the most physically sound and accurate standard for estimating reference evapotranspiration. It integrates radiation, temperature, humidity, and wind speed into a single physically based equation. Several studies have analysed the sensitivity of the FAO-56 PM method to its input variables and compared it with simpler empirical models to identify practical alternatives in data-limited situations.

Debnath *et al*. (2015) analysed the sensitivity of the FAO-56 Penman–Monteith method to climatic variables across five agro-ecological stations in India, representing semi-arid (Kovilpatti and Parbhani), humid (Mohanpur), and sub-humid (Ludhiana and Ranichauri) climate zones. Five years of daily climatic data comprising maximum and minimum temperature, solar radiation, relative humidity, and wind speed were used as inputs for ET_0 estimation. Sensitivity was evaluated by systematically increasing and decreasing each climatic variable by one unit up to five units while keeping other variables constant. Results showed that the change in ET_0 was linearly related to changes in all climatic variables, with the coefficient of determination exceeding 0.97 at all sites. ET_0 was found to be most sensitive to solar radiation at Kovilpatti, Mohanpur, and Ranichauri, while wind speed was the most sensitive variable at Ludhiana and Parbhani. ET_0 was least sensitive to relative humidity and minimum temperature at all sites. The study concluded that accurate measurement of solar radiation, wind speed, and maximum temperature is critical for reliable FAO-56 PM ET_0 estimation across diverse agro-ecological zones in

India, and that sensitivity patterns vary significantly from site to site and seasonally within a site.

Hadria *et al.* (2021) conducted a comparative assessment of reference evapotranspiration models at 22 weather stations in Morocco representing arid and semi-arid climatic conditions. Five empirical temperature-based ET_o estimates were compared against the standard FAO Penman–Monteith estimate using multiple statistical performance indices. A significant positive correlation was established between ET_o-PM and solar radiation, average temperature, and maximum air temperature across all stations. Among the empirical models evaluated, the Dorji estimate showed relatively better precision and stability, though it required advanced calibration for arid and semi-arid conditions. Following systematic calibration, the authors derived a new estimate designated as ET_o-Hadria, which demonstrated improvement in the quality and precision of ET_o assessment at approximately 68 per cent of the evaluated stations. The study highlighted that simple calibrated empirical estimates can function as transferable tools for ET_o assessment in regions where comprehensive meteorological data are unavailable, but also confirmed that the FAO-56 PM equation remains the most reliable benchmark and that its data requirements necessitate the development of data-driven alternatives.

Raja *et al.* (2024) evaluated the performance of different empirical ET_o models for the tropical highland region of Udhagamandalam (The Nilgiris) in Tamil Nadu, South India, using 60 years of meteorological data from 1960 to 2020. Eight temperature-based and ten radiation-based empirical models were evaluated against ET_o estimates derived from pan evaporation observations and the FAO Penman–Monteith method. A comprehensive set of statistical error metrics was used for model ranking. Results showed that both temperature-based and radiation-based models performed well for the study region; however, radiation-based models outperformed temperature-based models consistently. The superior performance of radiation-based models was attributed to the high humidity levels prevailing throughout the year in the region, with solar radiation serving as the dominant driver of the surface energy balance and hence ET_o. The study concluded that simple temperature-based and radiation-based empirical models using minimum meteorological information are adequate for ET_o estimation in tropical humid highland climates of South India, and can be practically applied in agricultural water management and irrigation scheduling. The findings have direct relevance to humid tropical regions of peninsular India such as Kerala, where similar climatic conditions prevail.

2.2 ESTIMATION OF REFERENCE EVAPOTRANSPIRATION USING EMPIRICAL MODELS WITH REDUCED INPUT VARIABLES

A major challenge in practical ET_0 estimation is the unavailability of complete meteorological datasets, particularly in developing and data-scarce regions. Empirical ET_0 models have been developed to address this limitation by utilising specific subsets of meteorological variables. These models are broadly classified into three categories: temperature-based models, which require only minimum and maximum temperature and extraterrestrial radiation; mass transfer-based models, which additionally incorporate relative humidity; and radiation-based models, which require temperature and solar radiation. The performance of these reduced-input empirical models and their comparison with machine learning approaches using the same restricted input variables have been the subject of several important studies.

Hadria *et al*. (2021) examined the performance variation of temperature-based empirical ET_0 models under data-limited conditions at 22 stations in Morocco, where full meteorological datasets were not always available. By relying only on temperature and related variables as inputs, the study showed that reduced-parameter empirical models performed inconsistently across different stations, with significant estimation errors arising from the absence of radiation and wind speed data. The Dorji model, which uses only temperature and extraterrestrial radiation, showed the best performance among the temperature-based approaches. Following local calibration using the FAO-56 PM equation as reference, all temperature-based models showed significant improvement in precision, demonstrating that the accuracy of reduced-input empirical models is strongly dependent on site-specific calibration. The study concluded that radiation-related inputs were the most critical missing factor in temperature-only models, and their absence significantly limits ET_0 estimation accuracy in arid climatic conditions.

Mandal *et al*. (2024) systematically evaluated three categories of reduced-input empirical ET_0 models at Pusa and Nagpur stations in India. Temperature-based empirical models -Hargreaves-Samani, Dorji, and Blaney-Criddle -used only T_{min}, T_{max}, T_{mean}, and extraterrestrial radiation (R_a). Mass transfer-based models -Dalton, Rohwer, and Meyer -additionally incorporated relative humidity (RH). Radiation-based models -Priestley-Taylor, Ritchie, and Makkink -used temperature and solar radiation (R_s). Among the three empirical categories, radiation-based models consistently outperformed temperature-based and mass transfer-based models at both stations, with the Ritchie model achieving R^2 of

0.912 and RMSE of 0.493 mm/day at Pusa. Notably, the inclusion of an additional variable (relative humidity) in mass transfer-based models over temperature-based models did not significantly improve empirical ET_0 estimation accuracy. All categories of empirical models were consistently outperformed by corresponding ML models trained on the same reduced input sets, clearly establishing the superiority of data-driven approaches even under limited input variable conditions.

Raja *et al*. (2024) evaluated eight temperature-based and ten radiation-based empirical ET_0 models in the tropical highland region of Udhagamandalam, Tamil Nadu, South India, using 60 years of meteorological data from 1960 to 2020. Temperature-based models used only minimum and maximum temperature as inputs, while radiation-based models additionally incorporated solar radiation. Both categories were evaluated against FAO-56 PM estimates and pan evaporation observations using a comprehensive set of statistical error metrics. Results showed that radiation-based models outperformed temperature-based models consistently across all seasons, which was attributed to the dominant role of solar radiation in driving ET_0 in humid tropical conditions. However, temperature-based models also demonstrated acceptable performance, confirming their utility when radiation data are unavailable. The study concluded that in humid tropical climates of South India, both radiation-based and temperature-based reduced-input empirical models can support practical ET_0 estimation and irrigation management. These findings are directly relevant to the study of ET_0 across the contrasting humid and semi-arid agro-climatic zones of Kerala in the present study.

2.3 ESTIMATION OF REFERENCE EVAPOTRANSPIRATION USING MACHINE LEARNING MODELS

Machine learning (ML) models have emerged as powerful alternatives to empirical ET_0 estimation methods owing to their ability to capture complex nonlinear relationships among meteorological variables without requiring explicit physical assumptions. Several studies have demonstrated that ML models consistently outperform empirical equations across diverse climatic regions, with research spanning multiple model types and input variable combinations.

Ayaz *et al*. (2021) evaluated machine learning techniques for estimating ET_0 with minimal climatic inputs at two climatically distinct stations -Hyderabad in India and Waipara in New Zealand. The FAO-56 Penman–Monteith model was adopted as the benchmark. Four ML models -Long Short-Term Memory neural networks (LSTM), Gradient Boosting Regressor (GBR), Random Forest (RF), and Support Vector Regression (SVR) -were developed and compared under varying input combinations. The results showed that 99 per cent prediction accuracy was achievable with all climatic inputs, while accuracy dropped to approximately 86 per cent with fewer parameters. The LSTM model outperformed all other models across all input combinations at both stations, followed by SVR and RF. The study demonstrated that even a two-parameter combination of temperature and relative humidity could yield promising ET_0 estimates, providing practical solutions for agricultural water management in semi-arid, data-scarce regions. Both LSTM and SVR were identified as the most robust models for ET_0 estimation with minimal climate data.

Chen *et al*. (2020) compared three deep learning models -Deep Neural Network (DNN), Temporal Convolution Network (TCN), and Long Short-Term Memory (LSTM) -alongside two classical ML models -Support Vector Machine (SVM) and Random Forest (RF) -for estimating daily ET_0 using limited meteorological data across the Northeast plain of China. Seven empirical equations spanning temperature-based, radiation-based, and humidity-based categories were also evaluated using the FAO-56 PM ET_0 as the reference standard. Two evaluation strategies were adopted: within-station training and testing, and a cross-station strategy in which stations in each group took turns being the test set for models trained on the remaining stations. Results showed that TCN and RF demonstrated R^2 values significantly higher than the best empirical model (Modified Hargreaves) by 0.048 and 0.035 respectively. In the cross-station strategy, TCN and LSTM outperformed empirical models with RMSE reductions of 0.069 and 0.079 mm/day respectively. Overall, all proposed ML and DL models outperformed empirical models in both strategies, with the TCN model emerging as the most consistently superior approach. The study established that DL models, particularly TCN and LSTM, can be effectively applied beyond their study areas with reliable performance.

Chang *et al*. (2025) presented a comprehensive review of machine learning models for reference crop evapotranspiration modelling and prediction, analysing 325 studies from the Web of Science database spanning 2001 to 2024. Major ML methods reviewed included ANN, ensemble learning approaches such as Random Forest and XGBoost, and deep learning models including CNN and LSTM. The review identified a significant acceleration in publications after 2020, with 54 papers published in 2023 alone. Temperature variables ($T_{m}^{I}{}_{n}$, T_{ma}^{x}, $T_{m}^{e}{}_{an}$) were identified as the most consistently critical inputs across all ML model types, followed by relative humidity, wind speed, and solar radiation. Optimal input variable combinations were found to vary by climatic zone, with temperature and relative humidity being more critical in arid regions, while solar radiation is more important in humid zones. The review noted that LSTM models inherently handle sequential dependencies and manage irrelevant inputs effectively, while Random Forest suffers performance degradation with numerous irrelevant features. The authors concluded that ML models offer more robust and precise ET_0 predictions than empirical equations, particularly in complex climatic environments, and identified input variable selection and spatial transferability as key future research directions.

2.3.1 Estimation of reference evapotranspiration using random forest model

Random Forest (RF) is a robust ensemble machine learning algorithm that constructs multiple decision trees during training and outputs the mean prediction for regression tasks. It offers significant advantages including resistance to overfitting, ability to handle nonlinear interactions among meteorological variables, capacity for variable importance assessment, and lower uncertainty in predictions compared to many individual models.

Wu *et al*. (2020) evaluated the potential of the Random Forest predictive model for simulating daily reference evapotranspiration at two stations in the arid oasis region of northwestern China. The FAO-56 Penman–Monteith ET_0 was used as the reference standard for training and evaluation. Input variables including maximum and minimum temperature, sunshine duration, wind speed, and relative humidity were tested in different combinations to assess the effect of input reduction on model performance. A Monte Carlo uncertainty analysis was incorporated to verify the reliability and stability of model outputs. Results demonstrated that the RF model was a reliable and accurate approach for ET_0 prediction in the arid oasis area under limited data conditions. Relative humidity was identified as the most influential meteorological variable governing ET_0 behaviour in the arid study region, followed by air temperature. The uncertainty analysis confirmed that the RF model exhibited lower uncertainty compared to other models, establishing it as a sound and practical modelling approach for ET_0 prediction in arid regions with limited but reliable weather datasets.

Das *et al*. (2023) developed a Random Forest-based ET_0 model for the semi-arid region of Punjab in northwestern India, focusing on the practical requirement for accurate ET_0 estimation from farm-level available meteorological data. The RF model was calibrated and validated using meteorological data collected from the weather station of

Punjab Agricultural University for 21 years (1990–2010) and nine years (2011–2019) respectively, with the FAO-56 PM model taken as the standard. For comparison, the Hargreaves–Samani (HS), Modified Penman, and Modified Hargreaves–Samani models were also evaluated. The RF model demonstrated the lowest mean absolute error (MAE) and root mean square error (RMSE) of 0.95 mm/day and 1.32 mm/day respectively, with an R^2 value of 0.92. The RF-predicted ET_0 values were subsequently used for irrigation scheduling of maize and wheat crops in the 2020–2021 growing seasons in field trials. Field trial results showed no appreciable yield reduction in either crop compared to FAO-56 PM-based irrigation scheduling, demonstrating the practical applicability of the RF model for irrigation management in the semi-arid region of Punjab. The study concluded that RF models trained on limited, farm-level meteorological inputs can substitute for the FAO-56 PM equation in data-constrained irrigation planning in Indian conditions.

Mandal *et al*. (2024) evaluated six ML models including tree-based, kernel-based, and neural network-based algorithms for ET_0 estimation across three input categories - temperature-based, mass transfer-based, and radiation-based -at Pusa and Nagpur meteorological stations in India. Within the evaluated models, the Extra Trees Regressor (ETR) and Gradient Boosting Regressor (GBR), both ensemble tree-based models closely related to the Random Forest approach, demonstrated consistently competitive performance across all three input categories. For temperature-based inputs, ETR achieved R^2 values of 0.826 and 0.832 at Pusa and Nagpur respectively, while GBR achieved 0.817 and 0.823 respectively. For radiation-based inputs, ETR reached R^2 values of 0.951 and 0.909, and GBR reached 0.951 and 0.900, at Pusa and Nagpur respectively. The study established that tree-based ensemble models offer reliable performance across all input categories and climatic conditions, though neural network-based models demonstrated marginal advantages particularly for radiation-based inputs. These findings underscore the robustness of ensemble tree-based approaches for ET_0 estimation under the three standardised input categories used in the present study.

2.3.2 Estimation of reference evapotranspiration using XGBoost model

Extreme Gradient Boosting (XGBoost) is a powerful ensemble machine learning algorithm that employs gradient descent optimisation to iteratively combine weak learners (decision trees) into a strong predictive model. XGBoost offers significant computational advantages over conventional boosting methods through built-in regularisation (L1 and L2) to prevent overfitting, parallel processing, and efficient handling of missing data. Its high scalability, predictive accuracy, and built-in feature importance ranking have made it one of the most widely adopted ML models for regression tasks involving meteorological time series data, including ET_0 estimation.

Mohammadi *et al*. (2024) proposed a novel explainable hybrid framework for estimating daily reference evapotranspiration by combining the XGBoost model with the Nelder–Mead (NM) optimisation algorithm at two meteorological stations, Shiraz and Fasa, in Iran. The study employed sixty meteorological variables categorised into solar and cloud-based, temperature-based, wind and humidity-based, and pressure-based groups to comprehensively assess their influence on ET_0 estimation. Three feature selection methods -Gradient Boosting Machine, Kendall's Tau, and Relief Algorithm -were applied to identify the most influential predictors, with variables having normalised weights equal to or greater than 0.9 selected as model inputs, resulting in the top ten per cent of variables being utilised. The XGBoost models were developed in three progressively refined levels: standalone XGBoost with selected inputs, wavelet-based hybrid XGBoost, and XGBoost coupled with the NM algorithm. Results indicated that the Kendall's Tau-based feature selection provided the most accurate predictions at Shiraz, achieving RMSE of 0.721 mm/day for solar and cloud-based variables. The XGBoost–NM hybrid model further demonstrated significant improvements, reducing RMSE to 0.091 mm/day and 0.155 mm/day for the testing period at Fasa and Shiraz respectively. The study established that integrating XGBoost with the NM optimisation algorithm produces a model that is both interpretable and highly accurate for ET_0 estimation, making it a practical and explainable alternative to black-box deep learning models for irrigation water management applications.

2.3.3 Estimation of reference evapotranspiration using CNN model

Convolutional Neural Networks (CNN) are deep learning architectures that can automatically extract local temporal features from time series data without manual feature engineering. In the context of ET_0 estimation, one-dimensional CNN models have demonstrated strong capability for learning patterns from sequences of meteorological inputs, often matching or exceeding LSTM performance while offering faster training times.

Chen *et al*. (2020) developed and evaluated the Temporal Convolution Network (TCN), a CNN-based architecture, alongside LSTM and DNN for daily ET_0 estimation from limited meteorological data across stations in Northeast China. The TCN model, which applies dilated causal convolutions across temporal sequences, consistently emerged as the best or jointly best-performing model across all three input categories -temperature-based, radiation-based, and humidity-based -in both within-station and cross-station evaluation strategies. For radiation-based inputs, TCN achieved R^2 values significantly higher than the best empirical model, with RMSE reductions of 0.069 mm/day in the cross-station strategy. The TCN architecture demonstrated that CNN-based temporal models can capture complex nonlinear meteorological patterns effectively and generalise well beyond the training station, making them highly suitable for ET_0 estimation in data-scarce regions where cross-station application is necessary.

Li *et al*. (2024) proposed improved CNN and LSTM models incorporating empirical mode decomposition (EMD) and wavelet threshold denoising (WD) as preprocessing techniques for predicting reference crop evapotranspiration at five meteorological stations in Northeast China. Daily and monthly ET_0 data from 1952 to 2020 were used, with FAO-56 PM ET_0 as the target standard. Two hybrid models -EMD–WD–CNN and EMD–WD–LSTM -were constructed by decomposing ET_0 time series into intrinsic mode function components using EMD, followed by wavelet denoising of high-frequency components, before feeding the processed sequences into CNN and LSTM architectures. Results demonstrated high consistency between predicted values and FAO-56 PM estimates, with the models achieving R^2 values of 0.86–0.95 for daily ET_0 and 0.91–0.95 for monthly ET_0 across the five stations. The EMD–WD–CNN model was found most suitable for daily scale ET_0 prediction across the region. The study demonstrated that combining signal decomposition with CNN architecture significantly improves ET_0 prediction accuracy by reducing noise in meteorological time series.

