2020 IEEE International Students' Conference on Electrical, Electronics and Computer Science

Machine Learning convergence for weather based crop selection

Sonal Jain
Department of Computer Science & Engineering
Indian Institute of Technology (ISM), Dhanbad
Dhanbad 826004, Jharkhand, India
Email: soni.shah014@gmail.com

Dharavath Ramesh*, Senior Member, IEEE
Department of Computer Science & Engineering
Indian Institute of Technology (ISM), Dhanbad
Dhanbad 826004, Jharkhand, India
Email: ramesh.d.in@ieee.org

2020 IEEE International Students' Conference on Electrical,Electronics and Computer Science (SCEECS) 978-1-7281-4862-5/20/$31.00 ©2020 IEEE 10.1109/SCEECS48394.2020.75

***Abstract*—Agriculture plays a vital role in Indian economy. It contributes 18% of total India's GDP. In India, most of the crops are solely dependent upon weather conditions. Hence, more yield of crops can be achieved by analysing agro-climate data using machine learning techniques. This paper proposes a crop selection method to maximize crop yield based on weather and soil parameters. It also suggests the proper sowing time for suitable crops using seasonal weather forecasting. Machine learning algorithms such as Recurrent neural network is used for weather prediction, and Random forest classification algorithm is used to select suitable crops. The result of proposed weather forecasting technique is compared with conventional Artificial neural network, which shows better performance results for each selected weather parameters.**

***Index Terms*—Agriculture, Recurrent neural network, Crop selection, Random forest, Artificial neural network, Weather prediction**

I. Introduction

Owing to increasing population, the demand for crop production is increasing. On the other hand, the available land area for agriculture is decreasing day by day. Moreover, one-third of the total food is wasted due to various factors such as climate change, selection of unsuitable crop, low fertile soil, pest, etc. [1]–[3]. Hence, techniques that provide maximum yield using optimal quantity of resources in limited land areas need to be adopted. This concept is widely known as precision agriculture (PA) [4]. Precision agriculture is a site-specific crop management technique that collects and analyzes the crop data, and based upon these analysis action such as application of appropriate amount of water, fertilizer and pesticides are taken.

In recent years, various data mining models are used in agriculture for prediction and decision making based upon weather and soil condition. Various such methods can be classified into two categories: First is statistics model such as multiple linear regression method, in which the structure of data need to be known or assumed in advance, and second is machine learning (ML) algorithms that map input data to output and learns from data itself. ML is used for solving various issues such as selection of suitable crops, weather prediction, crop disease prediction, crop yields prediction and developing automated irrigation system [5]. An estimation of crop sowing time based on available weather data is presented by Gümüşçü et al. [6] using ML techniques such as support vector machine (SVM), KNN and decision tree. Similarly, a machine learning based decision support system is presented by Navarro-Hellín et al. [7] to predict irrigation requirement for one week using two prediction techniques PLSR and ANFIS. A crop disease prediction system is presented by Patil et al. [8] for early disease detection in grapes using sensors. ML technique hidden Markov model (HMM) is applied to predict the disease based on leaf wetness, relative humidity and moisture. A detailed review of machine learning approaches for crop yield prediction and nitrogen state estimation is presented by Chlingaryan [9]. In a work presented by Dimitriadis [10], a plant-driven crop management technique is given using machine learning to check health status of plants and its water requirement. An impact of climate change in crop yield is presented in [11]. Hence, weather prediction plays a very important role in agriculture as to decide many activities such as crop irrigation, applying fertilizer, crop yield prediction, pest and disease control, etc. Various machine learning techniques such as decision tree, support vector machine, naïve Bayes and artificial neural network (ANN) are used for weather prediction [12]. A feature-based weather forecasting model using ANN is presented by Paras et al. [13]. In this manuscript, recurrent neural network (RNN) which is a variant of ANN is used for weather prediction.

