Internet of Things (IoT) and Machine Learning (ML) Assisted Reference Evapotranspiration (ET_O) Estimations

Rab Nawaz Bashir*, Rana Muhammad Saleem[2], Zahid Abbas[1], Haris Ali Khan[1], Dewan Muhammad Qaseem Hussain[1], Sarfaraz Natha[3]

[1]Department of Computer Science, COMSATS University Islamabad, Vehari Campus, Pakistan
[2]Department of Computer Science, University of Agriculture, Faisalabad, Pakistan
[3]Department of Software Engineering, Sir Syed University of Engineering & Technology, Karachi, Pakistan
*Corresponding author: rabnawaz@ciitvehari.edu.pk

Abstract

Reference Evapotranspiration (ET_O) is the amount of irrigation water required by a model crop to grow at its optimal level under prevailing environmental conditions. Freshwater is a scarce resource that needs to be used judiciously according to the ET_O rate for the conservation of irrigation water. Existing methods of ET_O rate are complex to be applied at the farmer level. There is a need for a solution to determine the ET_O rate from the available minimal set of environmental conditions. Internet of Things (IoT) and Machine Learning (ML) assisted monthly ET_O rate estimation from the directly sensed temperature of the crop field is proposed. IoT-assisted directly sensed temperature from the crop field is more effective in the determination of actual crop field ET_O rate. The proposed solution for ET_O rate determination can be very useful in Precision Agriculture (PA) applications. The study also compares the performance of Naïve Bayes, Support Vector Classifier (SVC), and K-Nearest Neighbors (KNN) in the determination of ET_O rate from local crop field temperature. The performance of the ML models is evaluated based on accuracy, F-measure, and recall for ET_O estimation. The proposed ET_O estimations are also compared against the Penman-Monteith method of ET_O measurements to benchmark the performance of the proposed solution.

Keywords—Internet of Things (IoT), Precision Irrigation (PI), Precision Agriculture (PA), Evapotranspiration (ET), Reference Evapotranspiration (ET_O), Machine Learning (ML), Support Vector Classifier (SVC), K-Nearest Neighbors (KNN), Naïve Bayes

1 Introduction

EVAPOTRANSPIRATION (ET_O) rate is the amount of irrigation water required for the successful growth of the crop under the prevailing environmental conditions. It is the total amount of water that evaporates from the soil and the plant surface. The amount of irrigation water required for a crop is determined based on ET_O. The standard ET_O rate of grass crops is called Reference ET_O [1]. ET_O is also used to determine the crop potential of a specific crop [2].

Water is an essential part of agriculture activities and a fundamental element of every ecosystem. Water is the main pillar of every ecosystem from forests to lakes. Freshwater resources are dwindling at an alarming rate. Water scarcity is the major challenge of the present era. Around two to three thousand per liter of water is required to serve the food requirements of one person for one day. Seventy (70%) percent of freshwater is used in agriculture. Judicious use of natural resources especially irrigation water is very important to feed humans. Climatic changes also have seriously affected the issue of fresh irrigation water scarcity. It is very important to act by every means to promote the conservation of freshwater especially used for irrigation purposes [3][1].

The efficient use of irrigation water is very important to support sustainable developments in agriculture [4]. The human population is growing at an alarming rate that would require more food in the future [5]. According to an estimate, the human population would be 9.6 billion in 2050 [6]. While the productivity

ISSN: 2523-0379 (Online), ISSN: 1605-8607 (Print)
DOI: https://doi.org/10.52584/QRJ.1902.13

in agriculture has decreased to a significant level due to reductions in available natural resources. The tremendous increase in the world population, coupled with the reduction in the availability of natural resources like arable land and irrigation water has seriously created the concern of food scarcity [5].

Water is a scarce resource that needs to be used efficiently for sustainable developments in agriculture. Around sixty (60%) percent of freshwater is used for agriculture purposes worldwide [4]. Most of the fresh irrigation water used in agriculture is wasted due to poor agriculture practices. Irrigation water is applied without considering the environmental conditions and ET_O rate. Efficient use of the irrigation water according to the prevailing ET_O rate is very important to conserve irrigation water while maintaining yield [7] [8].

Application of irrigation water according to the ET_O rate is important for its efficient use. Irrigation water is usually applied without ET_O rate which results in wastage of irrigation water. ET_O rates adjust the amount of irrigation water according to the prevailing environmental conditions.

There is an immense need to explore modern means to improve agriculture productivity with a decreased level of available natural resources [9]. The efficient use of irrigation water is a serious challenge, and many technologies are working to deal with the issue of scarcity of irrigation water [10]. Internet of Things (IoT) is a new paradigm with the capability to embed sensors and smart objects into the environment [11]. IoT is very successful for context-oriented applications in agriculture. The applications of IoT in health, traffic control, smart cities, smart homes, and agriculture have shown significant success. Many disciplines of human life are in a transition state due to the emergence of IoT-based applications. Agriculture is a strong candidate for IoT applications to effectively deal with the different issues.

The application of IoT and Wireless Sensor Networks (WSN) has emerged with the capability to provide exciting precision and smart agriculture applications [12]. IoT is a promising field that has the potential to provide the solution to many problems in agriculture and can play a crucial role in the intensification and modernization of agriculture [13]. Many IoT-assisted crop field monitoring, environmental monitoring, yield mapping, and smart farm management solutions have emerged in recent years [14]. The most profound applications of IoT in agriculture are related to predictive irrigation assessment, greenhouse environment monitoring, soil temporal site maps development, ambient crop environment monitoring, and yield assessment [12]. IoT application in agriculture proved to result in operational efficiency and productivity while conserving natural resources.

ET_O is a nonlinear complex process that is difficult to define in the relationship [3]. Data-driven Machine Learning (ML) assisted decisions are required to estimate the ET_O rate. Many solutions are proposed in recent years to determine the ET_O rate from meteorological conditions. There is a need for a solution that can determine the ET_O of a local location using minimal meteorological conditions.