Dong *et al*. (2024) developed a novel deep learning model designated MA-CNN-BiLSTM, which incorporates a Multi-Head Attention mechanism (MA), Convolutional Neural Network (CNN), and Bidirectional Long Short-Term Memory network (BiLSTM), for estimating ET_0 at 600 meteorological stations across four climatic zones of China during 1961–2020. The study evaluated three input categories: radiation-based (Rn-based), humidity-based (RH-based), and temperature-based (T-based) input combinations, and compared the MA-CNN-BiLSTM against CNN-BiLSTM, BiLSTM, LSTM, Multivariate Adaptive Regression Splines (MARS), and empirical models under both internal (within-station) and external (cross-station) cross-validation strategies. Results showed that MA-CNN-BiLSTM achieved superior performance with R^2 values ranging from 0.877 to 0.972 for internal strategy and 0.797 to 0.927 for external cross-validation across the four climatic zones. The Rn-based MA-CNN-BiLSTM excelled in temperate continental and mountain plateau zones, while RH-based inputs yielded best precision in other zones. The study conclusively demonstrated that incorporating a CNN layer for feature extraction alongside LSTM for temporal learning and an attention mechanism for input weighting delivers superior ET_0 estimation across diverse climatic conditions, with consistent performance under cross-station evaluation.

2.3.4 Estimation of reference evapotranspiration using hybrid deep learning models.

Hybrid deep learning architectures that combine the local feature extraction capability of Convolutional Neural Networks (CNN) with the sequential temporal learning strength of Long Short-Term Memory (LSTM) networks have demonstrated superior accuracy in ET_0 estimation compared to individual models. These CNN-LSTM hybrid architectures are particularly effective under limited data conditions and have been increasingly applied across diverse agro-climatic regions globally.

Sharma *et al*. (2022) presented two hybrid deep neural network models for estimating reference evapotranspiration at Ludhiana and Amritsar stations in Punjab, India. The study developed and compared the Convolution–Long Short Term Memory (Conv-LSTM) model, which performs convolution operations within LSTM cells, and the Convolutional Neural Network–LSTM (CNN-LSTM) model, which applies a convolution layer for feature extraction before LSTM layers for temporal learning. The investigation specifically focused on climate data scarcity conditions, testing different input combinations including maximum and minimum temperature, wind speed, solar radiation, relative humidity, vapour pressure, and sunshine hours. Temperature and radiation were identified as the prime inputs required for accurate ET_0 estimation. Both hybrid models outperformed empirical approaches including Hargreaves, Makkink, and Ritchie models, with Conv-LSTM emerging as the best overall model. The study concluded that hybrid CNN-LSTM architectures are highly effective for ET_0 estimation under data-scarce conditions at Indian meteorological stations, with significant implications for irrigation planning in data-limited Indian agricultural regions.

Woo (2025) developed a hybrid CNN–LSTM framework for estimating daily actual evapotranspiration (ET_a) directly from widely available station-type meteorological variables, using data from 167 FLUXNET sites spanning diverse global biomes. Input variables included radiation, temperature, vapour pressure deficit, wind speed, precipitation, atmospheric pressure, soil moisture, and static soil texture. The CNN–LSTM model achieved a validation R^2 of 0.92 and RMSE of 0.38 mm d^{-1} on withheld tower data from globally distributed sites not used during training. An ablation analysis confirmed that solar radiation, atmospheric pressure, wind speed, and vapour pressure deficit were the most informative input variables. The model was subsequently deployed with ERA5 reanalysis data to produce global ET estimates whose spatial patterns closely matched ERA5 benchmarks. The CNN–LSTM framework demonstrated strong spatial generalisation capability, performing consistently across diverse climatic regions. The study established that hybrid CNN–LSTM architectures trained on large, geographically diverse datasets can produce reliable ET estimates from routinely available meteorological observations across globally distributed locations.

Chen *et al*. (2026) developed a hybrid CNN–LSTM–Attention deep learning architecture for high-accuracy reference crop evapotranspiration estimation integrated with an optimised irrigation scheduling system. The model combined convolutional layers for local feature extraction, LSTM layers for capturing long-term temporal dependencies, and an attention mechanism to assign differential weights to input time steps. Key climate input parameters -average temperature, sunshine duration, and average wind speed -were selected through Grey Relational Analysis. The model was evaluated using data from 2012 to 2023 at Shaoxing in Zhejiang province, China, and outperformed traditional RNN, LSTM, and standard CNN–LSTM models in ET_0 estimation accuracy. When integrated with an Alternate Wetting and Drying irrigation scheduling system, the model achieved water savings of 16.5 per cent and 37.0 per cent during the 2022 and 2023 rice growing seasons respectively. The study demonstrated that incorporating an attention mechanism into the CNN–LSTM architecture further enhances ET_0 estimation accuracy and enables intelligent, data-driven irrigation scheduling in data-scarce and water-sensitive agricultural regions.

2.4 SPATIAL TRANSFERABILITY OF MACHINE LEARNING MODELS FOR ET_0 ESTIMATION

Spatial transferability refers to the ability of a machine learning model trained on data from one location to accurately predict ET_0 at geographically distinct locations with different climatic characteristics. This dimension of model performance is of critical practical importance for data-scarce regions where location-specific model training is not feasible, and has attracted increasing research attention in recent years.

Mandal *et al*. (2024) conducted a rigorous assessment of spatial transferability of ML models for ET_0 estimation at Pusa and Nagpur meteorological stations in India using three input categories: temperature-based (T_{min}, T_{max}, T_{mean}, R_a), mass transfer-based (additionally incorporating RH), and radiation-based (additionally incorporating R_s). Six ML models were evaluated, with transferability quantified as the ratio of external to internal model performance (T = External R^2 / Internal R^2), where a value of 1.0 represents perfect transferability. For models trained at Pusa and applied to Nagpur, CNN demonstrated the highest transferability of 1.0 for temperature-based inputs, while for radiation-based inputs, SVR and CNN achieved equal transferability of 0.94. The external performance of the best-performing Pusa-trained models consistently surpassed the performance of empirical models at Nagpur, with improvements of 5.7 per cent, 6.6 per cent, and 1.8 per cent for temperature-based, mass transfer-based, and radiation-based categories respectively. The study concluded that ML models, particularly neural network-based architectures, offer significant potential for spatial transferability across Indian agro-climatic regions, enabling reliable ET_0 estimation at data-scarce locations.

Dong *et al*. (2024) assessed the spatial transferability of the MA-CNN-BiLSTM model and four other deep learning models across 600 meteorological stations in four climatic zones of China, using both internal (within-station) and external (cross-station) cross-validation strategies. The internal strategy assigned 48 years of data for training and 12 years for testing within each station, while the external strategy used data from all stations in a zone for training and withheld a subset of stations for testing. Results showed that the internal strategy outperformed the external strategy by 2.74–106.04 per cent in R^2, reflecting the inherent challenge of cross-station generalisation. Despite this gap, the MA-CNN-BiLSTM model under external validation achieved R^2 values ranging from 0.797 to 0.927 across the four climatic zones, demonstrating that CNN-based hybrid architectures retain reasonable spatial transferability even when deployed at stations not seen during training. The Rn-based input combination showed the best external performance in temperate continental and mountain plateau zones. The study provided a large-scale, multi-zone empirical validation of the spatial transferability of CNN-BiLSTM architectures, confirming their robustness for ET_0 estimation across diverse climatic regions of China and supporting their potential application in data-scarce regions globally.

MATERIALS AND METHODS

CHAPTER III

MATERIALS AND METHODS

This chapter describes the materials and methods adopted for carrying out the study. It begins with a detailed description of the study area, including its geographical location and climatic characteristics. The meteorological data collected from the RARS Pattambi and Vellayani stations, which were used for the analysis, are then presented. Subsequently, the procedures employed for the computation of reference evapotranspiration and the selection of input variables are discussed. This chapter further outlines the conceptual framework of the study and the development of the machine learning models. Details regarding model training, validation, and performance evaluation are also provided. Finally, the methods adopted for assessing the spatial transferability of the machine learning models were described.

3.1 STUDY AREA

3.1.1 Study Area Details and Location

The study was conducted at two meteorological stations located in Kerala, - Pattambi in Palakkad district and Vellayani in Thiruvananthapuram district. These two locations were selected because of their contrasting agroclimatic conditions and data availability for reference evapotranspiration modelling.

Pattambi is situated in the Palakkad district of Kerala, lying in the valley of the Bharathapuzha River, one of the major west-flowing rivers of Kerala. It is geographically located at approximately 10.7870° N latitude and 76.1843° E longitude, at an elevation of approximately 63 m above mean sea level (MSL) and it belongs to Agroecological Zone 5 (Palakkad plains). The area falls within the transition zone between the Western Ghats and the plains, influenced by both the southwest and northeast monsoons. The geographic area around Pattambi is characterized by flat alluvial plains, primarily used for paddy cultivation and horticulture. The Pattambi block comes under the Malappuram-Palakkad agroclimatic zone, and the region experiences a humid tropical climate with an average annual rainfall ranging from 2000 to 2500 mm.

Vellayani is located in the Thiruvananthapuram district of Kerala, situated at approximately 8.3906° N latitude and 76.9719° E longitude, at an elevation of approximately 6 m above mean sea level (MSL). It falls under Agroecological Zone

1(coastal plain) and lies in the southern tip of the Indian peninsula, in close proximity to the Kerala Agricultural University -College of Agriculture, Vellayani, making it one of the most well-monitored agricultural meteorological stations in southern Kerala. The area is characterized by laterite soils and undulating terrain typical of the coastal and sub-coastal zones of Thiruvananthapuram. The region experiences a tropical humid climate, strongly influenced by the Arabian Sea branch of the southwest monsoon, with average annual rainfall ranging from 2500 to 2800 mm. The location map is shown in Fig 3.1

Fig 3.1 Location map of the study area (Pattambi and Vellayani)

3.1.2 Climate

The study areas Pattambi fall under tropical dry and wet climate where as vellayani is tropical humid climate. The average annual rainfall at Pattambi is about 2400 mm and at Vellayani is about 1800 mm and annual relative humidity is 72% and 83% where as the average annual wind speed is about 7 km/hr and 5 km/hr respectively.

3.2 DATA USED

Daily meteorological observations spanning 30 years (1995 - 2025) were obtained from RARS Pattambi and College of Agriculture Vellayani.It is shown in table 3.1.

Table 3.1 Data period details for Pattambi and Vellayani stations

	Pattambi	Vellayani
Total period of data	01-01-1995 to 31-12-2025	01-01-1995 to 31-12-2025
Total length of data in days (including missing variables)	11322	11323

The daily meteorological data for RARS Pattambi were obtained from the agrometeorological observatory maintained by the Regional Agricultural Research Station, Pattambi, Kerala Agricultural University. The daily meteorological data for Vellayani were obtained from the Agrometeorological Observatory of the College of Agriculture, Vellayani, Kerala Agricultural University. Both observatories maintain continuous daily records of standard meteorological variables measured using standard instruments following the guidelines of the India Meteorological Department (IMD). The daily meteorological dataset collected from both RARS Pattambi and Vellayani stations comprised the following climate variables:

1. Maximum Air Temperature (c; °C)
2. Minimum Air Temperature (T_{min}; °C)
3. Morning Relative Humidity (RH morning; %)
4. Afternoon Relative Humidity (RH afternoon; %)
5. Wind Speed (u; km h^{-1})
6. Sunshine Duration (n; hours day^{-1})
7. Rainfall (P; mm day^{-1})
8. Pan Evaporation (E_{pan}; mm day^{-1})

3.3 CONCEPTUAL FRAMEWORK OF THE STUDY

The overall flowchart of the methodology is shown in Fig.3.2. The step-by-step details are explained in the following subheads.

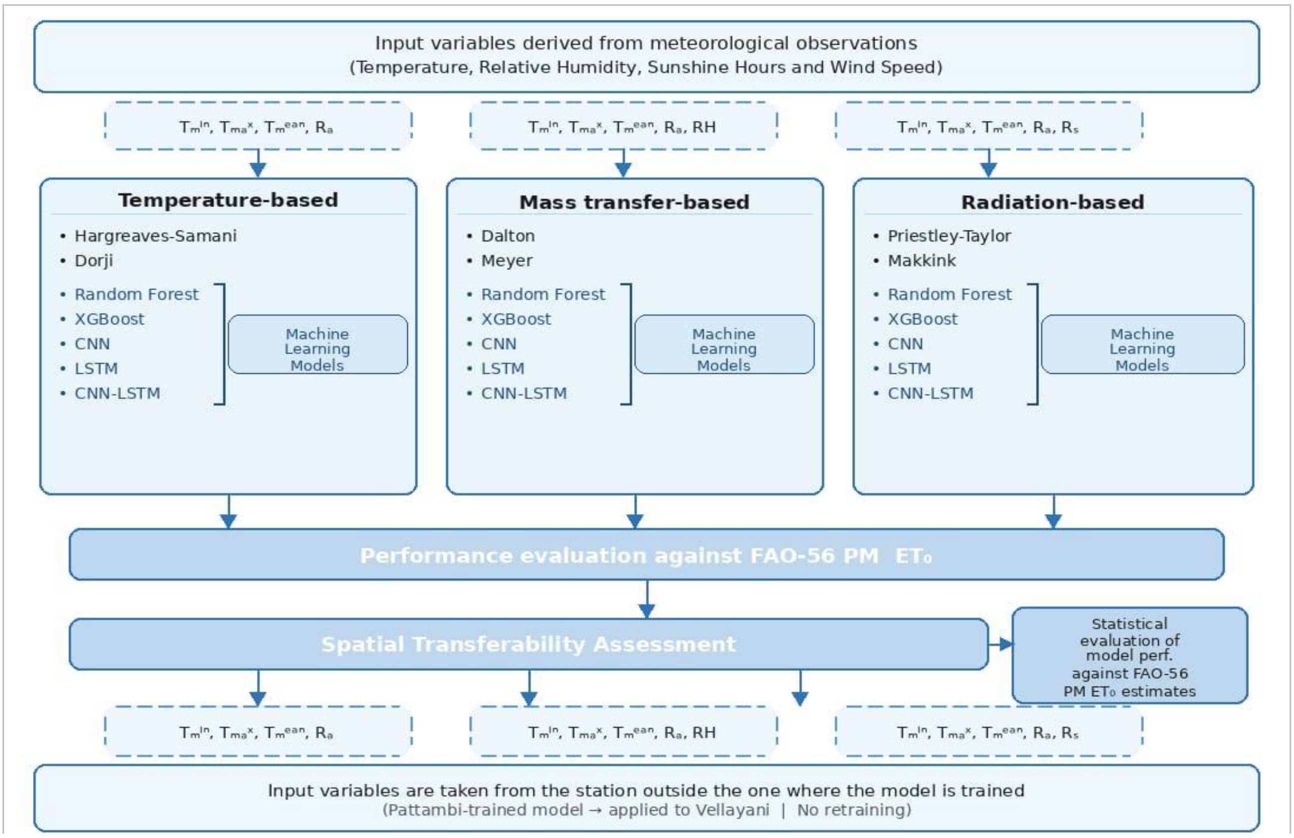

Fig 3.2. Overall flowchart of methodology

3.3.1 Various steps in the workflow

- ***Step 1 - Details of Study Area***

The study was conducted at two locations in Kerala: Pattambi (Palakkad) and Vellayani (Thiruvananthapuram), representing contrasting inland and coastal climatic zones. These diverse stations ensure models are tested under varied meteorological conditions and support spatial transferability analysis.

- ***Step 2 - Data Collection***

Daily meteorological data spanning 30 years (1995–2025) were collected from both stations, including maximum temperature, minimum temperature, mean temperature, relative humidity, wind speed, and sunshine hours. This long-term record captures wide seasonal and climatic variabilities essential for training robust ML models.

- ***Step 3 - Data Preprocessing***

Raw data were preprocessed through missing value treatment. Missing values in the dataset were filled using linear interpolation, where each missing value was estimated as the average of its immediate preceding and following values. This method preserved the continuity of the time-series data without introducing significant bias, and was applied only to short, isolated gaps.), Min-Max normalization (scaling inputs to 0–1), and chronological splitting into training and testing subsets were done. This pipeline ensures data quality, feature uniformity, and temporal integrity before model development.

- ***Step 4 - ET_0 Estimation using FAO-56 PM (Target Variable)***

Reference evapotranspiration (ET_0) was computed using the FAO-56 Penman–Monteith equation and used as the target variable for all models. This internationally recognized standard integrates temperature, humidity, wind speed, and solar radiation to produce a physically sound benchmark for model training and evaluation.

- ***Step 5 - Empirical Model Categorization***

Various empirical ET_0 estimation models were categorized based on the type of input variables they require. The categories include Full Variable-based models (requiring all meteorological inputs), Temperature-based models, Mass Transfer-based models, and Radiation-based models. This classification allows systematic comparison of models with different data requirements and complexity levels.

- ***Step 6 - Machine Learning Models***

Five different machine learning architectures were selected for ET_0 estimation. They are Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), a hybrid CNN-LSTM model, XGBoost (XGBoost), and Random Forest (RF). These models were chosen to represent a diverse range of approaches: deep learning, hybrid, and ensemble tree-based methods. The inclusion of both neural networks and ensemble methods allows a comprehensive evaluation of different computational strategies.

- ***Step 7 - ML Model development***

The machine learning models were developed by carefully defining training strategies, selecting appropriate hardware configurations, and using suitable software frameworks. Hyperparameter tuning and architecture design were carried out to optimize each model's performance. The development process ensured that models were trained consistently across both study locations for fair comparison. Proper software tools and computational resources were employed to handle the large 30-year daily dataset efficiently.

- ***Step 8 - Performance Assessment***

All models were evaluated using R^2, RMSE, and MAE metrics that collectively assess prediction accuracy and error magnitude. Performance was measured separately for training and testing phases at both Pattambi and Vellayani stations.