Crop yield rate depends upon various parameters such as the geography of area, soil type, soil nutrients, soil pH value, weather condition, etc. The combination of these parameters can be used for selection of suitable crops for a farm or land to gain maximum yield. In this manuscript, soil and weather parameters such as soil type, soil fertility, soil pH, maximum temperature, minimum temperature, rainfall are used to identify suitable crops for specified farm or land. Some factors that influence crop selection in agriculture are given below [14]:

1) *Weather condition*: Generally, most of the crops are solely dependent on weather conditions such as maximum and minimum temperature, rainfall, sunshine hours, relative humidity, etc. Hence, suitable crops need to be selected by considering the field weather condi-

tions.

2) *Soil condition*: Most of the crops grow better in one soil and give maximum yield based on soil type, soil drainage capacity, soil pH and soil fertility. Hence, crops that gives maximum yield in chosen land soil need to be planted.
3) *Cropping system*: Cropping system such as crop rotation plays an important role in enriching the soil, restore soil nutrients, prevent soil erosion and mitigates the build-up of pathogens. Hence, crop selection should to be done based on previous crop planted or other cropping systems such as mixed cropping, relay cropping and inter-cropping should be adopted.
4) *Water*: Crop selection can also be done based on water facilities available, some crop needs more water than others. Hence, if rainfall is not evenly distributed, proper irrigation facilities are required for better crop yield.

In this manuscript, the agro-climatic zones of Telangana state and crops planted on these zones are considered to train the model. However, the same model can be extended to different agro-climatic zones and crops. The organization of the remaining paper is as follows. Section 2 discusses work related to crop selection and crop yield prediction, section 3 explains dataset and the proposed methodology of weather prediction based crop selection. The outcome of weather prediction and classification algorithms are presented in section 4. Finally, section 5 concludes the proposed work and specifies the future direction.

II. Related Work

Numerous research has been performed in this area, which includes weather and soil data for suitable crop identification and crop yield prediction. A stochastic seasonal weather forecasting model is presented by Apipattanavis et al. [15], which is integrated with crop simulation model to predict expected yield and economic return. The model can also help to identify production risk associated with dry and wet climate. The implementation of this model is done for Maize crop in two regions of Argentina. An ensemble seasonal forecasting model is implemented in Europe to predict crop yield and to plan crop management activities beforehand [16]. The model is employed for wheat crop and shows better prediction accuracy than given by the operational system. It also uses probability distribution function to shows the reliability of prediction for a particular region. In [17], a crop yield prediction model using support vector regression is presented. First, the model predicts rainfall using modular artificial neural network (MANN) and then based upon amount of rainfall and some other features, the crop yield prediction is done for rice, maize, ragi and bajra crop in Visakhapatnam. A soil classification method is presented by Paul et al. [18], which classifies soil into three categories: low, medium and high based upon its nutrient content . Soil parameters such as pH, organic carbon, nitrogen, phosphorus, potassium and some other micro-nutrients are used for classification using KNN. The class of soil helps in identifying suitable crops for soil that gives maximum yield.

A crop selection method to maximize the production yield is presented in [19]. The proposed work also suggests sequence of crop planted to gain maximum yield over a year. The crops are classified into four categories based on their growing time such as: 1) seasonal crops 2) whole year crops 3) short time plantation crops and 4) long time plantation crops. The selection of suitable crop is done based on weather, soil type, crop type and water density parameters. After crop selection, it suggests a crop sequencing that gives maximum yield based on crop, sowing time, plantation day and predicted yield. Such a crop selection method is also presented by Tseng et al. [20] uses IoT to collect weather and soil data such as temperature, humidity, soil moisture, soil electrical conductivity, soil salinity, etc. from one or more farm. A 3D clustering is performed on collected farm data to analyze farming techniques used by farmer for a crop.

In literature, classification and prediction techniques such as decision tree, naïve Bayes, random forest are used for selection of suitable crops based upon weather and soil parameters. An ensemble technique using naïve Bayes, CHAID, random forest and K-nearest neighbour is presented in [21] which uses classification parameters such as root depth, texture, soil colour, drainage etc. However, this method is able to select only one crop while more than one crop can be suitable for specified land and weather. A naïve Bayes classification technique is presented in [22] to give suitability of crops such as chilli, rice, maize and cotton using weather parameters like rainfall, soil moisture, temperature and atmospheric pressure. This method also recommends the suitable time for harvesting and sowing of a particular crop.