The study presented in this paper proposes the ET_O estimation from only prevailing temperature data using ML capabilities. The study also compares the performance of Naïve Bayes, K-Nearest Neighbor (KNN), and Support Vector Classifier (SVC) in the estimation of the ET_O form only temperature data. The unique feature of the proposed solution is that it determines the ET_O rate according to the Food and Agriculture Organization (FAO) recommended Penman-Monteith ET determination method. The crop field directly sensed data also helps to optimize the ET_O rate according to prevailing conditions.

2 Literature Review

Different studies are reviewed for ET_O and irrigation water estimation using the Wireless Sensor Network (WSN) and IoT capabilities. Ali Rashid Niaghi et. al. proposed estimation of ET_O using different ML algorithms. The study also compares the performance of the Gene Expression Programming (GEP), Support Vector Classifier (SVC), Multiple Linear Regression (LR), and Random Forest (RF) ML models in estimation of the ET_O. The study uses climatic data of seventeen years. The study results showed that the RF model shows better performance and local climate data produce more compromising results as compared to spatial data [1].

Yamaç, et. al. [3] proposed ET_O of potato crop estimation using different meteorological conditions. The study also compares the performance of the KNN, Artificial Neural Networks (ANN), and Adaptive Boosting (AdaBoost) models to assess the ET of the potato crop. Tikhamarine et. al. Proposed ET_O modeling using five hybrid artificial neural network multi-verse optimizer (ANN-MVO), particle swarm optimizer (ANN-PSO), (ANN)-embedded grey wolf optimizer (ANN-GWO), whale optimization algorithm (ANN-WOA), an ant lion optimizer (ANN-ALO) in India and Algeria. The estimate ET_O by these hybrid models is compared against the Valiantzas-1, 2, and 3 [15]. Muhammad Adnan et. al. proposed a model of ET_O rate estimation with reduced meteorological conditions by applying ML models [16]. Pan, Shufen, et. al. compare the performance of ML and remote sensing models to determine the spatial and temporal variations in the ET_O rate [17]. Granata, Francesco proposed the ET_O rate determination from different combinations of meteorological conditions and compared the performance of M5P Regression Tree, Bagging, Random Forest (RF), and Support Vector Regression (SVR) in forecasting the ET_O rate [18].

Chen et. al. proposed the daily ET_O estimation from deep learning models. The study also determines the performance of Deep Neural Network (DNN), Long Short-Term Memory neural network (LSTM), and Temporal Convolution Neural Network (TCNN) [19]. Ferreira et. al. proposed daily and hourly ET_O rate estimation using the RF, XGBoost, ANN, and CNN models, with a limited set of meteorological conditions [20]. Milan Gocic' proposed a method of ET_O using the Penman-Monteith method and different microenvironmental conditions like temperature, humidity, and wind speed while applying the machine learning algorithms [21]. Feng et. al. suggested ET_O measurements using the ML model for efficient irrigation water applications [22]. Antonio Fernández-López et. al. proposed ET_O estimation from soil moisture conditions. The ET_O by the proposed method is compared against the Penman-Monteith method [23]. Mandeep et. al. proposed ET_O rate determination using H2O framework by Deep Learning-Multilayer Perceptron (DL), Generalized Linear Model (GLM), Random Forest (RF), and Gradient-Boosting Machine (GBM). The study also analyzes the abilities of these models to assess the ET_O rate [24].

The major contribution of the work presented in this paper is ET_O rate determination with only temperature conditions of the crop field. The proposed estimated ET_O is also acceptable for agriculture communities due to the estimation of ET_O rate by the FAO recommended penman-Monteith method. Naïve Bayes, SVC, and KNN models are also compared for performance in estimation of the ET_O rate with temperature data only.

3 Material & Method

In this section, the architecture of the IoT sensor node, the model of ET_O estimations, the sensor used, the prototypes, and the temperature data of the selected location are discussed. The proposed solution is based on the real-time field data using the IoT capability and use of FAO proposed Penman-Monteith method [2] for ET_O estimation. This real-time IoT data is used in ML algorithms to test the accuracy of the ML model to determine the ET_O rate.

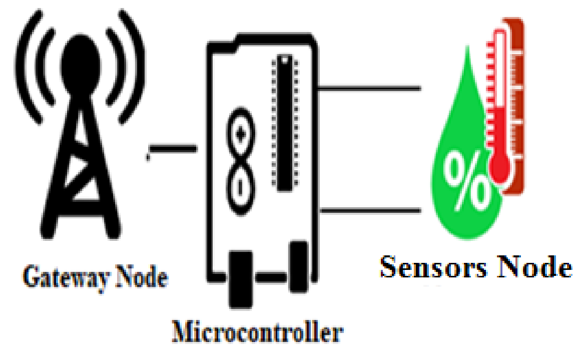

Fig. 1: Archtecture of sesnor node

3.1 Architecture for IoT Devices

The proposed solution is based on the deployments of sensor nodes in the crop field for real-time capture of microenvironment data from the crop fields. This data is transferred to the IoT server using wireless technologies with the help of gateway nodes. The architecture of the proposed IoT devices for real-time crop field environment data capture to determine ET_O rate is shown in Figure 1. The sensor node transmits data to the IoT server through the gateway node. The sensor node is made using the Arduino platform. The environmental data at the IoT server is processed according to the proposed solution to determine the ET_O rate.