- ***Step 9 - ML models vs Empirical models***

A comprehensive graphical and statistical comparison was performed between the machine learning models and the empirical models across all categories. This comparative analysis helps determine whether data-driven ML approaches outperform traditional physics-based empirical equations in estimating ET_0.

- ***Step 10 - Spatial transferability***

The spatial transferability of the best-performing models were assessed by evaluating their internal performance (trained and tested at the same station) and external performance (a model trained at one station was tested at the other station). This cross-location testing reveals how well the models generalize to unseen geographical and climatic conditions.

- ***Step 11 - ET_0 Prediction for 2025 and Validation***

The best-performing machine learning models were further used to predict ET_0 values for the year 2025 using the most recently available observed meteorological data. These predictions were then validated against ET_0 values computed using the FAO-56 Penman–Monteith equation from actual 2025 field data. This final validation step confirms the real-world forecasting ability and temporal generalizability of the developed models beyond the training period.

3.4 DETAILS OF REFERENCE EVAPOTRANSPIRATION MODELS USED IN THE STUDY

Reference evapotranspiration (ET_0) can be estimated using a range of methods, broadly classified into physically based methods and empirical methods. In the present study, the FAO-56 Penman–Monteith method was used as the standard reference for computing the target ET_0 values, The empirical models are grouped into 4 categories: 1. Full variable based 2. Temperature based 3. Mass transfer based 4. Radiation-based. All the empirical models are evaluated and compared with machine learning models. The models used in the present study are described in the following subsections.

3.4.1 FAO-56 Penman–Monteith Model (FAO-56 PM)

The FAO-56 Penman–Monteith (FAO-56 PM) model is universally accepted as the most physically sound and accurate method for estimating reference evapotranspiration and has been recommended by the Food and Agriculture Organization of the United Nations as the standard method for ET_0 computation. This ET_0 estimation method considered closest practical alternative to lysimeter derived ET_0 (Allen *et al*., 1998). This model integrates both the energy balance and aerodynamic components of evapotranspiration and has been validated across diverse climatic regions worldwide (Ayaz *et al*., 2021). In the present study, the reference evapotranspiration values estimated using the FAO-56 PM equation were used as the target variable (observed variable) for training and evaluating all empirical and machine learning models. The FAO-56 PM equation is expressed as:

$$ET_0 = \frac{[0.408\Delta(R_n - G) + \gamma(900 / T_{mean} + 273)\, u_2(es - ea)]}{[\Delta + \gamma(1 + 0.34u_2)]} \quad .....(1)$$

Where:

- ET_0 = reference evapotranspiration (mm day^{-1})
- R_n= net radiation at the crop surface (MJ m^{-2} day^{-1})
- G = soil heat flux density (MJ m^{-2} day^{-1}), assumed zero for daily computations
- T_{mean}= mean daily air temperature at 2 m height (°C)
- u_2 = wind speed at 2 m height (m s^{-1})
- es = saturation vapor pressure (kPa)
- ea = actual vapor pressure (kPa)
- (es − ea) = saturation vapor pressure deficit (kPa)
- Δ = slope of the saturation vapor pressure curve (kPa °C^{-1})
- γ = psychrometric constant (kPa °C^{-1})

3.4.2 Details of category-based empirical models

Temperature-Based Empirical Models

Temperature-based empirical models estimate ET_0 using minimum temperature, maximum temperature, mean temperature, and extraterrestrial radiation as their primary inputs. These models are applicable in data-scarce regions where only temperature records are reliably available and represent the simplest approach to ET_0 estimation (Mandal *et al.*, 2024). They rely on the diurnal temperature range as a surrogate for solar radiation and atmospheric dryness. Two temperature-based empirical models were evaluated in the present study.

i. Hargreaves–Samani Model

The Hargreaves–Samani model is one of the most widely used temperature-based empirical models for ET_0 estimation, validated across diverse climatic regions worldwide (Hargreaves and Samani, 1985). It estimates ET_0 from the diurnal temperature range and extraterrestrial radiation and is expressed as:

$$ET_0 = 0.408 \times 0.0023 \times R_{\mathrm{a}} \times (T_{\mathrm{mean}} + 17.8) \times (T_{\mathrm{max}} - T_{\mathrm{min}})^{0.5} \quad .....(2)$$

Where R_a is the extraterrestrial radiation (MJ m^{-2} day^{-1}), T_{mean} is the mean daily air temperature (°C), T_{max} is the maximum daily air temperature (°C), and T_{min} is the minimum daily air temperature (°C). The Hargreaves–Samani model has been widely recommended as the best performing temperature-based alternative to the FAO-56 PM method when humidity and wind speed data are unavailable (Allen *et al.*, 1998).

ii. Dorji Model

The Dorji model is a temperature-based empirical model developed as a simplified alternative for ET_0 estimation in regions with limited data availability (Dorji *et al.*, 2016). It estimates ET_0 using mean temperature and extraterrestrial radiation and is expressed as:

$$ET_0 = 0.408 \times 0.002 R_{\mathrm{a}} \times (T_{\mathrm{mean}} + 33.9) \times (T_{\mathrm{max}} - T_{\mathrm{min}})^{0.296} \quad .....(3)$$

Where all variables are as defined previously. The Dorji model provides a modified formulation of the temperature-range approach with adjusted empirical coefficients suited to different climatic conditions (Mandal *et al.*, 2024).

Mass Transfer-Based Empirical Models

Mass transfer-based empirical models estimate ET_0 based on the vapor pressure deficit between the evaporating surface and the overlying atmosphere, incorporating relative humidity and temperature as their primary inputs. These models are grounded in Dalton's law of evaporation, which states that the rate of evaporation is proportional to the vapor pressure deficit (Mohammadi *et al*., 2024). They are applicable in stations where humidity records are available but solar radiation measurements are absent. Two mass transfer-based empirical models were evaluated in the present study.

i. Dalton Model

The Dalton model is one of the earliest empirical approaches for estimating evaporation and evapotranspiration, based on the principle that the evaporation rate is proportional to the difference between saturation vapor pressure and actual vapor pressure (Dalton, 1802). The equation used in the present study is expressed as:

$$ET_0 = (0.3648 + 0.07223(u_2))(es - ea) \text{ .....(4)}$$

Where es is the saturation vapor pressure (hPa), ea is the actual vapor pressure (hPa), and u_2 is the wind speed at 2 m height (m s^{-1}).

ii. Meyer Model

The Meyer model is a mass transfer-based empirical model that estimates ET_0 from the saturation vapor pressure deficit incorporating a wind correction factor (Meyer, 1944). The equation used in the present study is expressed as:

$$ET_0 = (0.375 + 0.05026(u_2))(es - ea) \text{ .....(5)}$$

Where all variables are as defined previously. The Meyer model introduces modified empirical coefficients compared to the Dalton model and has been widely applied for ET_0 estimation in Indian agroclimatic conditions (Mandal *et al*., 2024).

Radiation-Based Empirical Models

Radiation-based empirical models estimate ET_0 primarily from solar radiation and temperature data, grounded in the energy balance principle wherein incoming solar radiation is the primary driver of evapotranspiration (Mandal *et al*., 2024). These models are particularly applicable in regions where sunshine duration data are available but humidity and wind speed records are absent or unreliable. Several studies have established that radiation-based models generally outperform temperature-based and mass transfer-based models in estimating ET_0 across diverse Indian meteorological stations (Mandal *et al.*, 2024). Two radiation-based empirical models were evaluated in the present study.

i. Priestley–Taylor Model

The Priestley–Taylor model is a simplified energy balance model that estimates potential evapotranspiration from net radiation and temperature, eliminating the aerodynamic term of the full Penman equation by introducing an empirical coefficient (Priestley and Taylor, 1972). It is expressed as:

$$ET_0 = 1.26 \times (\frac{\Delta}{\Delta} + \gamma) \times (R_n - G) \text{ .....(6)}$$

Where Δ is the slope of the saturation vapor pressure curve (kPa °C^{-1}), γ is the psychrometric constant (kPa °C^{-1}), R_n is the net radiation (MJ m^{-2} day^{-1}), and G is the soil heat flux density assumed zero for daily computations. The empirical coefficient 1.26 accounts for the advective component of evapotranspiration under well-watered conditions (Priestley and Taylor, 1972).

ii. Makkink Model

The Makkink model is a radiation-based empirical model that estimates ET_0 from solar radiation and temperature using the slope of the saturation vapor pressure curve and the psychrometric constant (Makkink, 1957). It is expressed as:

$$ET_0 = 0.61 \times (\frac{\Delta}{\Delta} + \gamma) \times (\frac{R_s}{\lambda}) - 0.12 \text{ .....(7)}$$

Where R_s is the solar radiation (MJ m^{-2} day^{-1}) and λ is the latent heat of vaporization (2.45 MJ kg^{-1}). The Makkink model has been evaluated and applied in various tropical and subtropical regions as a simple radiation-based alternative to the FAO-56 PM method (Mandal *et al.*, 2024). Furthermore, local calibration of all empirical models was conducted with the FAO-56 PM equation as the target. A simple linear regression model as recommended by Allen *et al.* (1998) was employed:

$$ET_0 - PM = a + b \times ET_0 - empirical \text{ .....(8)}$$

Where ET_o-PM represents the ET_o estimated using the FAO-56 Penman–Monteith method, ET_o-empirical denotes the ET_o estimated using the various empirical equations, and a and b are the regression coefficients for local calibration.

3.5 MACHINE LEARNING ALGORITHMS USED

In the present study, five machine learning algorithms were selected for ET_o estimation -Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Hybrid CNN-LSTM, Extreme Gradient Boosting (XGBoost), and Random Forest (RF). These five models represent three distinct architectural categories, namely deep learning algorithms (CNN and LSTM), hybrid deep learning algorithms (CNN-LSTM), and ensemble tree-based models (XGBoost and Random Forest), allowing a comprehensive evaluation of different machine learning paradigms under the same input and evaluation conditions. All models were implemented in Python using TensorFlow and Scikit-learn libraries in the Google Colab environment.

3.5.1 Ensemble Tree-Based algorithms

Ensemble tree-based models construct multiple decision trees during training and combine their outputs to produce a final prediction. These models are particularly well suited for regression tasks involving meteorological time series data owing to their robustness to overfitting, ability to handle nonlinear variable interactions, and capacity for feature importance assessment (Ayaz *et al.*, 2021).

i. Random Forest (RF)

Random Forest (RF) is an ensemble machine learning algorithm that constructs multiple decision trees using bootstrap sampling of the training data. Each tree is trained on a randomly selected subset of features, which increases model diversity and reduces overfitting. For regression problems, the final prediction is obtained by averaging the outputs of all decision trees (Ayaz *et al.*, 2021). Random Forest is widely used due to its robustness, ability to model nonlinear relationships, and capability to evaluate feature importance. Previous studies have reported reliable ETo prediction performance even with limited meteorological inputs (Wu *et al.*, 2022)

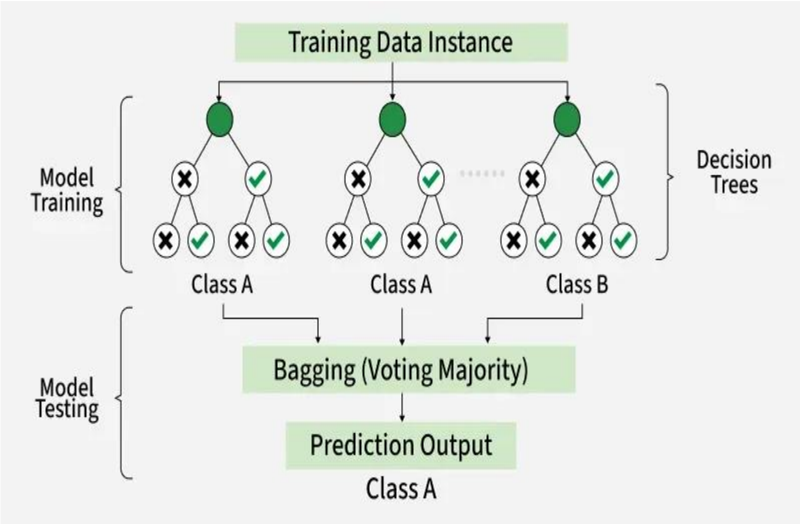

Fig 3.3 Architecture of Random Forest (RF) model

ii. Extreme Gradient Boosting (XGBoost)

Extreme Gradient Boosting (XGBoost) is an advanced ensemble machine learning algorithm based on the gradient boosting framework. It builds decision trees sequentially, with each new tree learning from the residual errors of the previous trees to improve prediction accuracy (Chang *et al*., 2025). The final prediction is obtained by combining the outputs of all trees in the ensemble. XGBoost incorporates regularization techniques to reduce overfitting and efficiently models complex nonlinear relationships. Due to its high predictive performance and feature importance analysis capability, it has been widely applied in hydrological and meteorological prediction studies. In the present study, XGBoost was implemented using the Python XGBoost library with hyper parameter optimization as described in Section 3.7.

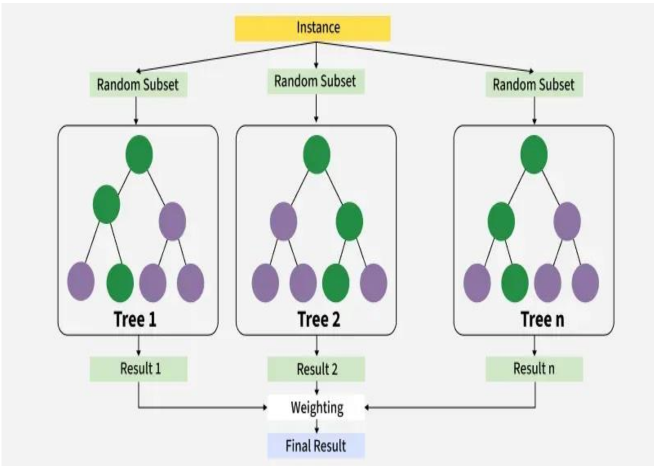

Fig 3.4 Architecture of XGBoost model

3.5.2 Deep Learning algorithms

Deep learning algorithms are neural network-based machine learning algorithms that learn hierarchial feature representations from data through multiple layers (Chen *et al*., wu2020). Their increasing use in regression analysis is supported by advances in computational power and the availability of large datasets. Neural networks can model complex nonlinear relationships between dependent and independent variables and effectively identify patterns in large datasets (Mandal *et al*., 2024). Therefore, they are well suited for estimating the nonlinear relationships between meteorological variables and ET_0; in this study, CNN and LSTM architectures were employed.

i. Convolutional Neural Network (CNN)

Convolutional Neural Network (CNN) is a deep learning architecture widely used for feature extraction and prediction tasks. In time-series regression, CNN applies convolutional filters to automatically learn temporal patterns from input data without manual feature engineering (Li *et al*., 2024). The architecture typically consists of Conv1D layers for feature extraction, MaxPooling layers for dimensionality reduction, and Dense layers for final prediction. CNN is effective in capturing nonlinear relationships and temporal dependencies in meteorological data, making it suitable for ET_0 estimation (Raja *et al*., 2024). In the present study, CNN was employed as both a standalone prediction model and the feature extraction component of the hybrid CNN–LSTM architecture.

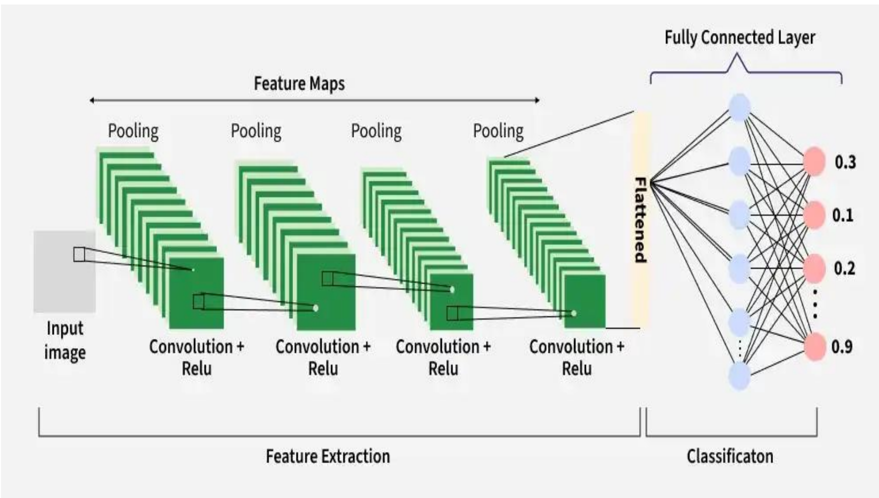

Fig 3.5 Architecture of CNN model

ii. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a specialized recurrent neural network designed to capture long-term dependencies in sequential data and overcome the vanishing gradient problem (Li *et al.*, 2024). LSTM employs memory cells and three gating mechanisms—input, forget, and output gates—to regulate the flow of information across time steps. This architecture enables the model to retain relevant temporal patterns while discarding less important information, making it effective for meteorological time-series prediction. Owing to its ability to model sequential and temporally correlated data, LSTM has been widely applied for ET_0 estimation (Sharma *et al.*, 2022). In the present study, LSTM was used both as a standalone deep learning model and as the sequential learning component of the hybrid CNN–LSTM architecture.

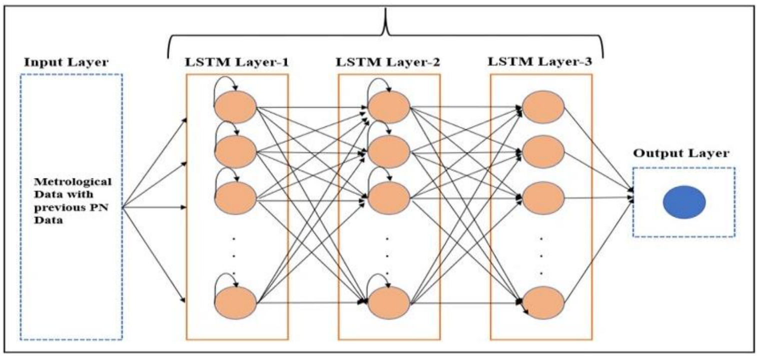

Fig 3.6 Architecture of LSTM mode

iii. Hybrid Deep Learning algorithms CNN-LSTM

Hybrid deep learning algorithms combine the architectural strengths of two or more individual neural network types to achieve superior performance compared to any single model. (Dong *et al*., 2024).In the context of ET_0 estimation from meteorological time series, hybrid architectures are particularly advantageous as they can simultaneously capture both local temporal features and long-term sequential dependencies in the input data (Debnath *et al*., 2015).