The research gap of previous work exhibited in crop selection is, either the method suggests only one crop for land or it does not consider the agriculture facilities available for farmers such as a rice crop can be grown in rabi season if appropriate irrigation facilities are available. The contribution of the proposed work is to provide an integration of seasonal weather prediction with crop classification method to suggest suitable crops based on weather and to advise crop sowing time using predicted weather parameters. The proposed method also suggests agriculture facilities required for a particular crop.

III. Proposed Methodology

The proposed methodology is divided into two phases: 1) seasonal weather prediction and 2) identification of suitable crop. It uses recurrent neural network (RNN) for seasonal weather forecasting. Subsequently, the classification of suitable crops is done via random forest classification algorithm. The schematic flow of proposed mechanism is demonstrated in Fig. 1.

A. Seasonal weather prediction

The dataset used for weather prediction is collected from NRSA Hyderabad station, which consists of five years data of weather parameters such as temperature, humidity, sun hours, wind speed, wind direction, etc. The data are collected from

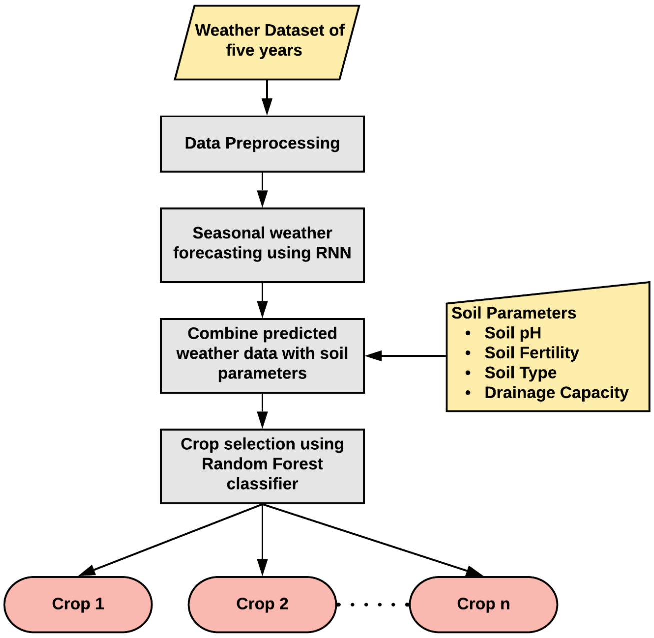

Fig. 1. Schematic flow of proposed weather forecasting based crop selection method

January 2014 to September 2019 at one hour resolution. For training the prediction model, initially data pre-processing is carried out which comprises filling missing values, data transformation and data normalization. The missing values of data are filled using linear interpolation method. Following this, hourly data is converted into daily average data consist of minimum temperature, maximum temperature calculated from temperature data. Subsequently, min-max normalization [23] is performed on dataset to convert it into a common scale. For weather prediction, recurrent neural network (RNN) is used which is a variant of neural network (NN) intended to handle sequence dependence. The advantage of RNN over conventional NN is, the output from previous step is forwarded to current step, hence the hidden state in RNN remembers the sequence while in conventional neural network all the values are treated independent of each other. The basic architecture of RNN is depicted in Fig. 2 consisting of an input layer, two hidden layers and an output layer. In this paper, long short-term memory (LSTM) recurrent neural network with feedback connection is used. The single cell of LSTM network is shown in Fig. 3.

The cell of LSTM consists of three gates: 1) input gate i 2) forget gate f and 3) output gate o. Whereas, g is an updated input at current time stamp t. The value of all these gates depend upon previous hidden state h_{i-1} and current input x_t, which can be given as: [24], [25]:

$$i_t = f(W_i x_t + U_i h_{t-1} + b_i) \quad (1)$$

$$f_t = f(W_f x_t + U_f h_{t-1} + b_f) \quad (2)$$

$$g_t = tanh(W_c x_t + U_c h_{t-1} + b_c) \quad (3)$$

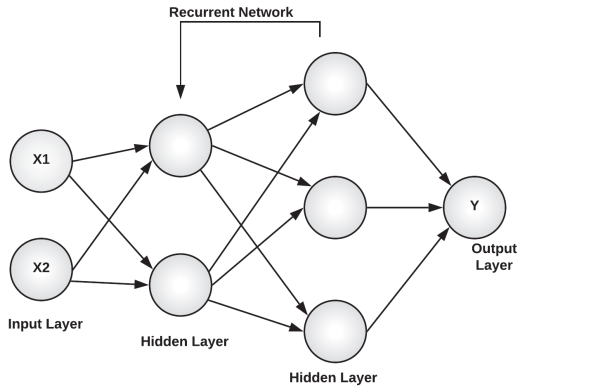

Fig. 2. Basic architecture of RNN

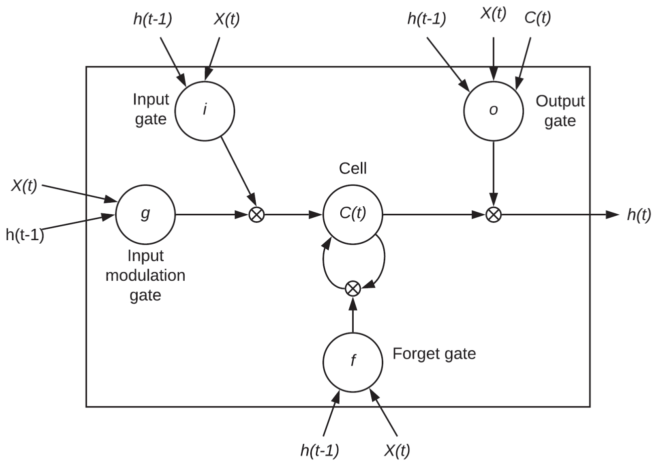

Fig. 3. The single cell of LSTM

The updated cell value is calculated as:

$$c_t = i_t.g_t + f_t.c_{t-1} \quad (4)$$

The output gate value comprises of current cell state, previous output and current input value.

$$o_t = f(W_o.x_t + U_o.h_{t-1} + V_o.c_t + b_o) \quad (5)$$

$$h_t = o_t.tanh(c_t) \quad (6)$$

Three different RNN models are trained one for each minimum temperature, maximum temperature and rainfall prediction. For estimating maximum temperature, the first three year dataset is used for training the model and last two year dataset is used for testing the prediction accuracy. The data can be represented as a vector series $\{x(1)\ldots, x(n)\}$ and for seasonal forecasting or 90 days ahead forecasting data is converted into matrix of size NXM of each row consists of one input feature and 90 output or target values is represented by:

$$\{x(t), x(t+1), x(t+2)\ldots\ldots\ldots x(t+90)\} \quad (7)$$

Here, n is the length of time-series data, size of N is $n-90$ and size of M is 91. The LSTM model used for prediction consists of one input layer with one value, one hidden layer consists of 4 neurons or LSTM cell and an output layer that predicts 90 values, is trained. Similarly, prediction is carried out for minimum temperature and rainfall. The prediction outcome of RNN is shown in results section. The prediction results generated using RNN are also compared with conventional ANN, which indicates better prediction accuracy than ANN.

B. Selection of suitable crop

In this manuscript, the crop and soil data of Telangana state are considered for analysis. Telangana state is divided into three agro-climatic zones: 1) north Telangana zone, 2) central Telangana zone and 3) southern Telangana zone. The major crops grown in these areas are rice, cotton, chilli, soybean and maize with 50.54 lakhs ha of total farm holding [26]. Northern zone has red soil, shallow black soil and deep calcareous soil. The central Telangana zone consists of mainly red soil and calcareous soil. Southern zone contains red soil with different texture, and alluvial soil and calcareous soil are also found in some regions [27].