3.2 Model of ET_O Estimation

The workflow of the proposed solution is shown in Figure 2 that shows the development of the ML model and prediction by the proposed solution. Directly sensed crop field daily temperature is processed to determine the mean monthly temperature ($Tmean$). Daily maximum temperature ($Tmax$) and daily minimum temperature ($Tmin$) are used to determine the mean monthly maximum (Tmx) and mean monthly minimum temperature (Tnx) by Equation 1 and Equation 2, respectively. Tmx and Tnx are further used to determine the mean monthly temperature ($Tmean$) by Equation 3.

$$Tmx = \frac{(\sum_i^n Tmax_i)}{n} \tag{1}$$

$$Tnx = \frac{(\sum_i^n Tmin_i)}{n} \tag{2}$$

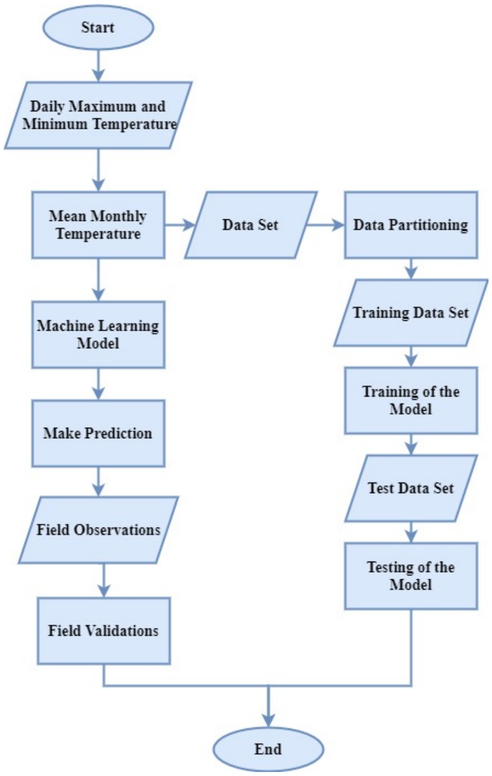

Fig. 2: Workflow of the proposed Evapotranspiration (ET_O)

Mean monthly temperature (Tmean) is obtained by Equation 3, where Tmx is the mean monthly maximum temperature and Tnx is the mean monthly minimum temperature.

$$Tmean = \frac{(Tmx + Tnx)}{2} \tag{3}$$

Mean monthly temperature ($Tmean$) is used in the ML model to determine the ET_O rate accordingly. The predictions by proposed solutions are verified by comparing them against the standard method of ET_O rate determination methods. The ML model is developed by observing the temperature conditions from the crop field and ET_O rate by the Penman-Monteith method. The dataset is partitioned into training and test data sets. The test data set is used to evaluate the performance of the ML model. Once the ML model is developed the directly sensed temperature conditions are used to predict the ET_O from mean monthly temperature ($Tmean$). The prediction made by the proposed solutions is also verified by comparing the ET_O rate determined by the Penman-Monteith methods.

The ET_O rate determined by the proposed solution is compared against the Penman-Monteith method of ET_O rate. Penman-Monteith method is used to determine the ET of a model crop in millimeters per day by Equation 4 [25].

$$\lambda ET = \frac{(\Delta(Rn - G) + PaCp\frac{(e_s - e_a)}{r_a})}{(\Delta + \gamma(\frac{(\Delta + \gamma(1 + r_s/r_a))}{r_a}))} \tag{4}$$

Where r_s and r_a are the surfaces and aerodynamic resistance, γ is the psychometric constant, G is the soil heat flux, $es - ea$ is the vapor deficit, pa is the mean air density, cp is the specific heat of the air, and Δ is the temperature and vapor pressure relationship. The modified Panman-Montieth method to determine the ET_O from climate data is expressed by Equation 5 [25].

$$ET_O = \frac{(0.408\Delta(R_n - G) + \gamma\frac{900}{(T+273)}u_2(e_s - e_a))}{\Delta + \gamma(1 + 0.34u_2)} \tag{5}$$

Where γ is the psychrometric constant, Δ is the slope of vapor pressure deficit, G is the soil heat flux density, Rn is the net radiation, T is the temperature, $u2$ is the wind speed, es is the saturated vapor pressure, ea is the actual vapor pressure and $es - ea$ is the saturation vapor pressure deficit. All these parameters of Equation 4 are derived from the air temperature, humidity, wind speed, latitude, and altitude of the location. The temperature, humidity, and wind speed are directly sensed from the crop field.

3.2.1 Determination of Psychrometric Constant (γ)

The psychrometric constant (γ) is in Kilopascal per degree Centigrade [$kPa°C^{-1}$] is obtained by Equation 6, [26].

$$\gamma = \frac{C_p P}{\epsilon\sigma} \tag{6}$$

Where, σ is the latent heat of vaporization that is 2.45 megajoule per kilogram (MJ kg^{-1}), C_p is the specific heat at constant pressure that is 1.01310^{-3} megajoule per kilogram per degree centigrade ($MJkg^{-1}°C^{-1}$), ϵ is the ratio molecular weight of vapors that is 0.622 and P is the atmospheric pressure in kilopascal (kPa) obtained by elevation of a location expressed by Equation 7, where g is the elevation of the location from sea level.

$$P = 101.30(\frac{(293 - 0.0065g)}{293})^{5.26} \tag{7}$$

3.2.2 The Slope of Vapor Pressure Curve Δ at a Particular Temperature

To determine the relationship between temperature and saturation vapor pressure, the slope of the vapor pressure curve Δ at a particular temperature is obtained in Equation 8, where $Tmean$ is the mean temperature obtained by Equation 3.

$$\Delta = \frac{4098[0.6108exp(\frac{17.27\times Tmean}{Tmean+237.3})}{(Tmean+237.3)^2} \quad (8)$$

3.2.3 Vapor Pressure Deficit $(es-ea)$, Maximum Temperature $(Tmax)$, Minimum Temperature $(Tmin)$, and Maximum Humidity $(Hmax)$ level

Saturation vapor pressure is determined by Equation 9 as a function of temperature. The mean saturation vapor pressure is determined by Equation 10 as the mean of daily maximum and daily minimum temperature.

$$e^o(T) = 0.6108exp[\frac{17.27\times T}{T+237.3}] \quad (9)$$

$$e_s = \frac{e^o(Tmax)+e^o(Tmin)}{2} \quad (10)$$

The actual vapor pressure ea is obtained from daily minimum temperature (Tmin) and daily maximum humidity $(Hmax)$ by Equation 11.