Hybrid CNN–LSTM is a deep learning architecture that combines the feature extraction capability of CNN with the sequential learning capability of LSTM (Sharma *et al*., 2022). In this framework, CNN layers extract important temporal features from meteorological input data, while LSTM layers learn long-term dependencies from the extracted feature sequences. The final Dense layer maps the learned representations to the predicted ET_0 value. By integrating the strengths of both architectures, CNN–LSTM provides improved prediction performance for complex time-series data and has been widely applied in hydrological and meteorological forecasting studies (Hadria *et al*., 2021). In the present study, the CNN–LSTM model was implemented using TensorFlow/Keras and evaluated for ET_0 prediction under different input categories and stations.

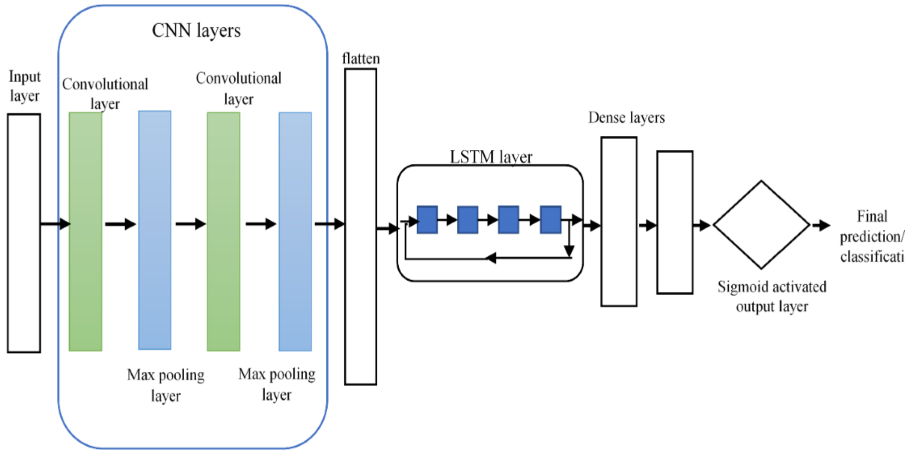

Fig 3.7 Architecture of Hybrid CNN-LSTM model

3.6 MACHINE LEARNING MODEL DEVELOPMENT STRATEGY

The machine learning model development in the present study was carried out through a systematic and structured procedure encompassing data normalization, dataset partitioning, model architecture configuration, hyperparameter,optimisation, and performance evaluation. The present section elaborates on the specific technical configurations and implementation details adopted for model development. The data for the period from 01-01-1995 to 31-12-2016 was used for training and data for the period 01-01-2017 to 31-12-2024 was used for testing, whereas the data for 2025 was used for validation of 2025 prediction. Details of training and testing datasets for both stations are shown in Table 3.2.

Table 3.2 Details of training and testing datasets used for Pattambi and Vellayani.

	Pattambi	Vellayani
Total period of data	01-01-1995 to 31-12-2025	01-01-1995 to 31-12-2025

Total length of data in days (including missing variables)	11322	11323
Training Period	01-01-1995 to 31-12-2016	01-01-1995 to 31-12-2016
Length of training period in days (% of N)	6939 (61.3%)	6940 (61.3%)
Testing Period	01-01-2017 to 31-12-2024	01-01-2017 to 31-12-2024
Length of testing period in days (% of N)	4383 (38.7%)	4383 (38.7%)
Prediction Period	01-01-2025 to 31-12-2025	01-01-2025 to 31-12-2025

3.6.1 Selection of Input Variables for Different Model Categories

The specific input variables employed for the ML models within each category are detailed in Table 3.3.

Table 3.3 Details of selected input variables for different model categories

Model Category	T_{min}	T_{min}	T_{mean}	R_a	RH	R_s
Full Inputs	✓	✓	✓	✓	✓	✓
Temperature-Based	✓	✓	✓	✓	–	–
Mass Transfer-Based	✓	✓	✓	✓	✓	–
Radiation-Based	✓	✓	✓	✓	–	✓

where R_a is the extraterrestrial radiation, RH is the relative humidity and R_s is the solar radiation.

3.6.2 Data Normalization

Before training the machine learning models, all input variables and the target variable (ET_o) were normalized to ensure a uniform scale. This preprocessing step minimizes the influence of variables with larger numerical ranges and enhances the stability and convergence of the learning algorithms (Li *et al.*, 2024). In this study, Min–Max normalization was employed to constrain all variables to a range of 0 to 1 while preserving the relationships among the original data values. The normalization method implemented in this study is defined by the following equation:

$$x_{norm} = \frac{(x_o - x_{min})}{(x_{max} - x_{min})}$$

where x_{norm}, x_o, x_{min}, x_{max} represent the normalized value, the actual value, the maximum observed value, and the minimum observed value, respectively.

3.6.3 Training and Testing Split Configuration

For the deep learning models -CNN, LSTM, and Hybrid CNN-LSTM -a three-way sequential split of 70 per cent training, 15 per cent validation, and 15 per cent testing was adopted. For the 30-year dataset from 1995 to 2024 comprising approximately 10,957 daily records, this corresponds to approximately 7,670 records for training, 1,644 records for validation, and 1,644 records for testing. The validation set was used for early stopping and hyperparameter tuning, while the test set was used exclusively for final performance evaluation.

For the XGBoost model, a two-way sequential split of 80 per cent training and 20 per cent testing was adopted, corresponding to approximately 8,766 records for training and 2,191 records for testing. For the Random Forest model, a year-based chronological split was adopted wherein records up to and including the year 2020 constituted the training set, records from 2021 to 2022 constituted the validation set, and records from 2023 onwards constituted the test set.

3.6.4 Software Libraries and Implementation Platform

i) Primary Programming Language and Cloud Computing Platform

All machine learning models were implemented using Python, a widely adopted open-source programming language preferred for its rich ML/DL library support, rapid prototyping capability, and extensive data handling features. The implementation was carried out on Google Colab, a cloud-based platform providing GPU-accelerated computing resources, eliminating the need for local hardware setup and enabling efficient training of deep learning models through its collaborative notebook environment.

ii) Libraries and Frameworks

TensorFlow and Keras were used for developing, training, and evaluating the CNN, LSTM, and CNN–LSTM models, with TensorFlow managing tensor computations and Keras providing the high-level API for neural network design. Scikit-learn was employed for Random Forest implementation, data preprocessing, and performance metric computation, while XGBoost was used for the Extreme Gradient Boosting model, offering a scalable and optimized gradient boosting suited for structured meteorological data. NumPy and Pandas facilitated numerical computation and time series data handling, whereas Matplotlib and Seaborn were used for visualization of predictions, scatter plots, and performance metric plots.

3.6.5 Model Architecture and Hyperparameter Configuration

Several hyperparameters including the number of layers, filters, neurons, learning rate, batch size, and dropout rate were varied and the configuration yielding the lowest validation RMSE was selected as the optimal model for each case. Among the deep learning models, Hybrid CNN-LSTM adopted the most complex architecture combining convolutional and recurrent layers, followed by CNN and LSTM with progressively simpler sequential structures with Dropout regularization. The ensemble tree-based models XGBoost and Random Forest were configured with an optimized number of estimators and controlled tree depth to ensure adequate learning capacity while preventing overfitting.

3.7 PERFORMANCE EVALUATION METRICS

The predictive accuracy of both the empirical models and the machine learning models developed in the present study was evaluated using three widely accepted statistical performance metrics the Coefficient of Determination (R^2), the Root Mean Square Error (RMSE), and the Mean Absolute Error (MAE). These metrics were computed by comparing the model-predicted reference evapotranspiration (ET_o) values against the corresponding FAO-56 Penman-Monteith (FAO-56 PM) ET_o values, which served as the standard reference in this study.

The Coefficient of Determination (R^2) is one of the most widely used dimensionless statistical metrics for evaluating the performance of predictive models. It quantifies the proportion of the total variance in the observed ET_o values that is explained by the model predictions. R^2 ranges from 0 to 1, where a value of 1 denotes a perfect agreement between predicted and observed values with no residual variance, while a value of 0 indicates that the model has no explanatory power over the observed data. Values closer to 1 are indicative of superior model performance.

The Root Mean Square Error (RMSE) is a widely used error metric that quantifies the average magnitude of the discrepancies between the model-predicted and observed ET_o values. RMSE is expressed in the same units as the variable of interest (mm day^{-1}). Due to the squaring of individual errors prior to averaging, RMSE assigns disproportionately greater weight to large errors compared to smaller ones, making it particularly sensitive to outliers and extreme prediction errors. Consequently, a model with consistently small errors will exhibit a lower RMSE, and a lower RMSE value is generally indicative of better model performance.

The Mean Absolute Error (MAE) measures the average magnitude of the absolute differences between model-predicted and observed ET_o values, without imposing a disproportionate penalty on larger errors. Unlike RMSE, MAE treats all individual errors equally regardless of their magnitude, which makes it a more robust and interpretable measure of average model accuracy, particularly in datasets where outliers may be present. A lower MAE value indicates better predictive accuracy. MAE is expressed as:

$$R^2 = \frac{1 - \Sigma i\,(Oi - Pi)2}{\Sigma i\,(Oi - \text{ŏ})2}$$

$$\text{RMSE} = \frac{\sqrt{[}\,\Sigma i\,(Oi - Pi)2}{N}$$

$$\text{MAE} = \frac{\Sigma i\,|\,Oi - Pi\,|}{N}$$

where O_i is the observed FAO-56 PM ET_o value on the th day (mm day^{-1}).

P_i is the model-predicted ET_o on the th day (mm day^{-1}).

ŏ is the mean of all observed ET_o values (mm day^{-1}).

N is the total number of observations in the testing dataset.

3.8 SPATIAL TRANSFERABILITY ASSESSMENT

One of the key objectives of the present study was to assess the spatial transferability of the developed machine learning models across two geographically distinct and climatically contrasting stations in Kerala. The transferability analysis determines whether a machine learning model trained on meteorological data from one station can be effectively applied to predict ET_o at a different station without any retraining. This is particularly important for data-scarce regions of Kerala where complete meteorological records may not be available at all locations, and reliable ET_0 estimation is still required for irrigation planning and water resource management.

Following the methodology proposed by Cao *et al*. (2024) and adopted by Mandal *et al*. (2024), the models developed and trained specifically on the RARS Pattambi dataset were evaluated in two ways -first on the Pattambi test dataset itself, referred to as internal performance, and second on the Vellayani test dataset without any retraining, referred to as external performance. The spatial transferability of each model was quantified using the transferability score T, defined as the ratio of external performance to internal performance:

$$T = \frac{External\ R^2}{Internal\ R^2} \quad \text{.....(Eq. 15)}$$

A higher value of T closer to 1.0 indicates superior spatial transferability of the model across locations. A score significantly lower than 1.0 indicates that the model loses considerable predictive accuracy when applied to the new location. The transferability was assessed separately for each of the three input categories—full input-based, mass transfer-based, and radiation-based-to identify which combination of model and input category generalises best across the two contrasting agroclimatic zones of Kerala.Temperature-based models were excluded from the analysis due to their poor performance under Kerala conditions, as indicated by their low R^2 values.

3.9 ET_0 PREDICTION FOR 2025

Following the completion of model training and spatial transferability on the 1995–2024 dataset, the finalised machine learning models were applied to predict daily reference evapotranspiration for the year 2025 at both RARS Pattambi and Vellayani stations. The actual meteorological data recorded for the year 2025 were fed as inputs to the developed models and the predicted ET_0 values were compared against the empirically estimated FAO-56 Penman–Monteith for the same 2025 meteorological records. The performance of the models in predicting ET_0 for 2025 was evaluated using R^2, RMSE, and MAE. The predicted ET_0 values by machine learning models for 2025 were checked for agreement with the FAO-56 PM computed values, at both stations with scatter plots, demonstrating the practical forecasting capability of the developed models beyond the training period.

RESULTS AND DISCUSSION

CHAPTER IV
RESULTS AND DISCUSSION

The results and discussion of the project work entitled "Evapotranspiration Estimation Using Empirical and Machine Learning Models: A Comparative Assessment of Accuracy and Transferability" are presented under the following four subheads. The first section compares the performance of machine learning (ML) and empirical models using the full set of meteorological variables as inputs. The second section evaluates model performance using reduced-input-variable categories, viz., temperature-based, mass-transfer-based, and radiation-based datasets. The third section examines the spatial transferability of machine learning models developed at Pattambi and Vellayani. The final section presents ET_o predictions for the year 2025 using the developed machine learning models.

4.1 MACHINE LEARNING MODEL ARCHITECTURE AND ITS HYPERPARAMETER CONFIGURATION

The final machine learning model architectures, Python libraries used and hyperparameter configurations tuned are presented in Table 4.1. The selected configurations provided the best validation performance and may be used for subsequent model training, evaluation and prediction.

Table 4.1 Details of Python libraries, architectures and hyperparameters of ML models

Convolutional Neural Network (CNN)	Python Library	TensorFlow / Keras (tf.random.set seed(42))
	Scaler	StandardScaler (fit on training set only - no leakage)
	Architecture	Conv1D(64, k=3, ReLU) → Conv1D(128, k=3, ReLU) → MaxPooling1D(2) → Conv1D(64, k=2, ReLU) → Flatten → Dense(128, ReLU) → Dropout(0.2) → Dense(64, ReLU) → Dropout(0.1) → Dense(1)
	Input shape	(TIME STEPS=7, n features=8)
	Output layer	Dense(1); Activation: linear (regression output)
	Optimizer	Adam (Learning rate = 0.001)
	Loss function	Mean Squared Error (MSE)
	Training setup	Epochs: 100; Batch size: 64
	Early stopping	monitor = val loss; patience = 15; restore best weights = True
	Sequence window	TIME STEPS = 7 (sliding window of 7 days)

Long Short-Term Memory (LSTM)	Python Library	TensorFlow / Keras (tf.random.set seed(42))
	Scaler	StandardScaler (fit on training set only - no leakage)
	Architecture	LSTM(128, return sequences=True) → Dropout(0.2) → LSTM(64, return sequences=False) → Dropout(0.2) → Dense(64, ReLU) → Dropout(0.1) → Dense(32, ReLU) → Dense(1)
	Input shape	(TIME STEPS=7, n features=8)
	Output layer	Dense(1); Activation: linear (regression output)
	Optimizer	Adam (Learning rate = 0.001)
	Loss function	Mean Squared Error (MSE)
	Training setup	Epochs: 150; Batch size: 64
	Early stopping	monitor = val loss; patience = 15; restore best weights = True
	Sequence window	TIME STEPS = 7 (sliding window of 7 days)
Hybrid CNN-LSTM	Python Library	TensorFlow / Keras (tf.random.set seed(42))
	Scaler	StandardScaler (fit on training set only - no leakage)
	Architecture	Reshape → TimeDistributed(Conv1D(64, k=2, ReLU)) → TimeDistributed(Conv1D(32, k=2, ReLU)) → TimeDistributed(MaxPooling1D(2)) → TimeDistributed(Flatten) → LSTM(64) → Dropout(0.2) → Dense(64, ReLU) → Dropout(0.1) → Dense(32, ReLU) → Dense(1)
	Sequence config	N SUBSEQ = 2 sub-sequences; SUB LEN = 4; TIME STEPS = N SUBSEQ × SUB LEN = 8
	Input shape	(TIME STEPS=8, n features=8) reshaped to (2 sub-sequences, 4 timesteps, 8 features)
	Output layer	Dense(1); Activation: linear (regression output)
	Optimizer	Adam (Learning rate = 0.0001)
	Loss function	Mean Squared Error (MSE)
	Training setup	Epochs: 200; Batch size: 32
	Early stopping	monitor = val loss; patience = 15; restore best weights = True
Extreme Gradient Boosting (XGBoost)	Python Library	xgboost (XGBRegressor)
	Scaler	MinMaxScaler (fit on training set only)
	Model configuration	n estimators = 800; learning rate = 0.03; max depth = 5; min child weight = 2; subsample = 0.85; colsample bytree = 0.85
	Objective function	reg:squarederror (Mean Squared Error regression)
	Random state	random state = 42
Random Forest (RF)	Python Library	Scikit-learn (RandomForestRegressor)
	Scaler	No scaling applied (RF is scale-invariant)
	Model configuration	n estimators = 500; max depth = 12; min samples split = 5; min samples leaf = 2; max features = 'sqrt'

	Parallel processing	n jobs = -1 (all available CPU cores)
	Random state	random state = 42

4.2 PERFORMANCE OF FULL INPUT BASED MACHINE LEARNING MODELS

The five machine learning models were trained and evaluated on the full set of input variables, comprising T_{max}, T_{min}, T_{mean}, RH, wind speed, R_a, and R_s. This phase established the upper performance benchmark for all subsequent reduced-input comparisons. The performance metrics for all five ML models at Pattambi and Vellayani are presented in Table 4.2 and the corresponding scatter plots comparing FAO-56 Penman Monteith vs estimated ET_0 by ML models are shown in Fig. 4.1. The empirically estimated values by FAO-56 Penman-Monteith were considered as the target variable for machine learning models as well as the observed value for comparison.