Crop selection model takes input of soil parameters such as soil type, soil pH, soil fertility and its water holding capacity. The soil and predicted weather parameters are used collectively to choose suitable crops for land. The crop selection can be done either annual or seasonal. In seasonal crop selection method, the model suggests one or more crop that is suitable for a season and also suggests crop requirement such as irrigation, and the proper sowing time of a particular crop based on predicted weather. The sample dataset used for crop selection model is shown in Table I:

The dataset consists of 10 crops which are rice, cotton, maize, sunflower, castor, chilli, redgram, greegram, jowar and soyabean. However, the model can be extended for different type of land and crops. A random forest classification method [28] is used to classify crops based on weather and soil parameters. A random forest classifier is an ensemble classifier which uses collection of decision trees created from subset of training dataset. The outcome of random forest is decided by aggregating output of each decision tree. The class having maximum votes after aggregation becomes the outcome of random forest. However, to select more then one crop suitable for a land, a slight modification is done in the algorithm. Such that, if more than threshold value of decision trees in random forest gives the same crop as classification, the crop is added into list of suitable crops. The threshold value th can be defined as:

$$th = 2 * \left(\frac{Total\ number\ of\ trees\ in\ random\ forest}{Total\ number\ of\ classes} \right) \tag{8}$$

The sample outcome of classification model is depicted in Table III.

IV. Experiment Results

The study consist of 1987 weather data points of five years from January 2014 to September 2019. Prediction is done for maximum temperature, minimum temperature and rainfall. The seasonal prediction results for maximum temperature, minimum temperature and rainfall are shown in Fig. 4, 5 and 6 respectively.

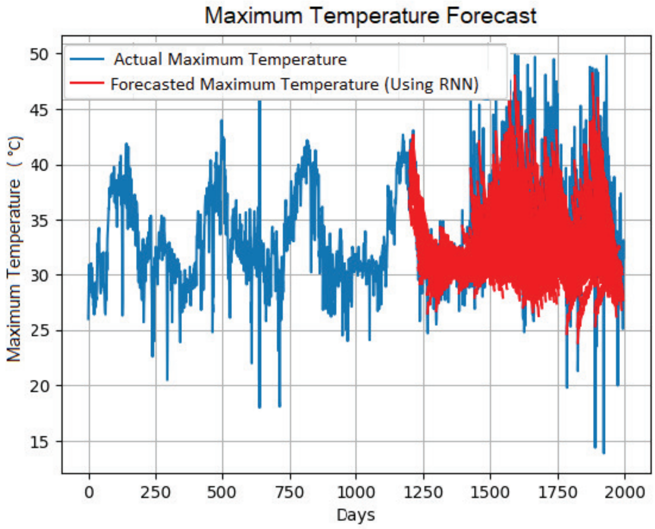

Fig. 4. Maximum Temperature forecasting for 90 days

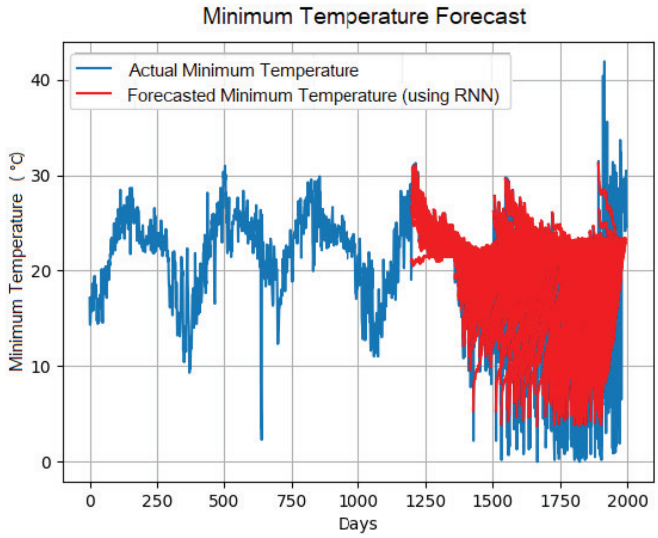

Fig. 5. Minimum Temperature forecasting for 90 days

From prediction results, it is noted that RNN able to predict time-series data more accurately with an acceptable error. The performance of RNN model is compared with ANN using Root Mean Square Error (RMSE) and comparison results are shown in Table II, which depicts better prediction accuracy of RNN than conventional ANN. The comparison graph of RMSE between ANN and RNN for each day ahead forecasting