$$e_a = e^o(Tmin)\frac{Hmax}{100} \quad (11)$$

Vapor pressure deficit $(es-ea)$ is obtained by the difference of saturation vapor pressure (es) and actual vapor pressure (ea) by Equation 12.

$$Vapour\ Pressure\ Deficit = e_s - e_a \quad (12)$$

3.2.4 The net radiation (Rn)

The net radiation (Rn) is the difference of radiation falling on the earth's surface and reflected by the earth's surface. The net radiation (Rn) is obtained by Equation 13.

$$Rn = Rns - Rnl \quad (13)$$

Where Rns is the fraction of the radiation absorbed by the earth's surface and Rnl is the radiation reflected by the earth's surface, obtained by Equation 14.

$$Rns = (1-\alpha)\times R_s \quad (14)$$

For grass as a reference crop to determine the reference evapotranspiration the albedo (ϵ) value is taken 0.23. Now the Rns is obtained by Equation 15.

$$Rns = 0.77\times R_s \quad (15)$$

Where Rs is solar radiation, obtained by the difference in air temperature by Hargreaves' radiation formula given in Equation 16.

$$Rs = K_{Rs}\sqrt{(Tmax-Tmin)}Ra \quad (16)$$

Were Ra is the terrestrial radiation, $Tmax$ is the maximum temperature, $Tmin$ is the minimum temperature and K_{Rs} is the adjustment coefficient. Terrestrial radiation is the radiation striking at the earth's surface. The energy changes with the angle of striking and its value is $0.082MJm^{-2}min^{-1}$ when striking perpendicular to the earth's surface. Its value is the function of latitude, date, and time of day and obtained from Table 5, by converting the latitude in degree radians by Equation 17.

$$Radian = \frac{\pi}{180}(Decimaldegree) \quad (17)$$

Rnl is the net longwave energy loss from the earth's surface, obtained by Stefan-Boltzmann's law by Equation 18. The rate of energy loss is proportional to the temperature raised to power fourth.

$$Rnl = \sigma[\frac{(TmaxK^4+TminK^4)}{2}](0.34-0.14\sqrt{e_a})$$
$$(1.35\frac{R_s}{R_{so}}-0.35) \quad (18)$$

Where ϵ is the Stefan-Boltzmann constant $(4.90310^{-9}MJK{-}4m^{-2}day^{-1})$, ea is the actual vapor pressure, Rs/R_{so} is the relative shortwave radiation (limited to $\leq$ 1.0), the $Tmax, K$ is the maximum absolute temperature, $Tmin, K$ is the minimum absolute temperature in Kelvin scale obtained by Equation 19.

$$K = ^\circ C + 273.16 \quad (19)$$

Now Rs/R_{so} is the ratio of the solar radiation (Rs) to the clear-sky solar radiation (R_{so}), where Rs is the solar radiation that reaches the earth surface in a specific period, and R_{so} is the solar radiation that reaches the earth's surface in cloudy conditions. The difference in maximum and minimum temperature is used to determine the level of clouds at a particular location. The Hargreaves' radiation formula is used to determine the Rs of different locations by Equation 20.

$$Rs = K_{Rs}\sqrt{(Tmax-Tmin)}Ra \quad (20)$$

Where Ra is the daily extraterrestrial radiation (Ra) that is the function of altitude of a location obtained by Equation 21 and Table 5, in the appendix.
The clear sky radiation R_{so} is expressed as a function of Ra and elevation of the location and obtained by Equation 21.

$$Rso = (0.75 + 210^{-5}g)Ra \quad (21)$$

3.2.5 The soil heat flux (G)

The soil heat flux (G) is the energy used in the heating of the soil. G is very low as compared to Rn and is often ignored.

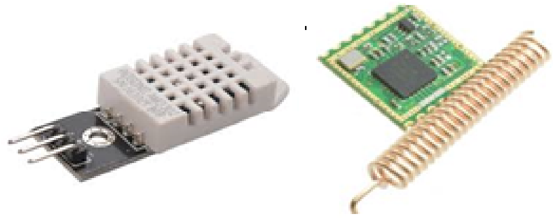

Fig. 3: DHT22 temperature sensor and LoRa module

Fig. 4: Experiment site

3.3 Equipment and Applications

DHT22 is used as a temperature sensor that is cheap and provides accurate measurements. It is a cost-effective, durable sensor shown in Figure 3. DHT22 operates with a voltage range of 3.5 to 5.5 Volts. It only requirs 0.3mA current for observation of temperature. It can observe the temperature from -40 to 80 °C with an accuracy of ±0.5°C. It can read humidity from 0 to 100% with ±1% accuracy.

X1278 wireless module is the latest LoRaTM modulation technology from SEMTECH'. It uses LoRa spread spectrum technology for long-distance wireless communication. LoRa is a long-distance, low interference, low power usage wireless technology that makes it a perfect candidate for IoT applications in agriculture. LoRa module is shown in Fig 3.

3.4 Implementation Detail

This section describes the site selected for experiment purposes, the prototype developed, the application, and detail about the ML model is explained.

3.4.1 Implementation Site

The proposed solution is implemented in district Vehari that is the agricultural intensive area of the province of Punjab in Pakistan. Pakistan is an agricultural country, situated in south Asia. The geographical location of Pakistan is shown in Figure 4. Vehari is situated at Latitude 30°N and Longitude 72.34°at an altitude of 135 m (443 ft) from the sea level. Vehari is the hub of agricultural activities with arid and dusty environmental conditions. Summer lasts from April to Late October with frequent temperatures hit to 45°C to 50°C in summer.