Table 4.2 Performance of ML models developed based on full input variables

Station	Model	R^2	RMSE (mm/day)	MAE (mm/day)
Pattambi	CNN	0.959	0.219	0.164
	LSTM	0.960	0.216	0.159
	CNN-LSTM	0.964	0.203	0.162
	XGBoost	0.982	0.143	0.085
	Random Forest	0.983	0.137	0.089
Vellayani	CNN	0.957	0.200	0.152
	LSTM	0.964	0.184	0.139
	CNN-LSTM	0.965	0.179	0.140
	XGBoost	0.996	0.059	0.044
	Random Forest	0.987	0.114	0.080

The results indicated that XGBoost achieved the best performance at both stations, with the highest R^2 and the lowest prediction errors. At Pattambi, XGBoost recorded an R^2 of 0.982, RMSE of 0.143 mm/day, and MAE of 0.085 mm/day, while at Vellayani, it achieved an R^2 of 0.996, RMSE of 0.059 mm/day, and MAE of 0.044 mm/day. Random Forest ranked second at both locations, whereas among the deep learning models, CNN–LSTM hybrid model consistently outperformed LSTM and CNN. Overall, the performance ranking was of the order: XGBoost > Random Forest > CNN–LSTM > LSTM > CNN, with all models achieved $R^2 > 0.95$, indicating excellent predictive capability under full-input conditions.

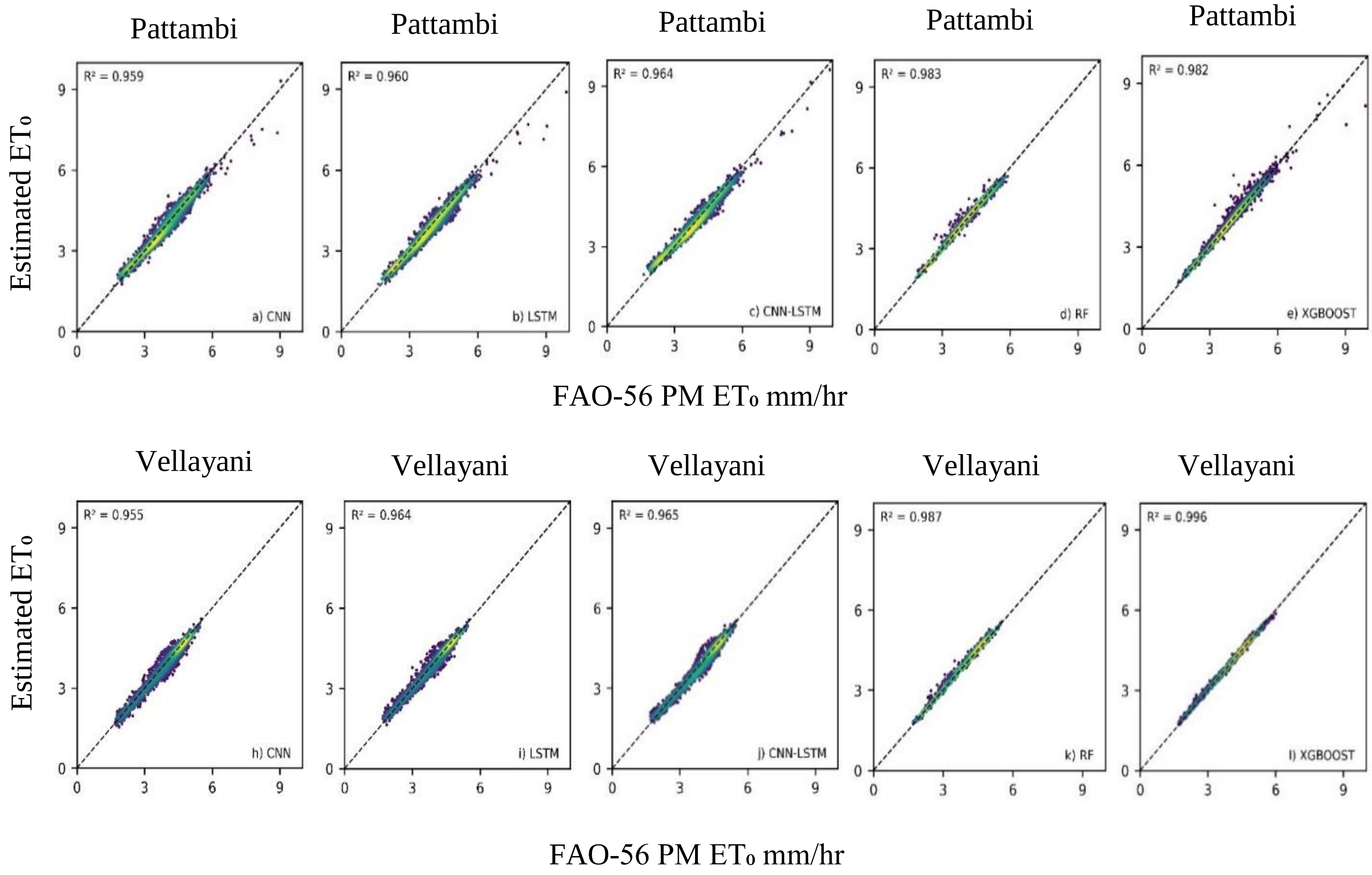

Fig. 4.1 Scatter plots of estimated ET_0 by ML models vs estimated ET_0 by FAO-56 PM for full input-based models.

4.3 PERFORMANCE OF TEMPERATURE-BASED MODELS

Two temperature-based empirical models, Hargreaves–Samani (HS) and Dorji (DO) and five machine learning models were evaluated and compared against FAO-56 PM ET_0 at the two stations. The results are summarized in Table 4.3 and corresponding scatter plots are presented in Fig. 4.2.

Table 4.3 Performance of temperature-based empirical and machine learning models

Station	Model Category	Model	R^2	RMSE (mm/day)	MAE (mm/day)
Pattambi	Empirical	Hargreaves–Samani (HS)	0.471	0.825	0.624
	Empirical	Dorji (DO)	0.439	0.850	0.652
	Machine Learning	CNN	0.557	0.728	0.550

	Machine Learning	LSTM	0.574	0.726	0.548
	Machine Learning	CNN-LSTM	0.559	0.726	0.547
	Machine Learning	XGBoost	0.557	0.732	0.547
	Machine Learning	Random Forest	0.544	0.736	0.552
Vellayani	Empirical	Hargreaves–Samani	0.255	0.693	0.535
	Empirical	Dorji	0.321	0.662	0.506
	Machine Learning	CNN	0.421	0.613	0.486
	Machine Learning	LSTM	0.456	0.593	0.448
	Machine Learning	CNN-LSTM	0.458	0.595	0.450
	Machine Learning	XGBoost	0.482	0.581	0.430
	Machine Learning	Random Forest	0.491	0.589	0.438

The two empirical models showed relatively low performance accuracy at both stations, with R^2 ranging from 0.255 to 0.471. At Pattambi, the Hargreaves–Samani model (R^2 = 0.471, RMSE = 0.825 mm/day) slightly outperformed the Dorji model (R^2 = 0.439, RMSE = 0.850 mm/day). But at Vellayani, the trend reversed, with the Dorji model (R^2 = 0.321, RMSE = 0.662 mm/day) performing better than Hargreaves–Samani (R^2 = 0.255, RMSE = 0.693 mm/day).

The temperature-based machine learning models consistently outperformed the empirical models at both stations. At Pattambi, LSTM achieved the highest R^2 (0.574) with RMSE = 0.726 mm/day and MAE = 0.548 mm/day, followed closely by CNN-LSTM and

CNN. At Vellayani, Random Forest gave the best performance ($R^2 = 0.491$, RMSE = 0.589 mm/day, MAE = 0.438 mm/day), with XGBoost showing comparable results ($R^2 = 0.482$, RMSE = 0.581 mm/day). Across both stations, machine learning models delivered consistently higher R^2 and lower error values than the empirical models, confirming their superior ability to model ET_0 from temperature-based inputs alone. The performance gap was widest at Vellayani, suggesting that data-driven approaches offer the greatest advantage where simple empirical relationships are least reliable.

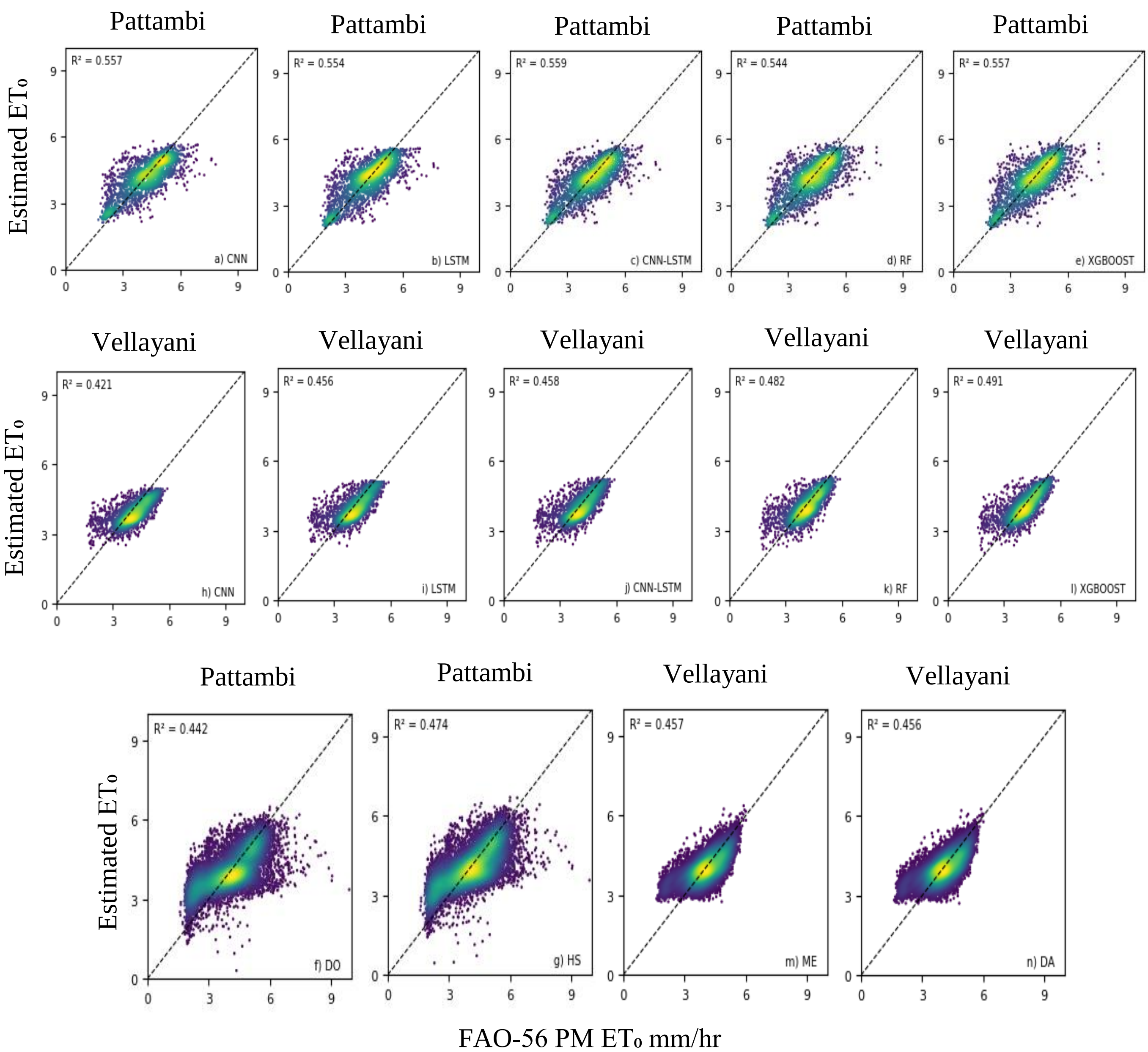

Fig. 4.2 Scatter plots of estimated ET_0 by ML models vs estimated ET_0 by FAO-56 PM for temp-based models

4.4 PERFORMANCE OF MASS TRANSFER-BASED MODELS

Mass transfer-based two empirical models and five machine learning models were evaluated and compared against FAO-56 PM ET_0 at the two stations. The performance machine learning models developed are summarised in Table 4.4 and the corresponding scatter plots are presented in Fig. 4.3.

Table 4.4 Performance of mass transfer-based empirical and machine learning models

Station	Model Category	Model	R^2	RMSE (mm/day)	MAE (mm/day)
Pattambi	Empirical	Dalton (DA)	0.455	0.594	0.467
	Empirical	Meyer (ME)	0.457	0.592	0.467
	Machine Learning	CNN	0.759	0.546	0.438
	Machine Learning	LSTM	0.756	0.537	0.406
	Machine Learning	CNN-LSTM	0.760	0.533	0.403
	Machine Learning	XGBoost	0.758	0.532	0.391
	Machine Learning	Random Forest	0.757	0.531	0.395
Vellayani	Empirical	Dalton	0.597	0.720	0.582
	Empirical	Meyer	0.598	0.720	0.584
	Machine Learning	CNN	0.613	0.487	0.374
	Machine Learning	LSTM	0.554	0.486	0.364
	Machine Learning	CNN-LSTM	0.622	0.484	0.363
	Machine Learning	XGBoost	0.654	0.465	0.348
	Machine Learning	Random Forest	0.657	0.470	0.354

Both Dalton (DA) and Meyer (ME) empirical models showed nearly identical performance at both stations, with Meyer marginally outperforming Dalton at Pattambi (R^2 = 0.457 vs 0.455). The mass transfer-based empirical models performed better at Vellayani ($R^2 \approx 0.60$) than at Pattambi ($R^2 \approx 0.46$), likely due to the higher and more consistent relative humidity at the humid coastal region at Vellayani station, and both models substantially outperformed the temperature-based empirical models, indicating that incorporating relative humidity improves ET_0 estimation accuracy Sharma *et al.* (2022). All ML models showed improved performance compared to empirical models. At Pattambi, CNN-LSTM achieved the highest R^2 (0.760), while Random Forest showed the lowest RMSE (0.531 mm/day) and XGBoost the lowest MAE (0.391 mm/day). At Vellayani, Random Forest model performed best with R^2 = 0.657, RMSE = 0.465 mm/day and MAE = 0.348 mm/day. Overall, the inclusion of the variable relative humidity significantly improved ET_0 estimation across both model categories (Chen *et al.*, 2026) .

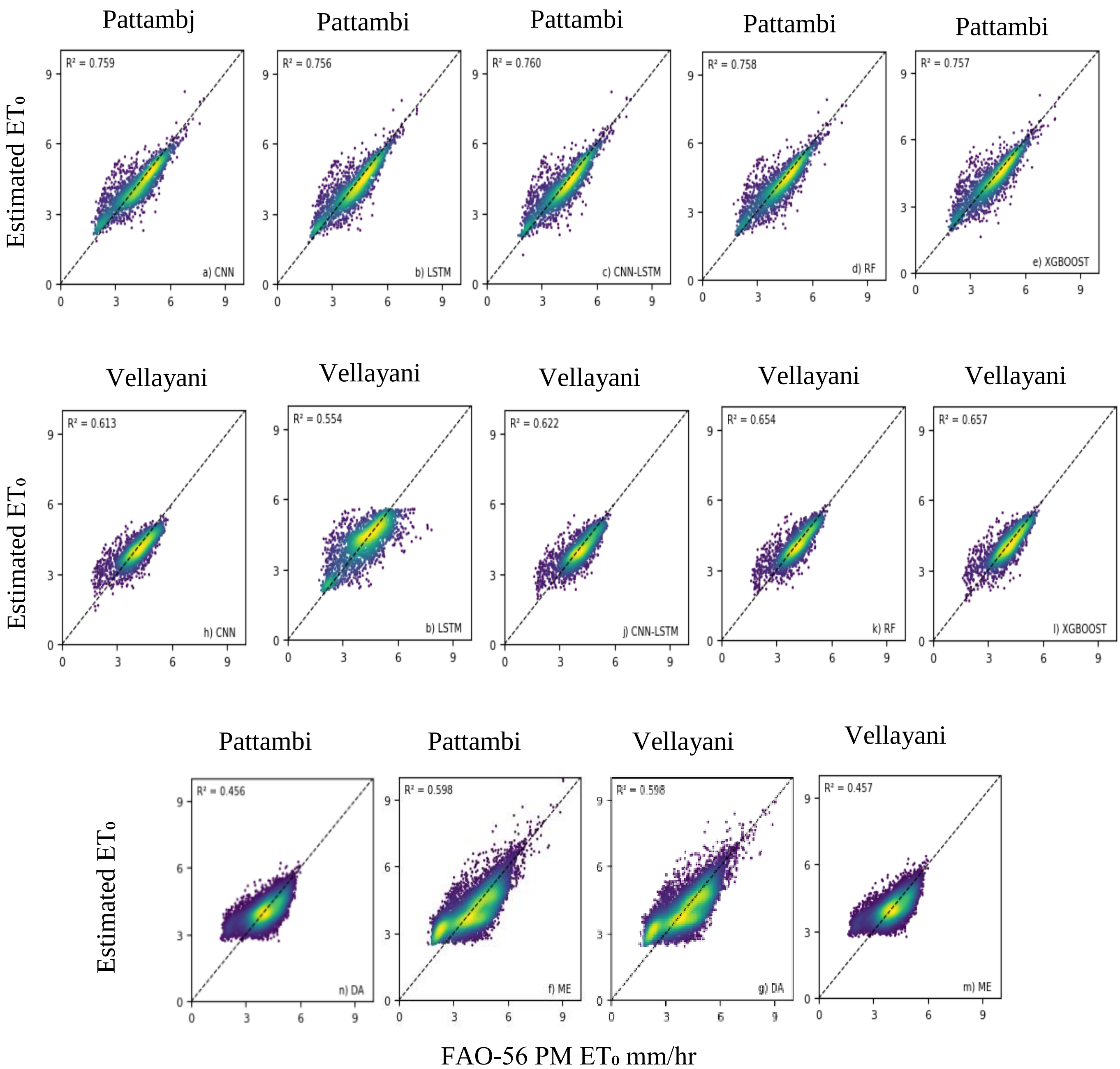

Fig. 4.3 Scatter plots of estimated ET_0 by ML models vs estimated ET_0 by FAO-56 PM for mass transfer based models

4.5 PERFORMANCE OF RADIATION-BASED MODELS

Priestley–Taylor (PT) and Makkink (MA) empirical models and five ML models were evaluated and tested at both stations and their performance metrics are presented in Table 4.5 and scatter plots are shown in Fig.4.4