TABLE I
SAMPLE DATASET FOR TRAINING CLASSIFICATION MODEL

Minimum Temperature (°C)	Maximum Temperature (°C)	Rainfall (mm)	Soil Type	Soil pH	Water Capacity	Soil Fertility	Crop
24	31	550	Red	6.0	Well drained	Medium	Castor
25	35	754	Loamy	7.0	Well drained	High	Chilli
22	33	830	Black	6.5	Well drained	High	Cotton
18	34	895	Alluvial	6.0	Water logged	High	Rice
19	36	830	Calcareous	7.0	Well drained	Low	Jowar

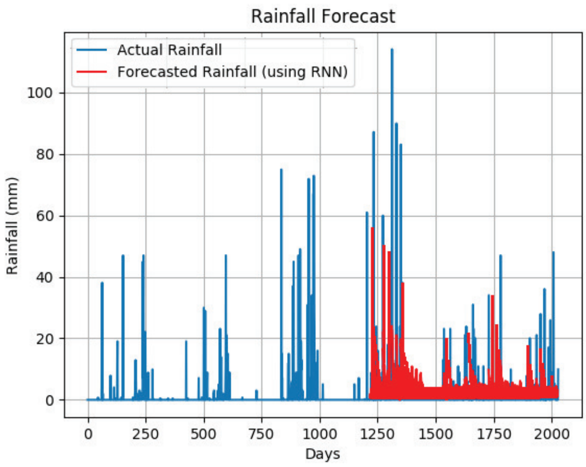

Fig. 6. Rainfall forecasting for 90 days

TABLE II
COMPARISON TABLE OF RNN AND ANN

Algorithm	Average RMSE (%)		
	Maximum Temperature	***Minimum Temperature***	***Rainfall***
ANN	6.015	12.20	9.20
RNN	5.002	7.12	8.17

is shown in Fig. 7, 8, 9; Which depicts with increasing forecasting days the error in forecasting also increases.

Following weather prediction, the crop selection is done using random forest classifier. The suggested crops for rabi season by random forest classifier based on predicted weather parameters and soil parameters of land are given in Table III. The proposed method also gives proper sowing time for crops and agriculture facilities required. From Table III, it can be noted that rice and maize crops grown in rabi season require irrigation facility while red gram crop can be dependent on rainfall. The sowing time for rice and maize crop is given as third week of December and November respectively and similarly, the sowing time for red gram is given as first week of October.

V. CONCLUSION AND FUTURE WORK

The proposed method represents a novel weather based crop selection system to select crops for land, based on predicted weather parameters and its soil parameters. The

TABLE III
SUGGESTED CROPS FOR RABI SEASON

Suggested Crop	Water Requirement	Sowing Time
Rice	Irrigated	Third week of December
Maize	Irrigated	Third week of November
Red Gram	Rainfed	First week of October