Fig. 5: Sensor node deployment in the field

3.4.2 Hardware Prototype

An Arduino-based hardware prototype is developed to sense the temperature directly from the field, as shown with the sensor node in Figure 5. The sensor node also captures the humidity and wind speed from the crop field to evaluate the proposed solution against the Panmen-Montieth method. The sensor node captures the crop field environmental conditions directly from the crop field. The meteorological data is preprocessed into mean monthly environmental conditions to make monthly predictions. The processing of temperature data at the IoT server is shown in Fig. 6.

4 Data Analysis

Temperature is the important influencing factor of ET_O rate. ET_O rate is directly related to temperature due to the high evaporation and transpiration rate of the crop. Direct sensed temperature from the crop field is also important to make an accurate estimation of the ET_O rate Daily maximum temperature ($Tmax$), daily minimum temperature ($Tmin$) from May to November is captured to determine the mean monthly temperature ($Tmean$) and ET_O rate of the second cropping season. The data from the year 2016 to 2019 is used to make predictions. The temperature data of the selected site for the year 2016 to 2019 is shown in Figure 7, Figure 8, Figure 9 and Figure 10, respectively.
It is observed that the period from May to August is very hot and then temperature tends to decrease in the selected area. $Tmean$ also shows similar behavior to the daily temperature observations. The temperature in each month for each selected Year has a similar

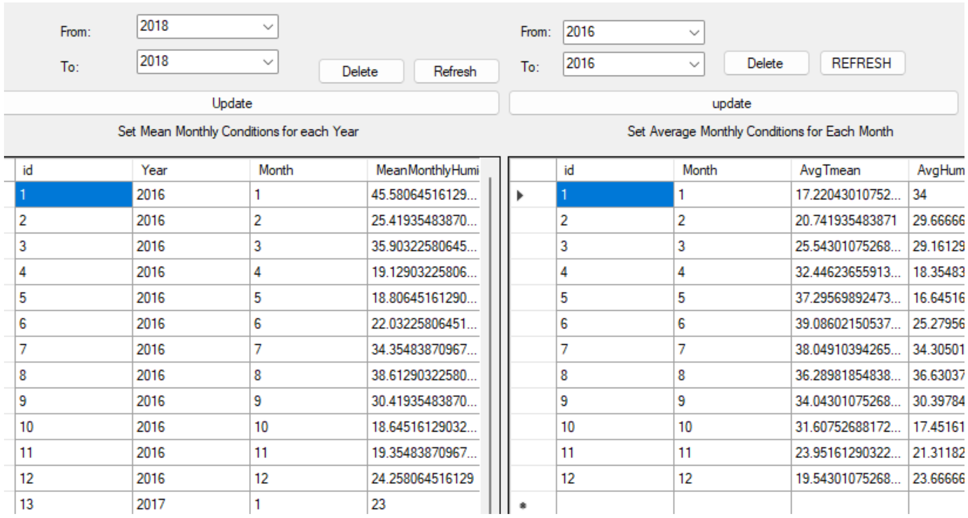

Fig. 6: Temperature analysis module

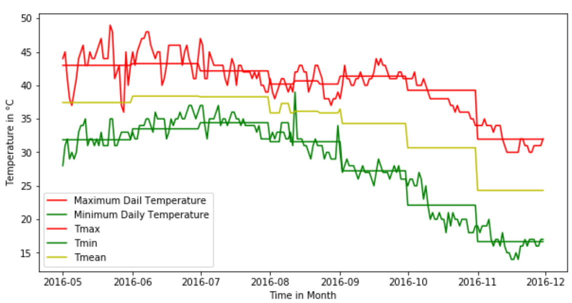

Fig. 7: The temperature of the Year 2016

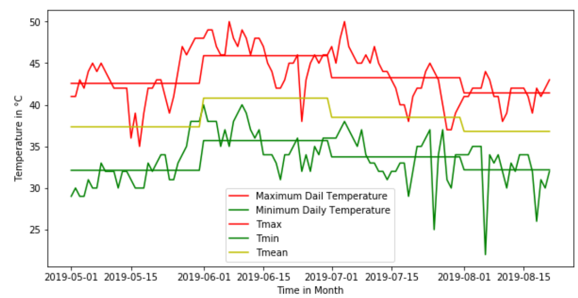

Fig. 10: The temperature of the Year 2019

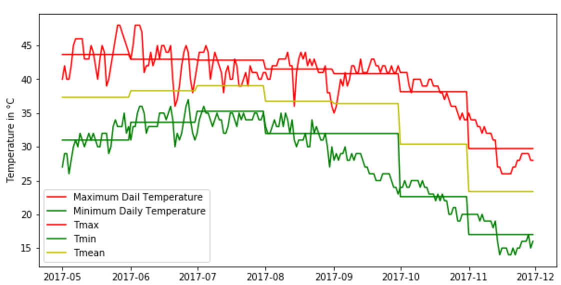

Fig. 8: The temperature of the Year 2017

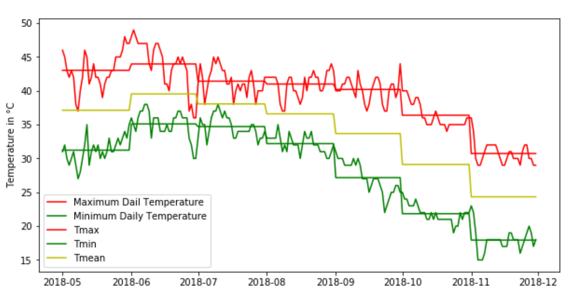

Fig. 9: The temperature of the Year 2018

mean temperature. Mean monthly temperature is the true reflection of the prevailing temperature conditions in the area.