Table 4.5 Performance of radiation-based empirical and machine learning models

Station	Model Category	Model	R^2	RMSE (mm/day)	MAE (mm/day)
Pattambi	Empirical	Priestley–Taylor	0.725	0.595	0.423
	Empirical	Makkink	0.841	0.453	0.306
	Machine Learning	CNN	0.870	0.436	0.284
	Machine Learning	LSTM	0.878	0.411	0.133
	Machine Learning	CNN-LSTM	0.875	0.414	0.255
	Machine Learning	XGBoost	0.875	0.415	0.242
	Machine Learning	Random Forest	0.557	0.403	0.239
Vellayani	Empirical	Priestley–Taylor (PT)	0.938	0.200	0.152
	Empirical	Makkink (MA)	0.947	0.185	0.143
	Machine Learning	CNN	0.960	0.152	0.114
	Machine Learning	LSTM	0.972	0.133	0.101
	Machine Learning	CNN-LSTM	0.972	0.133	0.101
	Machine Learning	XGBoost	0.971	0.130	0.097

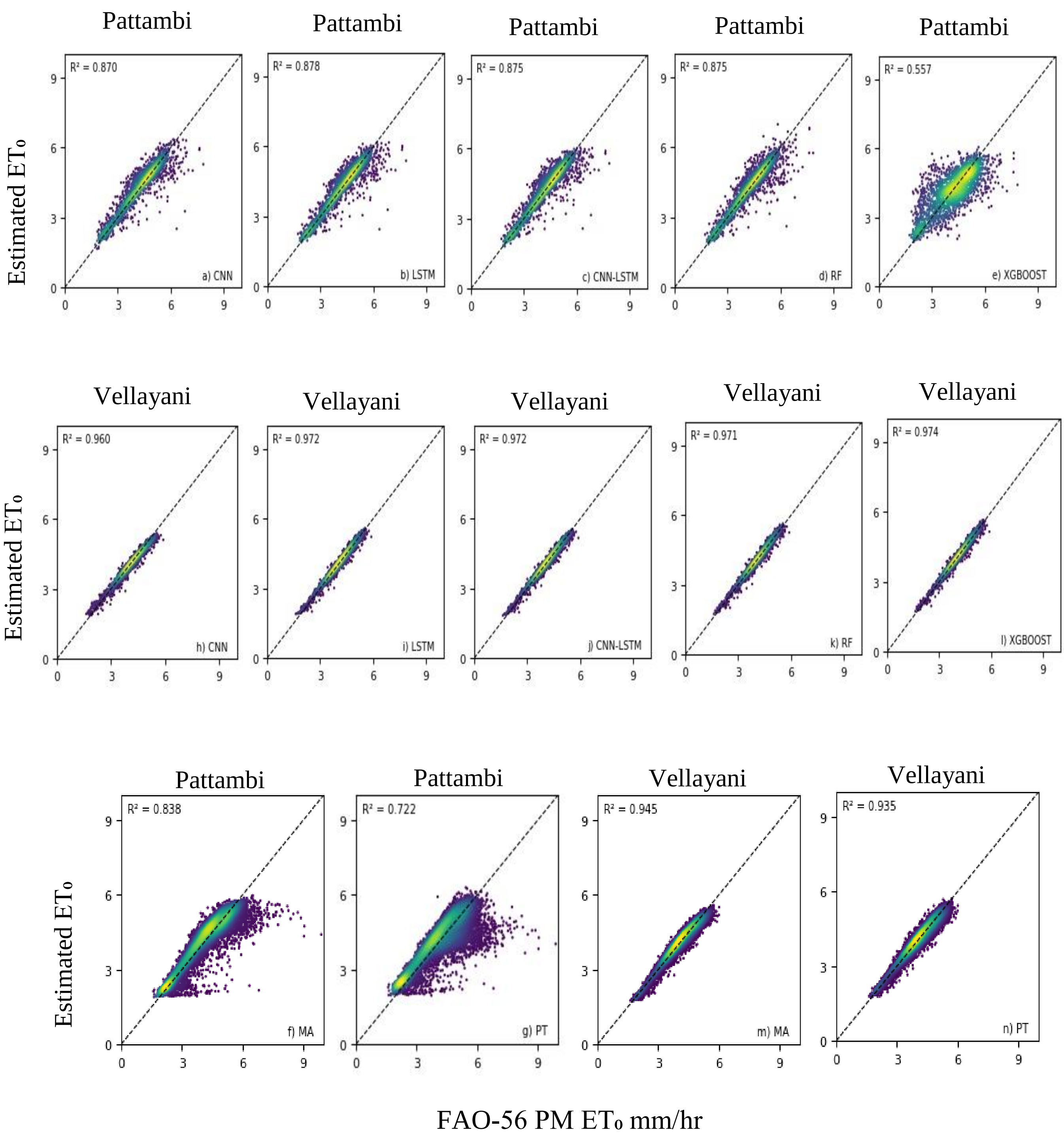

Fig. 4.4 Scatter plots of estimated ET_0 by ML models vs estimated ET_0 by FAO-56 PM for radiation based models

Radiation-based empirical models clearly outperformed temperature- and mass-transfer-based models at both stations. At Pattambi, Makkink (R^2 = 0.841) considerably outperformed Priestley–Taylor (R^2 = 0.725), while at Vellayani both models performed excellently (Makkink R^2 = 0.947, Priestley–Taylor R^2 = 0.938) - consistent with the strong solar radiation– ET_0 relationship in Thiruvananthapuram's humid coastal climate. These results confirm that radiation-based empirical models are the most accurate of the three empirical categories.

Under radiation-based inputs, all ML models showed improved performance compared with the temperature- and mass-transfer-based categories. At Pattambi, LSTM led (R^2 = 0.878), followed closely by CNN-LSTM and Random Forest (R^2 = 0.875 each), while XGBoost lagged notably (R^2 = 0.557). At Vellayani, XGBoost topped the group (R^2 = 0.974), with LSTM and CNN-LSTM close behind (R^2 = 0.972 each) and Random Forest slightly lower (R^2 = 0.971). Overall, radiation-based inputs improved ML model performance compared to the other two input categories.

4.6 COMPARATIVE SUMMARY OF PERFORMANCE OF EMPIRICAL AND MACHINE LEARNING MODELS

The overall comparison of model performance across the four input variable categories indicated that the inclusion of more relevant meteorological variables significantly improved ET_0 estimation accuracy. Among the four categories, the Full Input Variable-based and Radiation-based models exhibited the best predictive performance at both Pattambi and Vellayani stations. In the case of full Input variable category, Random Forest achieved the highest prediction accuracy with an R^2 of 0.983 at Pattambi and 0.996 at Vellayani, while XGBoost also demonstrated excellent performance (R^2 = 0.983 and 0.987, respectively). In the Radiation-based category, LSTM achieved the highest R^2 (0.878) at Pattambi, whereas XGBoost produced the best performance at Vellayani (R^2 = 0.974, RMSE = 0.130 mm day^{-1} and MAE = 0.097 mm day^{-1}). The Mass Transfer-based models showed moderate prediction accuracy, while the Temperature-based models consistently recorded the lowest R^2 values and comparatively higher RMSE and MAE. These findings demonstrated that the availability of comprehensive meteorological inputs, particularly radiation-related variables, together with advanced machine learning algorithms, substantially enhances the accuracy of reference evapotranspiration estimation compared with empirical approaches. This was in conformity with the findings of (Li *et al.*, 2024).

Table 4.6 Summary of performance indicators of all empirical and ML models

Model Category	Model	Pattambi			Vellayani		
		R^2	RMSE (mm/day)	MAE (mm/day)	R^2	RMSE (mm/day)	MAE (mm/day)
Full input variable based	CNN	0.959	0.219	0.164	0.957	0.200	0.152
	LSTM	0.960	0.216	0.159	0.964	0.184	0.139
	CNN-LSTM	0.964	0.203	0.162	0.965	0.179	0.140
	XGBoost	0.982	0.143	0.085	0.996	0.059	0.044
	Random Forest	0.983	0.137	0.089	0.987	0.114	0.080
Temperature-based	Hargreaves-Samani	0.474	0.825	0.624	0.456	0.693	0.535
	Dorji	0.442	0.850	0.652	0.457	0.662	0.506
	CNN	0.557	0.728	0.550	0.421	0.613	0.486
	LSTM	0.554	0.726	0.548	0.456	0.593	0.448
	CNN-LSTM	0.559	0.726	0.547	0.458	0.595	0.450
	XGBoost	0.557	0.732	0.547	0.482	0.581	0.430
	Random Forest	0.544	0.736	0.552	0.491	0.589	0.438
Mass transfer-based	Dalton	0.456	0.594	0.467	0.598	0.720	0.582
	Meyer	0.598	0.592	0.467	0.457	0.720	0.584
	CNN	0.759	0.546	0.438	0.613	0.487	0.374
	LSTM	0.756	0.537	0.406	0.554	0.486	0.364
	CNN-LSTM	0.760	0.533	0.403	0.622	0.484	0.363
	XGBoost	0.758	0.532	0.391	0.654	0.465	0.348
	Random Forest	0.757	0.531	0.395	0.657	0.470	0.354
Radiation-based	Priestley-Taylor	0.72	0.595	0.423	0.935	0.200	0.152
	Makkink	0.838	0.453	0.306	0.945	0.185	0.143
	CNN	0.870	0.436	0.284	0.960	0.152	0.114
	LSTM	0.878	0.411	0.133	0.972	0.133	0.101
	CNN-LSTM	0.875	0.414	0.255	0.972	0.133	0.101
	XGBoost	0.875	0.415	0.242	0.974	0.130	0.097
	Random Forest	0.557	0.403	0.239	0.971	0.135	0.103

4.7 SPATIAL TRANSFERABILITY OF MACHINE LEARNING MODELS

The spatial transferability score (T) represents the fraction of internal accuracy retained upon transfer to the other station. To assess spatial transferability, each of the five machine learning models (CNN, LSTM, CNN-LSTM, RF, and XGBoost) developed at one station was tested forte same station (internal performance) and the other station (external performance). The transferability across different input categories viz. full input based, mass transfer-based, and radiation-based, was evaluated for both Pattambi and Vellayani. The temperature-based models were excluded due to poor baseline performance because of Kerala's humid tropical conditions (Sharma *et al.*, 2022).

4.7.1 Full input-based machine learning models

Generally, it was found that internal R^2 values were uniformly high (0.96–0.98), and external R^2 remained above 0.85 for all the machine learning models. When all input-based machine learning models trained at Pattambi and tested at Vellayani, showed that all five models retained a high proportion of their internal accuracy, with T values ranging from 0.89 (CNN) to 0.95 (XGBoost). In the case of ensemble tree-based models, RF and XGBoost, showed the highest transferability, suggesting that when all the full set of climatic variables are available, these models capture relationships that are less station-specific and more physically generalizable. Transferability and Comparison of internal and external model performance (R^2 is used as the performance metric) at Pattambi are shown in Fig 4.5

Transferability and Comparison of Internal and External Model Performance (R 2 is used as the performance metric) for Vellayani are shown in Fig 4.6. It was found that transferability was comparably strong, with the deep learning models (CNN, LSTM, CNN-LSTM) all recorded an identical T value of 0.94, while XGBoost achieved 0.91 and RF recorded the lowest value of 0.85. The transferability of ML models from Pattambi to Vellayani was found comparatively poor than Vellayani to Pattambi. This indicated that RF model’s tree-based partitioning may be more sensitive to the specific range and distribution of the training data, making it more susceptible to performance loss when the target station's climatic regime falls outside the value ranges (Chang *et al.*, 2025).

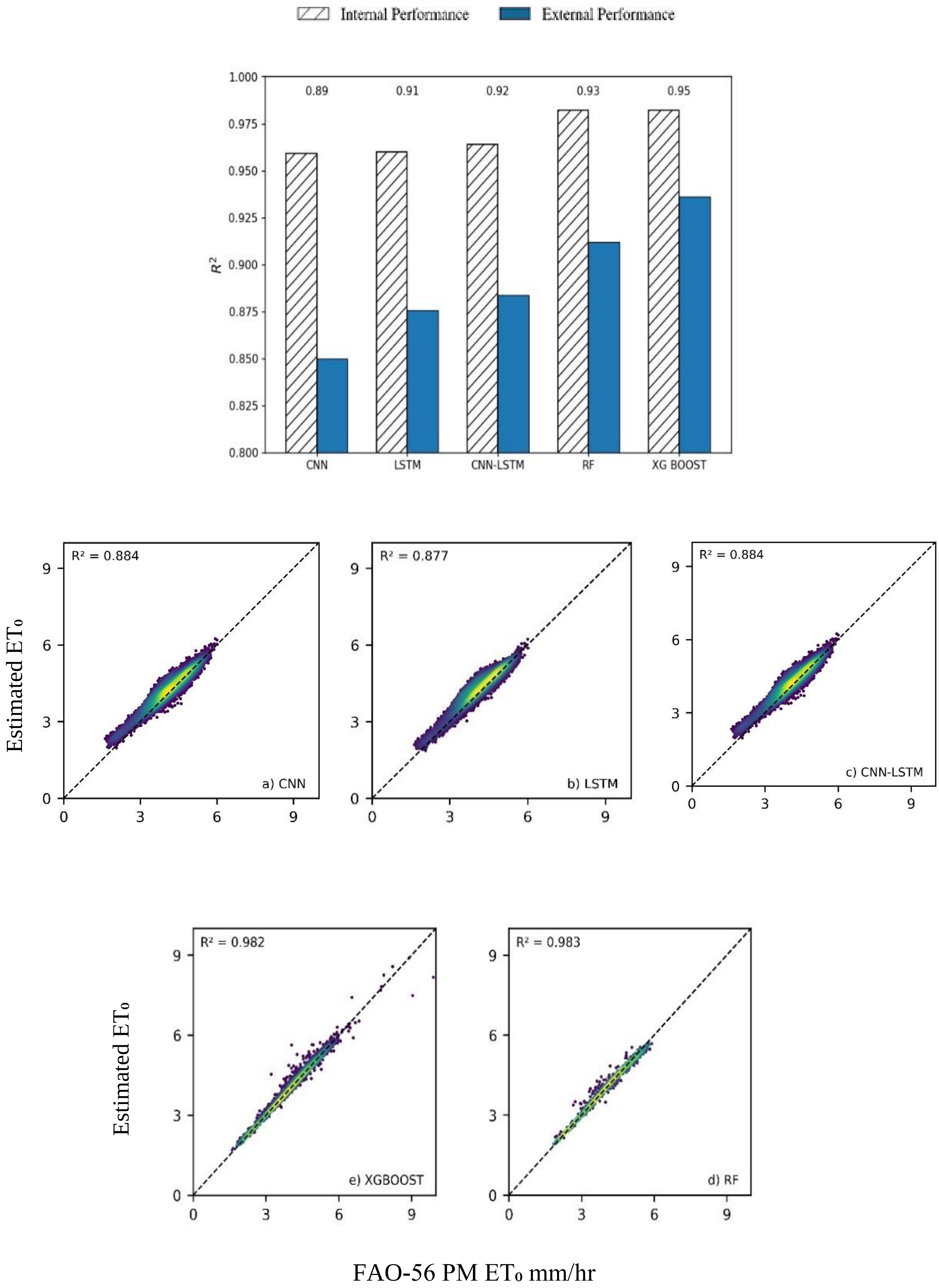

Fig 4.5 Transferability and Comparison of Internal and External Model Performance of Pattambi for full input based variables

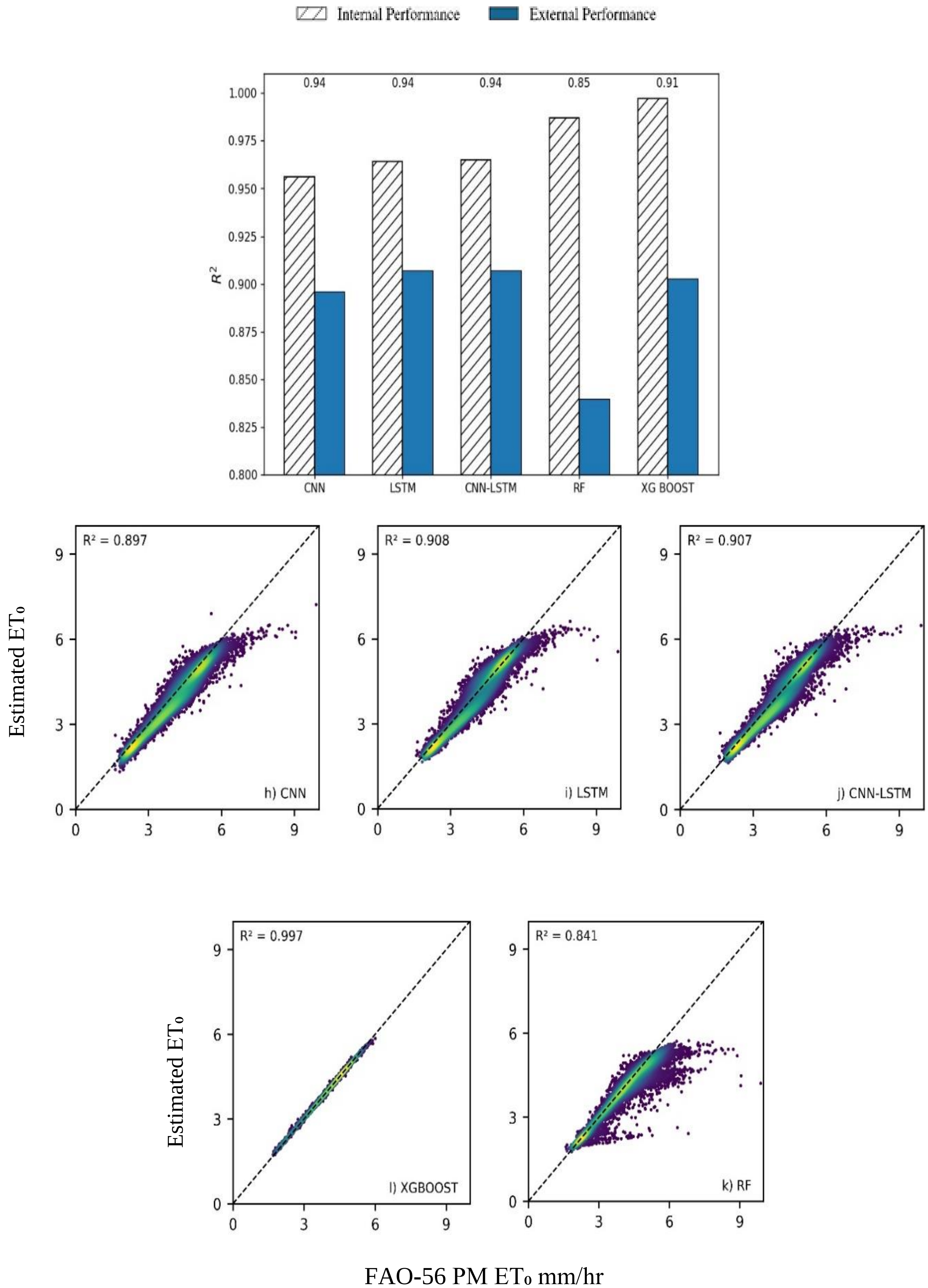

Fig 4.6 Transferability and Comparison of Internal and External Model Performance of Vellayani for full input based variables

4.7.2 Mass transfer-based machine learning models

Mass transfer-based models are restricted to the inputs of humidity and wind speed, producing a marked decline in transferability, with T values falling to between 0.47 and 0.66 (Fig 4.7). The poorest transferability observed across all mass transfer-based ML models. Notably, T decreased progressively from the simpler models (CNN, T = 0.66) to the more complex ones (XGBoost, T = 0.47), implying that the higher-capacity models were fitting station-specific humidity–wind relationships at Pattambi that did not hold at coastal Vellayani, where sea-breeze influence alters the vapour pressure–wind dynamics considerably (Ayaz *et al.*, 2021).