4.1 Machine Learning (ML) Model

For Machine Learning (ML) model, the data set is based on the directly sensed temperature, humidity, and wind speed. Apart from environmental conditions the latitude, altitude of the location, and month of the year are also used to predict the ET_O. The environmental data from 2016 to 2019 is used in the training and testing of the ML model to predict the ET_O from prevailing environmental conditions. The dataset is partitioned into the environment feature matrix and response vector (R). The environment feature matrix is expressed used to predict the response vector. The environment feature matrix is comprised of vectors of dependent variables that are maximum daily temperature ($Tmax$), minimum daily temperature ($Tmin$), mean monthly temperature ($Tmean$), month, latitude, and altitude of the location. Each vector of the environment feature matrix contains values of independent variables represented by Equation 22. Each vector of the environment feature matrix is used to determine the value of the response vector (R) by classification of different ET_O levels ("Low", "Medium", "High"), expressed by Equation 23.

$$M = Tmin, Tmax, Tmean, Month, Latitude, Altitude \tag{22}$$

$$R = ET_O \tag{23}$$

The dataset is also partitioned into two set in the ratio of 80:20 for training and testing of the model. 20% of the dataset is used to evaluate the ML model. Sciket Learn library of ML is used for the implementation of the ML model. Yellow Brick library of python is used to evaluate the ML model on the test dataset.
Following sections present a review of the machine learning algorithms.

4.1.1 Gaussian Naïve Bayes algorithm

Naïve Bayes is a supervised ML algorithm based on the Bayes theorem. Naïve Bayes ML algorithm assumes that every data feature is independent of the others. This algorithm determines the probability of an event by Equation 24.

$$P(c|x) = \frac{P(x)P(x|c)}{P(x)} \tag{24}$$

Where $P(c|x)$ is the probability of event c given that event x has already occurred, is the probability of event x, is the probability of event c, and $P(x|c)$ is the probability of event x given that event c has already occurred. Gaussian Naïve Bayes algorithm is used to continuous data of each class is normally distributed.

4.1.2 Support Vector Classifier (SVC)

SVC is a supervised algorithm that is also used for classification problems. In SVC each data item is plotted into n-dimensional space, with n number of feature sets. A hyperplane is used to differentiate between classes, and to perform classification. The distance between planes is determined by Equation 25.

$$\cos\theta = \frac{\vec{a}.\vec{b}}{||\vec{a}||.||\vec{b}||} \tag{25}$$

4.1.3 K-Nearest Neighbors (KNN)

KNN is a supervised ML algorithm that assumes that similar data is more likely to exist in proximity. This algorithm is used to solve regression as well as classification problems. The distance between the data point is determined.

Temperature (oC)	
Range	Class
>30	Hot
15-30	Mild
<15	Cool

TABLE 1: Temperature Classes [27]

Evapotranspiration (ET_O) Classes	
Range	Class
>=8	High
3-7	Medium
<3	Low

TABLE 2: Evapotransppiration (ET_O) Classes

4.2 Classes for Machine Learning (ML) Model

To implement the ML algorithm, the Scikit-learn libraries of python are used. The classification of the temperature is shown in Table 1 and ET_O in Table 2. The temperature and ET_O classification is done according to FAO recommendations [27]. The selected temperature classification has a subtle impact on the ET_O rate.

5 Evaluations & Findings

The performance of the ML algorithms is determined using the accuracy of the different algorithms with test data. The training and test data ratio is set to 80:20 for evaluation purposes. The SVC model shows the highest accuracy of 87.5% from the test data set. The accuracy of the selected ML models is compared in Table 3. Along with accuracy f-measure, precision, and recall for each feature set are also determined. Precision is the measure of the relevance among the correct prediction among the total predictions, and recall is the ratio of correct predictions out of total predictions made. These measures are evaluated using the yellow brick library of the python programming language. The precision, recall, and f-measure of predictive features of the model are shown in Figure 11, Figure 12, and Figure 13 for different ML models.

The precision, recall and F-measure of predictive features of the Naïve Bayes model are shown in Figure 11. The maximum precision is achieved in the case of the "Medium" case and maximum recall in the case of the "Low" predictive feature. The precision, recall, and

Accuracy of Machine Learning (ML) Algorithm	
Linear SVC accuracy	87.5
Naive-Bayes accuracy	82.5
K-Neighbors accuracy score	86.8

TABLE 3: Comparison of machine learning algorithms

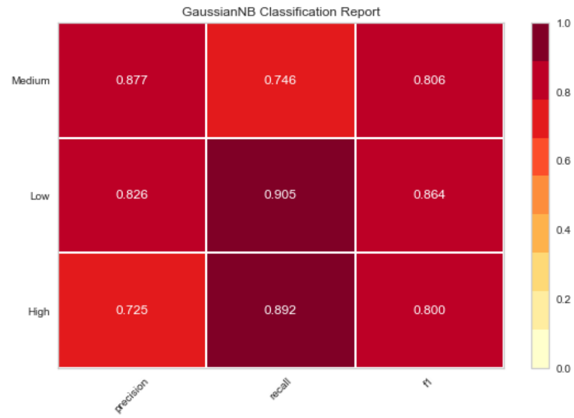

Fig. 11: Gaussian NB classification report

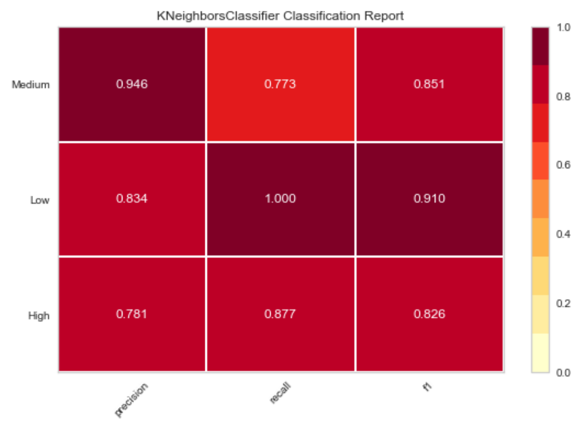

Fig. 12: K-Nearest Neighbour (KNN) classification report

F-measure of predictive features of the KNN model are shown in Figure 12. The maximum precision is achieved in the case of the "Medium" case and maximum recall in the case of the "Low" predictive feature.

The precision, recall, and F-measure of predictive

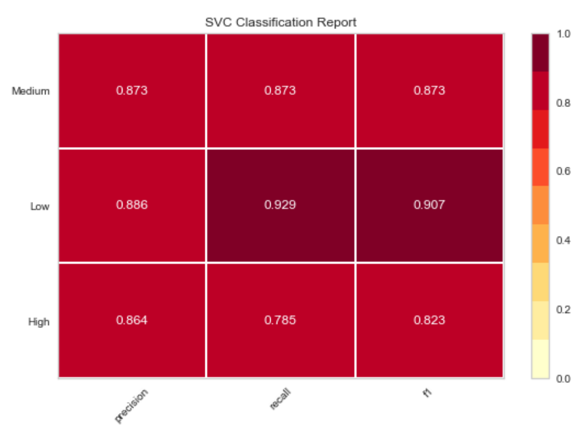

Fig. 13: Linear SVC Classification report

	Ali Rashid Niaghi et. al.				Proposed Approach
	LR	RF	SVM	GEP	SVC
RMSE	0.97	0.82	0.77	0.90	0.29

TABLE 4: Comparison of the proposed model against existing approach

features of the SVC model are shown in Figure 13. The maximum precision is achieved in the case of "Low" and maximum recall in the case of the "Low" predictive feature. The high accuracy and precision show that the ML model is precise to make estimations regarding the ET_O.

5.1 Comparison of Performance with the Existing Approach

Ali Rashid Niaghi et. al. [1] proposed local environmental conditions-based ET_O estimations using four machine learning algorithms. This approach uses temperature and spatial conditions to predict the ET_O and compare the performance of the ML algorithms. The study compares the performance of Gene Expression Programming (GEP), Linear Regression (LR), Random Forest (RF), and Support Vector Machine (SVM). These models are compared based on the Root Mean Squared Errors (RMSE) and shown in Table 4. RMSE is the difference between the observed and predicted values. Along with these observations, the RMSE of the proposed solution using SVC is 0.29 which is significantly lower than the solution proposed by Ali Rashid Naighi [1].

5.2 Field Validations

The predicted ET_O by the proposed solution is compared against the Penman-Monteith, the standard ET_O rate method. The difference in ET_O by proposed solution and Penman-Monteith method for each selected month from the year 2016 to 2019 is shown in Figure 14. The predictions are made by the SVC model being the most accurate model for predicting the ET_O from temperature data. It is observed that the difference in ET_O observations is very low for each month and gradually reduces year by year. The maximum difference in ET by proposed solution and Pan-Monteith method is 2.7 mm day-1 in October of 2016. The minimum difference is 1.1 mm day-1 in November of 2019. The average difference for selected months in 2016 is 2.17 mm day-1, 1.8 mm day-1 in 2017, 1.6 mm day-1 in 2018, and 4.6-mm day-1 in 2019. In 2019, the average difference in ET_O is smaller as compared to the previous years. The difference in ET_O

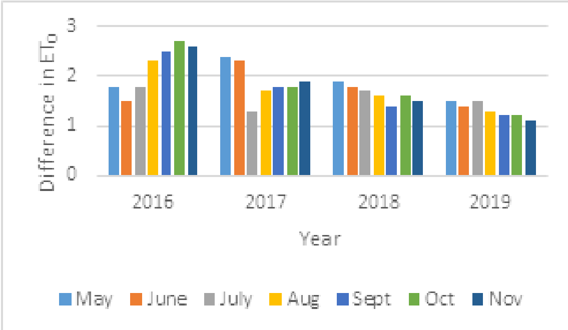

Fig. 14: The difference in ET_O observations by proposed solution and by Penman Montieth Method

gradually decreases with each year due to an increase in performance of the ML model over the time

6 Conclusion

Reference Evapotranspiration (ET_O) determination from minimal environmental conditions is very important in Precision Agriculture (PA) applications. ET_O estimation from the temperature of the crop field using IoT and Machine Learning (ML) approach is proposed in this paper. Performance of the Support Vector Classifier (SCV), K-Nearest Neighbors (KNN), and Naïve Bayes ML models are also compared. SVC is more accurate in ET_O estimation from temperature data as compared to KNN‘ and Gaussian Naïve Bayes classifiers. The ET_O rate predictions made by SVC are compared against the standard Penman-Monteith method. The difference in ET_O rate by the proposed method and standard method is very low.

Appendix

Latitude	Jan	Mar	May	Jul	Aug
0	36.2	37.9	34.8	33.9	35.7
2	35.4	37.8	35.4	34.6	36.1
8	32.8	37.2	37.1	36.5	37.2
12	30.9	36.5	38.0	37.6	37.8
16	28.9	38.1	38.7	38.5	38.1
32	19.9	24.8	40.0	40.7	37.9
34	18.7	23.7	40.0	40.8	37.6
36	17.5	22.6	40.0	40.8	37.4
38	16.2	21.5	39.9	40.8	37.0
48	10.1	23.3	38.8	40.4	34.9

TABLE 5: Daily extraterrestrial radiation (R_a) against different latitude

References

[1] A. R. Niaghi, O. Hassanijalilian, and J. Shiri, “Estimation of Reference Evapotranspiration Using Spatial and Temporal Machine Learning Approaches,” Hydrology 2021, Vol. 8, Page 25, vol. 8, no. 1, p. 25, Feb. 2021.

[2] FAO, “Crop Water Needs Chapter 2,” Crops, 2010. http://www.fao.org/3/s2022e/s2022e07.htm3.1 influence of climate on crop water needs (ET_O) (accessed Aug. 30, 2019).

[3] S. S. Yamaç and M. Todorovic, “Estimation of daily potato crop evapotranspiration using three different machine learning algorithms and four scenarios of available meteorological data,” Agricultural Water Management, vol. 228, p. 105875, Feb. 2020.