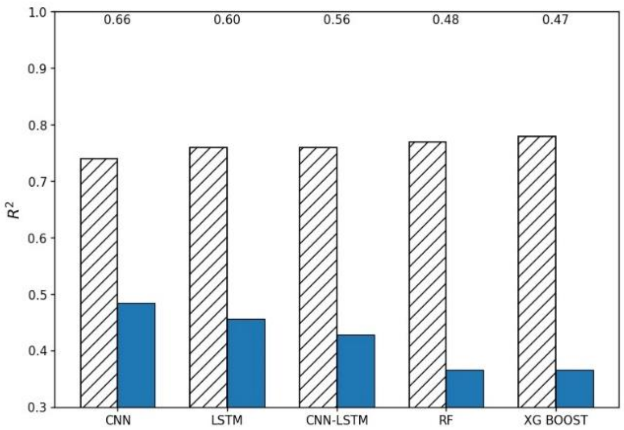

Fig 4.7 Transferability and Comparison of Internal and External Model Performance of Pattambi for mass transfer based models

Transferability in the reverse direction (Fig 4.8) was found improved (T = 0.72–0.82), though internal R^2 itself was moderate range (0.61–0.66) at Vellayani under mass transfer based ML models. The higher spatial transferability observed in Pattambi-trained ML models indicates that inland stations tend to capture more generalised vapour pressure deficit dynamics, which are less influenced by local coastal effects such as sea breeze and high ambient humidity. As a result, models trained at Pattambi generalise better when applied to the coastal station at Vellayani, whereas models trained at Vellayani struggle to replicate the drier, more thermally driven mass-transfer conditions typical of inland locations.

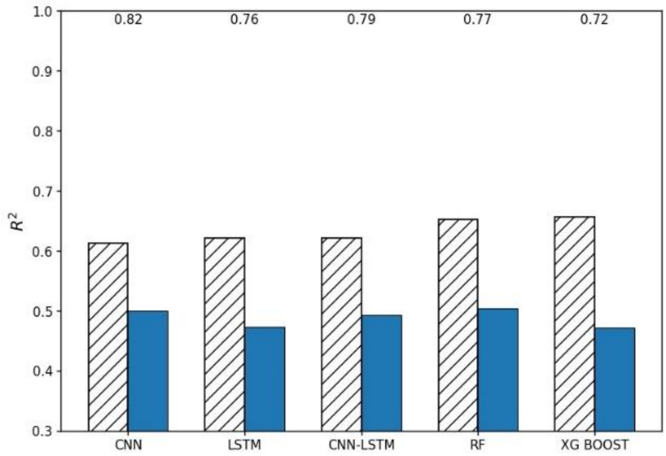

Fig 4.8 Transferability and Comparison of Internal and External Model Performance of Vellayani for mass transfer based models

4.7.3 Radiation-based machine learning models

All radiation-based ML models showed external performance exceeding one, yielding T values in the range of 1.04 to 1.06. Internal R^2 at Pattambi ranged from 0.87–0.88, while external R^2 at Vellayani reached 0.91–0.93. This indicated that radiation-driven ET_0 relationships learned from Pattambi's data generalised exceptionally well to Vellayani, likely reflecting the greater spatial consistency of solar radiation-evapotranspiration dynamics compared to wind- and humidity-driven processes, which are more strongly modulated by local, station-specific factors such as coastal exposure (Sharma *et al.*, 2022).

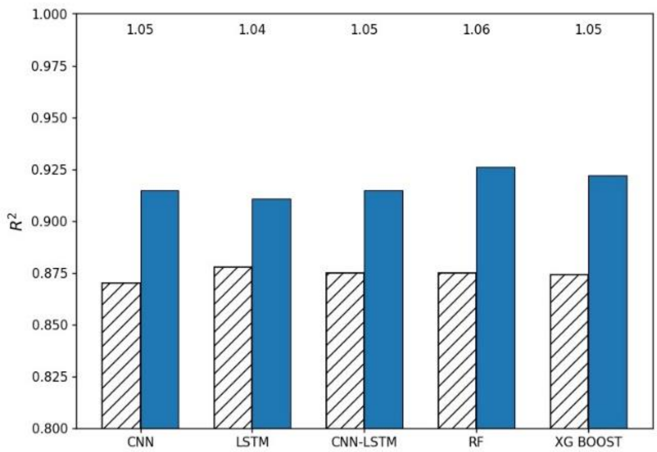

Fig 4.9 Transferability and Comparison of Internal and External Model Performance of Pattambi for radiation based models

In the opposite direction, transferability declined to 0.75–0.80, with internal R^2 at Vellayani being very high (0.96–0.97) but external R^2 at Pattambi dropping to 0.72–0.78. This directional asymmetry mirrors the pattern observed for mass transfer-based inputs and reinforces that models trained on the coastal station (Vellayani) tend to transfer less effectively inland than models trained on the inland station (Pattambi) transfer toward the coast.

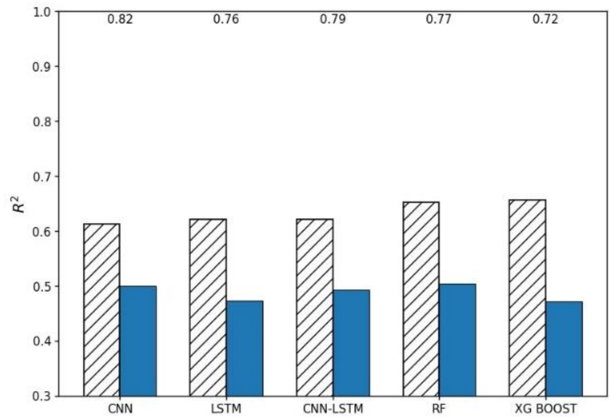

Fig 4.10 Transferability and Comparison of Internal and External Model Performance of Vellayani for radiation based models

4.8 ET_0 PREDICTION FOR THE YEAR 2025

Following model development and spatial transferability assessment, the finalised machine learning models were used to predict daily ET_0 for 2025 at both RARS Pattambi and Vellayani, using actual 2025 meteorological records as input. The predicted ET_0 values were compared against FAO-56 Penman–Monteith estimated ET_0 for 2025 and the agreement was visualised through scatter plots of predicted versus observed ET_0. Among the different types full input-based ML models and radiation-based ML models performed well. Hence, prediction was done based on full input-based model and the radiation-based model. A summary table comparing the performance metrics of the ET_0 prediction across both stations is presented in Table 4.7.

Table 4.7 Performance Metrics of ET_0 Prediction Models for 2025

Category	Model	Pattambi			Vellayani		
		R^2	RMSE	MAE	R^2	RMSE	MAE
Full Input Variable	CNN	0.958	0.216	0.164	0.920	0.253	0.201
	LSTM	0.969	0.186	0.135	0.946	0.213	0.167
	CNN-LSTM	0.957	0.236	0.192	0.933	0.239	0.196
	RF	0.986	0.132	0.081	0.980	0.128	0.097
	XGBoost	0.994	0.090	0.062	0.995	0.065	0.049
Radiation-based	CNN	0.954	0.225	0.167	0.973	0.150	0.121
	LSTM	0.959	0.212	0.143	0.980	0.130	0.113
	CNN-LSTM	0.960	0.210	0.147	0.982	0.124	0.106
	RF	0.960	0.209	0.143	0.982	0.123	0.100
	XGBoost	0.960	0.209	0.141	0.979	0.131	0.106

4.8.1 Prediction based on full input-based Category

Under the full variable-based input category, all five models showed close clustering along the 1:1 line at both stations, confirming strong agreement between predicted and FAO-56 PM ET_0 for 2025 has shown in fig 4.11. At Pattambi, XGBoost (R^2 = 0.994) and RF (R^2 = 0.986) achieved the highest accuracy, followed by LSTM (R^2 = 0.969), CNN (R^2 = 0.958), and CNN-LSTM (R^2 = 0.957), with the deep learning models showing marginally wider scatter above 5 mm day^{-1}. At Vellayani, also the same performance with XGBoost (R^2 = 0.995) and RF (R^2 = 0.980) while the deep learning models showed comparatively greater dispersion with LSTM (R^2 = 0.946) and CNN-LSTM (R^2 = 0.933). CNN recorded the lowest accuracy (R^2 = 0.920), with visible scatter in the mid-range (3–5 mm day^{-1}).

Overall, the ensemble tree-based models consistently outperformed the deep learning architectures at both stations, with XGBoost emerging as the most accurate model (R^2 = 0.994–0.995). The deep learning models, though still achieving $R^2 > 0.92$ in all cases, showed slightly lower accuracy at Vellayani than at Pattambi, indicating marginally greater year-to-year variability in the coastal stations for 2025.

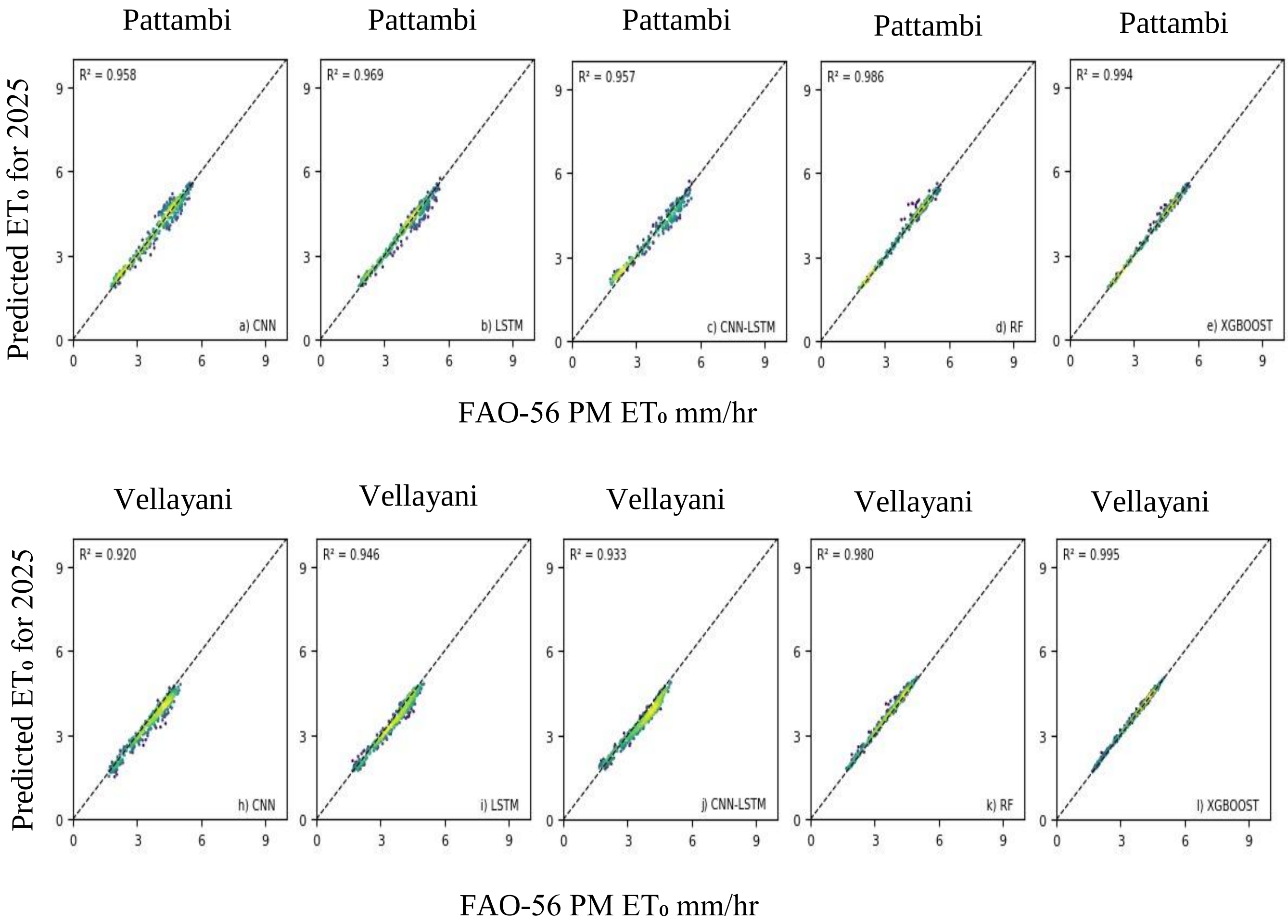

Fig 4.11 Scatter plot of predicted ET_0 for 2025 vs FAO-56 PM for full input based variables

4.8.2 Radiation-Based Category

Under the radiation-based input category, all five models again showed close alignment with the 1:1 line at both stations, though with marginally lower accuracy than the full variable-based category which shown in Fig 4.12. At Pattambi, all the models performed identically, with CNN-LSTM, RF, and XGBoost recorded $R^2 = 0.960$, LSTM $R^2 = 0.959$, and CNN $R^2 = 0.954$. The narrow spread across models indicates limited sensitivity to model architecture when using solar radiation and temperature as primary inputs. At Vellayani, the accuracy improved across all models relative to Pattambi, with CNN-LSTM and RF both reaching $R^2 = 0.982$, LSTM $R^2 = 0.980$, XGBoost $R^2 = 0.979$, and CNN being the lowest $R^2 = 0.973$.

Overall, the radiation-based models performed consistently well at both stations ($R^2 = 0.954–0.982$), with Vellayani showing slightly higher and more tightly clustered predictions than Pattambi. Unlike the full variable-based category, where the ensemble tree-based models clearly outperformed the deep learning architectures, the radiation-based models showed comparable performance across all five architectures. This suggested that radiation and temperature drove the estimation; model choice has a smaller influence on 2025 prediction accuracy than the full input-based category variables (Mandal *et al.*, 2026).