[4] S. A. Nikolidakis, D. Kandris, D. D. Vergados, and C. Douligeris, “Energy efficient automated control of irrigation in agriculture by using wireless sensor networks,” Computers and Electronics in Agriculture, vol. 113, pp. 154–163, 2015.

[5] O. Elijah, T. A. Rahman, I. Orikumhi, C. Y. Leow, and M. N. Hindia, “An Overview of Internet of Things (IoT) and Data Analytics in Agriculture: Benefits and Challenges,” IEEE Internet of Things Journal, vol. 4662, no. c, pp. 1–17, 2018.

[6] FAO, “Global agriculture towards 2050,” High Level Expert Forum-How to feed the world 2050. pp. 1–4, 2009.

[7] F. Rojo, E. Kizer, S. Upadhyaya, S. Ozmen, C. Ko-Madden, and Q. Zhang, “A Leaf Monitoring System for Continuous Measurement of Plant Water Status to Assist in Precision Irrigation in Grape and Almond crops,” IFAC-PapersOnLine, vol. 49, no. 16, pp. 209–215, 2016.

[8] P. Afrasiabikia, A. Parvaresh Rizi, and M. Javan, “Scenarios for improvement of water distribution in Doroodzan irrigation network based on hydraulic simulation,” Computers and Electronics in Agriculture, vol. 135, pp. 312–320, 2017.

[9] M. A. Uddin, A. Mansour, D. Le Jeune, and E. H. M. Aggoune, “Agriculture internet of things: AG-IoT,” in 2017 27th International Telecommunication Networks and Applications Conference, ITNAC 2017, vol. 2017-Janua, pp. 1–6, 2017.

[10] H. Navarro-Hellin, J. Martínez-del-Rincon, R. Domingo-Miguel, F. Soto-Valles, and R. Torres-Sánchez, “A decision support system for managing irrigation in agriculture,” Computers and Electronics in Agriculture, vol. 124, pp. 121–131, 2016.

[11] K. Zeinab and E. Ahmed, “Internet of Things Applications , Challenges and Related Future Technologies Internet of Things Applications , Challenges and Related Future Technologies,” vol. 67, no. February, pp. 126–148, 2017.

[12] X. Shi et. al., “State-of-the-art internet of things in protected agriculture,” Sensors (Switzerland), vol. 19, no. 8, 2019.

[13] A. Tzounis, N. Katsoulas, T. Bartzanas, and C. Kittas, “Internet of Things in agriculture, recent advances and future challenges,” Biosystems Engineering, vol. 164, pp. 31–48, 2017, doi: 10.1016/j.biosystemseng.2017.09.007.

[14] T. Ojha, S. Misra, and N. S. Raghuwanshi, “Wireless sensor networks for agriculture: The state-of-the-art in practice and future challenges,” Computers and Electronics in Agriculture, vol. 118, pp. 66–84, 2015.

[15] Y. Tikhamarine, A. Malik, A. Kumar, D. Souag-Gamane, and O. Kisi, “Estimation of monthly reference evapotranspiration using novel hybrid machine learning approaches,” Hydrological sciences journal, vol. 64, no. 15, pp. 1824–1842, Nov. 2019.

[16]] M. Adnan, M. Ahsan, A.-R. -, and M. Nazir, “Estimating Evapotranspiration using Machine Learning Techniques,” International Journal of Advanced Computer Science and Applications, vol. 8, no. 9, pp. 108–113, 2017.

[17] S. Pan et. al., “Evaluation of global terrestrial evapotranspiration using state-of-the-art approaches in remote

sensing, machine learning and land surface modeling," Hydrology and Earth System Sciences, vol. 24, no. 3, pp. 1485–1509, Mar. 2020.

[18] F. Granata, "Evapotranspiration evaluation models based on machine learning algorithms—A comparative study," Agricultural Water Management, vol. 217, pp. 303–315, May 2019.

[19] Z. Chen, Z. Zhu, H. Jiang, and S. Sun, "Estimating daily reference evapotranspiration based on limited meteorological data using deep learning and classical machine learning methods," Journal of Hydrology, vol. 591, p. 125286, Dec. 2020.

[20] L. B. Ferreira and F. F. da Cunha, "New approach to estimate daily reference evapotranspiration based on hourly temperature and relative humidity using machine learning and deep learning," Agricultural Water Management, vol. 234, p. 106113, May 2020.

[21] M. Gocić et. al., "Soft computing approaches for forecasting reference evapotranspiration," Computers and Electronics in Agriculture, vol. 113, pp. 164–173, 2015.

[22] Y. Feng, Y. Peng, N. Cui, D. Gong, and K. Zhang, "Modeling reference evapotranspiration using extreme learning machine and generalized regression neural network only with temperature data," Computers and Electronics in Agriculture, vol. 136, pp. 71–78, 2017.

[23] A. Fernández-López, D. Marín-Sánchez, G. García-Mateos, A. Ruiz-Canales, M. Ferrández-Villena-García, and J. M. Molina-Martínez, "A Machine Learning Method to Estimate Reference Evapotranspiration Using Soil Moisture Sensors," Applied Sciences 2020, Vol. 10, Page 1912, vol. 10, no. 6, p. 1912, Mar. 2020.

[24] M. K. Saggi and S. Jain, "Reference evapotranspiration estimation and modeling of the Punjab Northern India using deep learning," Computers and Electronics in Agriculture, vol. 156, pp. 387–398, Jan. 2019.

[25] "Chapter 2 - FAO Penman-Monteith equation." https://www.fao.org/3/x0490E/x0490e06.htm (accessed Dec. 15, 2021).

[26] "Chapter 3 - Meteorological data." https://www.fao.org/3/x0490E/x0490e07.htmTopOfPage(accessed Dec. 15, 2021)

[27] J. Doorenbos and W. O. Pruitt, "Guidelines for predicting crop water requirements," FAO Irrigation and Drainage Paper, vol. 24, p. 144, 1977.