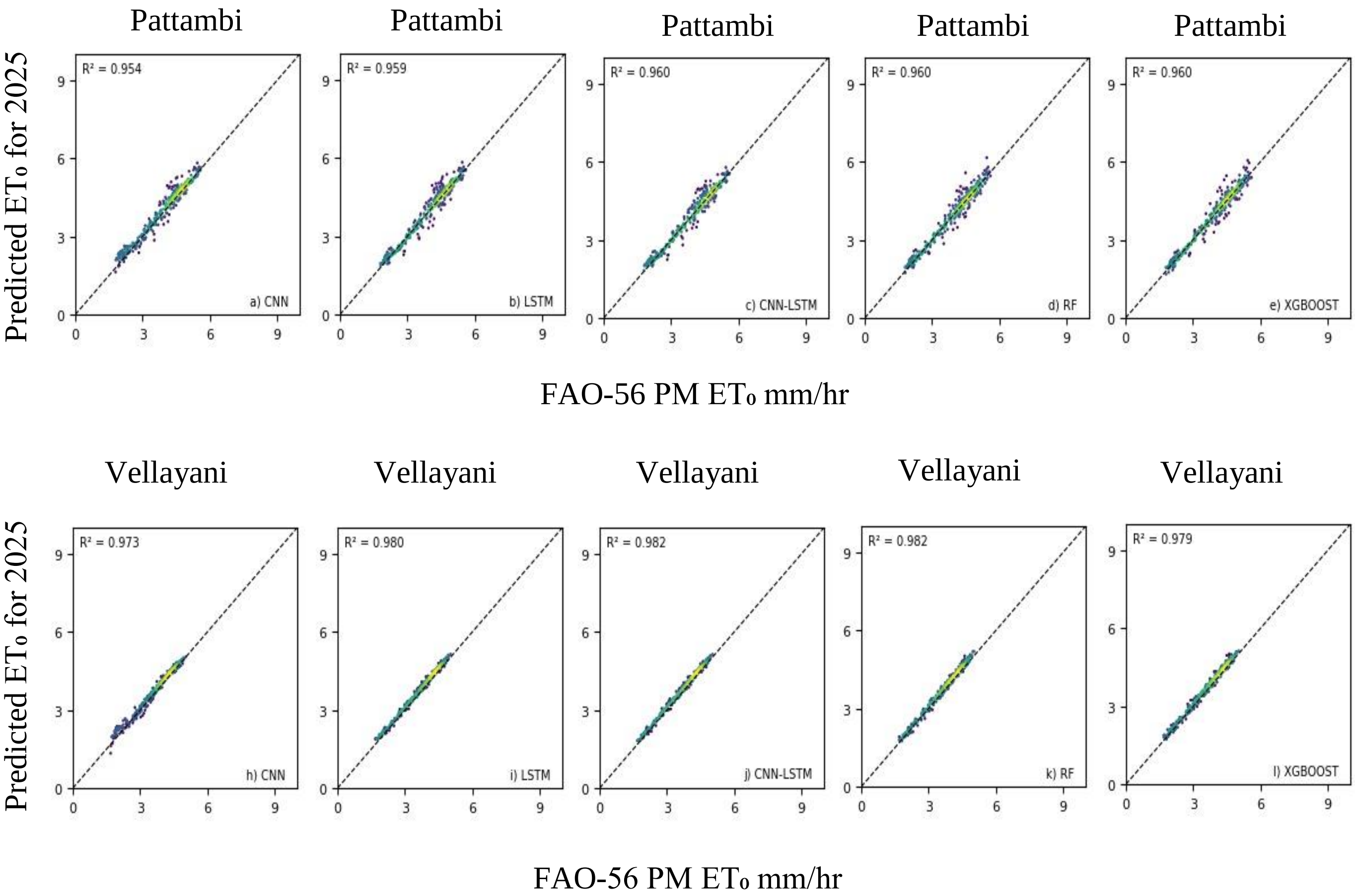

Fig 4.12 Scatter plot of predicted ET_0 for 2025 vs FAO-56 PM for radiation based models

SUMMARY AND CONCLUSION

CHAPTER V

SUMMARY AND CONCLUSION

The present study entitled *"Evapotranspiration Estimation Using Empirical and Machine Learning Models: A Comparative Assessment of Accuracy and Transferability"* was undertaken to estimate reference evapotranspiration (ET_0) at two contrasting agroclimatic stations in Kerala, namely RARS Pattambi in Palakkad district (inland, semi-arid) and the College of Agriculture, Vellayani in Thiruvananthapuram district (coastal, humid). The study also aimed to compare the performance of machine learning (ML) models against conventional empirical approaches and to evaluate the spatial transferability of the developed ML models across the two locations. Daily meteorological data spanning 30 years (1995–2025) were collected from the agrometeorological observatories at both stations, comprising maximum and minimum air temperature, relative humidity (morning and afternoon), wind speed, sunshine duration, rainfall, and pan evaporation. The FAO-56 Penman–Monteith (FAO-56 PM) equation was adopted as the standard reference to compute ET_0 values, which served as the target variable for machine learning model training and evaluation. Five machine learning models- Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), hybrid CNN-LSTM, XGBoost, and Random Forest (RF) were developed under four input categories: full input variable-based, temperature-based, mass transfer-based, and radiation-based. Totally six empirical models were evaluated under the reduced-input categories: Hargreaves–Samani and Dorji (temperature-based), Dalton and Meyer (mass transfer-based), and Priestley–Taylor and Makkink (radiation-based). Model performance was assessed using R^2, RMSE, and MAE, and spatial transferability was quantified using a transferability score (T = External R^2 / Internal R^2). The best-performing models were further used to predict ET_0 for the year 2025 and validated against FAO-56 PM estimated values.

The results of the study revealed that machine learning models consistently and substantially outperformed empirical models across all input categories at both stations. Under full input variable conditions, all ML models achieved excellent predictive accuracy ($R^2 > 0.95$ at both stations), with RF and XGBoost emerging as the best performer at pattambi and vellyani ($R^2 = 0.983$ at Pattambi, $R^2 = 0.996$ at Vellayani). Among the deep learning architectures, CNN-LSTM outperformed the standalone LSTM and CNN models. Performance was notably lower under the temperature-based input category, where the empirical models achieved R^2 values of only 0.25–0.47, while ML models improved the accuracy to 0.42–0.57, with LSTM performing best at Pattambi and Random Forest at Vellayani. Mass transfer-based models showed moderate improvement over temperature-based models, with ML models achieving R^2 of 0.76–0.76 at Pattambi and 0.61–0.66 at Vellayani, while the empirical Dalton and Meyer models remained below with $R^2 = 0.60$. Radiation-based models delivered the strongest performance among the reduced-input categories: at Pattambi, LSTM led with $R^2 = 0.878$, while at Vellayani XGBoost achieved $R^2 = 0.974$, with all ML models exceeding $R^2 = 0.96$. The radiation-based empirical models, particularly Makkink ($R^2 = 0.841$ at Pattambi, $R^2 = 0.947$ at Vellayani), also performed substantially better than the other empirical categories, confirming the stronger physical relationship between solar radiation and ET_0 in Kerala's climate. The following are the specific findings of the study:

- Under the full input variable category, RF and XGBoost achieved the highest prediction accuracy at Pattambi ($R^2 = 0.983$, RMSE = 0.137 mm/day) and vellayani ($R^2 = 0.996$, RMSE = 0.059 mm/day) respectively. Establishing ensemble tree-based models as the most accurate approaches when complete meteorological data are available.

- Temperature-based ML models consistently outperformed the Hargreaves–Samani and Dorji empirical models at both stations. The performance gap was widest at Vellayani, where the coastal humid climate limits the reliability of simple temperature-based empirical relationships.
- Mass transfer-based ML models showed significant improvement over empirical Dalton and Meyer models, with CNN-LSTM achieving the highest R^2 (0.760) at Pattambi and Random Forest performing best at Vellayani (R^2 = 0.657). The empirical mass transfer models performed better at Vellayani than at Pattambi, likely due to the more consistent relative humidity conditions at the coastal station.
- Radiation-based models, both empirical and ML, delivered the best performance among all reduced-input categories. The Makkink model outperformed Priestley–Taylor at both locations, while LSTM (R^2 = 0.878) and XGBoost (R^2 = 0.974) were the best-performing ML models at Pattambi and Vellayani, respectively.
- Across all input categories, the overall performance ranking of ML models was XGBoost > Random Forest > CNN-LSTM > LSTM > CNN, with radiation-based inputs producing the best results and temperature-based inputs producing the lowest accuracy.

The spatial transferability analysis revealed distinct directional asymmetry in model transferability between the two contrasting stations. Under the full input variable category, all ML models transferred well in both directions, with T values of 0.89–0.95 (Pattambi to Vellayani) and 0.85–0.94 (Vellayani to Pattambi), demonstrating that comprehensive meteorological inputs ensure robust cross-station generalisation. Mass transfer-based models showed the poorest transferability in the Pattambi-to-Vellayani direction (T = 0.47–0.66), attributable to the marked differences in sea-breeze-driven vapour pressure dynamics between the inland and coastal stations. Interestingly, the reverse direction (Vellayani to

Pattambi) yielded substantially better transferability (T = 0.72–0.82) for mass transfer-based models, indicating that the generalised vapour pressure deficit conditions at the coastal station transfer more readily to the inland environment than the reverse. Radiation-based ML models trained at Pattambi demonstrated exceptional transferability to Vellayani (T = 1.04–1.06), with external R^2 at Vellayani actually exceeding internal R^2 at Pattambi, reflecting the greater spatial consistency of solar radiation–ET_0 relationships compared to humidity- or wind-driven processes. The reverse direction transferability for radiation-based models (T = 0.75–0.80) was comparatively weaker, confirming that Pattambi-trained models consistently generalised better toward the coast than Vellayani-trained models generalised inland.

The ET_0 predictions for 2025 using the best-performing models validated these findings well. Under the full input category, XGBoost achieved the highest prediction accuracy at both Pattambi (R^2 = 0.994, RMSE = 0.090 mm/day) and Vellayani (R^2 = 0.995, RMSE = 0.065 mm/day), while radiation-based models showed consistently high performance across all five ML architectures (R^2 = 0.954–0.982), with minimal differences between models. These results confirmed that radiation-based inputs, when combined with any of the ML architectures evaluated in this study, provide a reliable and transferable framework for ET_0 estimation in the humid tropical climatic zones of Kerala, even in the absence of humidity and wind speed data.

Overall, the study demonstrated that machine learning models, particularly XGBoost and Random Forest under full input conditions, and LSTM and CNN-LSTM under radiation-based inputs, are highly effective tools for reference evapotranspiration estimation at both inland and coastal tropical stations in Kerala. Radiation-based input combinations consistently yielded the best trade-off between data requirement and prediction accuracy, making them particularly suitable for deployment in data-scarce conditions. The spatial transferability analysis confirmed that models trained at the inland station of Pattambi transfer more effectively to the coastal environment at Vellayani than in the reverse direction, a finding with direct implications for extending ET_0 estimation to ungauged or data-limited locations in the region. The successful validation of ET_0 predictions for 2025 further establishes the practical utility and temporal generalisability of the developed models. In short, the findings of this study provide a sound scientific basis for adopting ML-based ET_0 estimation frameworks for improved irrigation planning and water resource management across the diverse agroclimatic zones of Kerala.

REFERENCES

REFERENCES

Ahmadi, S.H., Andersen, M.N. and Plauborg, F. *et al*. 2023. A review of machine learning approaches for evapotranspiration estimation. *Agric. Water Manag.* 284: 108324. Available: https://doi.org/10.1016/j.agwat.2023.108324 .

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M. 1998. Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper No. 56. Food and Agriculture Organization, Rome, Italy. Available: https://www.fao.org/4/x0490e/x0490e00.htm .

Allen, R.G., Walter, I.A., Elliott, R.L., Howell, T.A., Itenfisu, D., Jensen, M.E. and Snyder, R.L. 2005. The ASCE Standardized Reference Evapotranspiration Equation. American Society of Civil Engineers, Reston, Virginia, USA. Available: https://doi.org/10.1061/9780784408056 .

Ayaz, A., Rajesh, M., Singh, S.K. and Rehana, S. 2021. Estimation of reference evapotranspiration using machine learning models with limited data. *AIMS Geosci.* 7(3): 268-290. Available: https://doi.org/10.3934/geosci.2021016 .

Chang, Y., Zhang, C., Huang, J., Chang, H., Wang, C. and Huo, Z. 2025. Machine learning for reference crop evapotranspiration modeling: a state-of-the-art review and future directions. *Agronomy* 15: 2038. Available: https://doi.org/10.3390/agronomy15092038 .

Chen, J., Chen, K. and Zhu, Y. 2026. Hybrid CNN-LSTM-attention model for enhanced reference crop evapotranspiration estimation toward optimized irrigation scheduling. *J. Agric. Food Res.* 25: 102545. Available: https://doi.org/10.1016/j.jafr.2025.102545 .

Debnath, S., Adamala, S. and Raghuwanshi, N.S. 2015. Sensitivity analysis of FAO-56 Penman-Monteith method for different agro-ecological regions of India. *Environ. Process.* 2: 689-704. Available: https://doi.org/10.1007/s40710-015-0107-1 .

Dong, J., Xing, L., Cui, N., Zhao, L., Guo, L., Wang, Z., Du, T., Tan, M. and Gong, D. 2024. Estimating reference crop evapotranspiration using improved convolutional bidirectional long short-term memory network by multi-head attention mechanism in the four climatic zones of China. *Agric. Water Manag.* 292: 108665 Available: https://doi.org/10.1016/j.agwat.2023.108665.

Hadria, R., Benabdelouhab, T., Lionboui, H. and Salhi, A. 2021. Comparative assessment of different reference evapotranspiration models towards a fit calibration for arid and semi-arid areas. *J. Arid Environ.* 184: 104318. Available: https://doi.org/10.1016/j.jaridenv.2020.104318 .

Khosravi, K., Choubin, B. and Mosavi, A. *et al*. 2024. Estimation of reference evapotranspiration using artificial intelligence models under varying climatic conditions. *Appl. Water Sci.* 14: 25. Available: https://doi.org/10.1007/s13201-023-02058-2 .

Li, M., Zhou, Q., Han, X. and Lv, P. 2024. Prediction of reference crop evapotranspiration based on improved convolutional neural network (CNN) and long short-term memory network (LSTM) models in Northeast China. *J. Hydrol.* 645: 132223. Available: https://doi.org/10.1016/j.jhydrol.2024.132223 .

Mandal, N., Roshni, T. and Chanda, K. 2026. Empirical and ML models for estimation of evapotranspiration: comparison of performance and evaluation of spatial transferability. *Soft Comput.* 30: 2701-2719. Available: https://doi.org/10.1007/s00500-026-11216-y .

Mohammadi, B., Chen, M., Nikoo, M.R., Cheraghalizadeh, M., Yu, Y., Zhang, H. and Yu, R. 2024. An explainable hybrid framework for estimating daily reference evapotranspiration: combining extreme gradient boosting with Nelder-Mead method. *J. Hydrol.* 644: 132130 Available: https://doi.org/10.1016/j.jhydrol.2024.132130.

Raja, P., Sona, F., Surendran, U., Srinivas, C.V., Kannan, K., Madhu, M., Mahesh, P., Annepu, S.K., Ahmed, M., Chandrasekar, K., Suguna, A.R., Kumar, V. and Jagadesh, M. 2024. Performance evaluation of different empirical models for reference evapotranspiration estimation over Udhagamandalam, The Nilgiris, India. *Sci. Rep.* 14: 12429. Available: https://doi.org/10.1038/s41598-024-60952-4 .

Sharma, G., Singh, A. and Jain, S. 2022. A hybrid deep neural network approach to estimate reference evapotranspiration using limited climate data. *Neural Comput. Appl.* 34: 4013-4032. Available: https://doi.org/10.1007/s00521-021-06661- .

Sun, X., Zhang, B., Dai, M., Gao, R., Jing, C., Ma, K., Gu, S., Gu, L., Zhen, W. and Gu, X. 2024. Research on methods for estimating reference crop evapotranspiration under incomplete meteorological indicators. *Front. Plant Sci.* 15: 1354913. Available: https://doi.org/10.3389/fpls.2024.1354913 .

Tabari, H., Grismer, M.E. and Trajkovic, S. 2013. Comparative analysis of 31 reference evapotranspiration methods under humid conditions. *Irrig. Sci.* 31(2): 107-117. Available: https://doi.org/10.1007/s00271-011-0295- .

UNESCO [United Nations Educational, Scientific and Cultural Organization]. 2024. *United Nations World Water Development Report 2024: Water for Prosperity and Peace.* UNESCO, Paris, France Available: https://unesdoc.unesco.org/ark:/48223/pf0000388950.

Woo, D.K. 2026. Estimating actual evapotranspiration from widely available meteorological data with a hybrid CNN-LSTM. *Agric. Water Manag.* 323: 110078. Available: https://doi.org/10.1016/j.agwat.2025.11007 .

Wu, M., Feng, Q., Wen, X., Deo, R.C., Yin, Z., Yang, L. and Sheng, D. 2020. Random forest predictive model development with uncertainty analysis capability for the estimation of evapotranspiration in an arid oasis region. *Hydrol. Res.* 51(4): 648-665. Available: https://doi.org/10.2166/nh.2020.01 .

Wu, Y., Zhang, L. and Chen, X. *et al*. 2024. Improved reference evapotranspiration estimation using machine learning and satellite-derived variables. *Sci. Total Environ.* 947: 174480. Available: https://doi.org/10.1016/j.scitotenv.2024.174480 .

EVAPOTRANSPIRATION ESTIMATION USING EMPIRICAL AND MACHINE LEARNING MODELS: A COMPARATIVE ASSESSMENT OF ACCURACY AND TRANSFERABILITY

by

AFNA HAMEEM H (2022-02-007)

ABHIRAM M A (2022-02-010)

SREENANDANA P V (2022-02-013)

PROJECT REPORT

Submitted in partial fulfilment of the requirement for the degree of

BACHELOR OF TECHNOLOGY

in

AGRICULTURAL ENGINEERING

Faculty of Agricultural Engineering and Technology

Kerala Agricultural University

DEPARTMENT OF IRRIGATION AND DRAINAGE ENGINEERING

KELAPPAJI COLLEGE OF AGRICULTURAL ENGINEERING AND

FOOD TECHNOLOGY TAVANUR - 679573

MALAPPURAM, KERALA, INDIA

2026

ABSTRACT

Evapotranspiration Estimation Using Empirical and Machine Learning Models: A Comparative Assessment of Accuracy and Transferability

This study compared empirical and machine learning (ML) models for estimating reference evapotranspiration (ET_0) at two agroclimatically contrasting stations in Kerala: RARS Pattambi (inland, semi-arid) and the College of Agriculture, Vellayani (coastal, humid). Using thirty years (1995–2025) of daily meteorological data, the FAO-56 Penman–Monteith equation served as the benchmark. Five ML models - CNN, LSTM, CNN-LSTM, XGBoost, and Random Forest—were developed under four input configurations (full variable, temperature-based, mass transfer-based, and radiation-based) and compared against six empirical models, with accuracy assessed via R^2, RMSE, and MAE, and spatial transferability quantified through a transferability score.

ML models consistently outperformed empirical approaches across all input categories. With complete inputs, all ML models achieved $R^2 > 0.95$, with XGBoost RF performing best. Under reduced inputs, radiation-based models performed strongest, with LSTM and XGBoost exceeding an R^2 of 0.87, while the empirical Makkink model also performed comparatively well, reflecting the strong physical link between solar radiation and ET_0 in Kerala's climate. Temperature-based models performed weakest overall.

Transferability analysis showed that models trained at Pattambi transferred more effectively to Vellayani than the reverse, particularly for radiation-based inputs, where external validation exceeded internal accuracy. Mass transfer-based models exhibited the weakest cross-station transferability, reflecting differing humidity dynamics between the inland and coastal environments. Validation using 2025 data confirmed these patterns.

Overall, ML models - particularly XGBoost and Random Forest with complete data, and LSTM/CNN-LSTM under radiation-based inputs - offer a reliable and transferable framework for ET_0 estimation. Radiation-based inputs are recommended for data-scarce settings, and the demonstrated inland-to-coastal transferability supports extending these models to ungauged locations, providing a scientific foundation for improved irrigation planning and water resource management across Kerala.