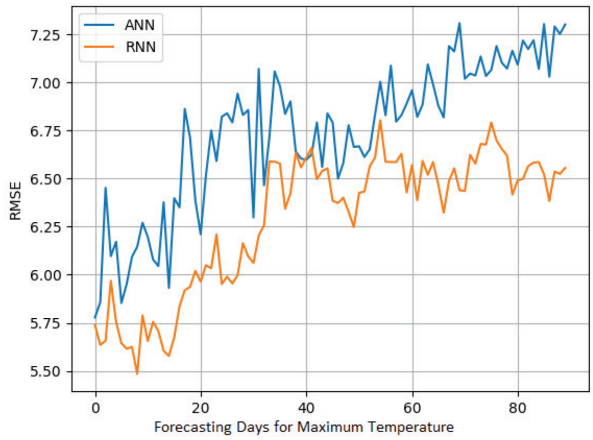

Fig. 7. RMSE for Maximum temperature day ahead forecasting

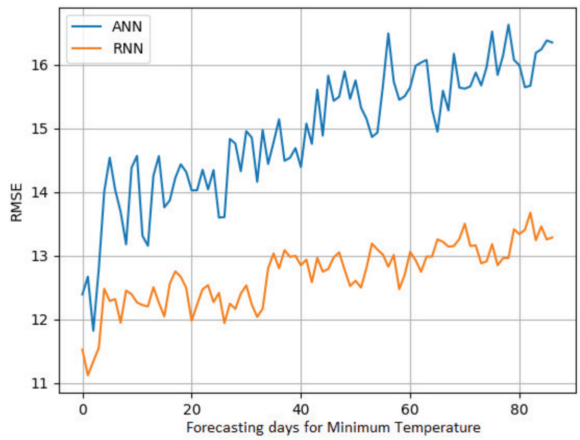

Fig. 8. RMSE for Minimum temperature day ahead forecasting

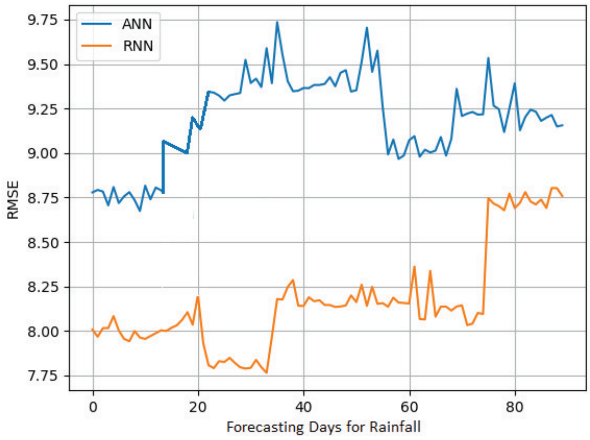

Fig. 9. RMSE for Rainfall day ahead forecasting

classification model is trained by considering the weather and soil parameters of agro-climatic zones of Telangana state. The seasonal weather forecasting is carried out using RNN. The prediction results are also compared with conventional ANN, which shows better prediction accuracy. A random forest classifier with additional threshold parameter is used to get more than one suitable crop for a land. Moreover, the proposed method also suggests the proper sowing time for each crop based upon predicted weather parameters.

The future work is to improve the efficiency of algorithm by using IoT devices to collect precise weather and soil data of a farm. The soil parameters such as the ratio of N-P-K nutrient in soil and soil temperature can be considered for better accuracy of the crop selection system.

Acknowledgements

This work is supported by Science and Research Engineering Board (Department of Science & Technology), Govt. of India under Grant no. EEQ/2018/000108. We also acknowledge NRSA Hyderabad station for providing the required dataset for experiment.

References

[1] L. Madhusudhan, "Agriculture role on indian economy," *Business and Economics Journal*, vol. 6, no. 4, p. 1, 2015.

[2] V. Kumar, V. Dave, R. Bhadauriya, and S. Chaudhary, "Krishimantra: agricultural recommendation system," in *Proceedings of the 3rd ACM Symposium on Computing for Development*. ACM, 2013, p. 45.

[3] "Key facts on food loss and waste you should know!" http://www.fao.org/save-food/resources/keyfindings/en/.

[4] A. Srinivasan, *Handbook of precision agriculture: principles and applications*. CRC press, 2006.

[5] K. Kaur, "Machine learning: applications in indian agriculture," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 4, pp. 342–344, 2016.

[6] A. Gümüşçü, M. E. Tenekeci, and A. V. Bilgili, "Estimation of wheat planting date using machine learning algorithms based on available climate data," *Sustainable Computing: Informatics and Systems*, 2019.

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

[8] S. S. Patil and S. A. Thorat, "Early detection of grapes diseases using machine learning and iot," in *2016 Second International Conference on Cognitive Computing and Information Processing (CCIP)*. IEEE, 2016, pp. 1–5.

[9] A. Chlingaryan, S. Sukkarieh, and B. Whelan, "Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review," *Computers and electronics in agriculture*, vol. 151, pp. 61–69, 2018.

[10] S. Dimitriadis and C. Goumopoulos, "Applying machine learning to extract new knowledge in precision agriculture applications," in *2008 Panhellenic Conference on Informatics*. IEEE, 2008, pp. 100–104.

[11] Y. Kang, S. Khan, and X. Ma, "Climate change impacts on crop yield, crop water productivity and food security–a review," *Progress in natural Science*, vol. 19, no. 12, pp. 1665–1674, 2009.

[12] D. Chauhan and J. Thakur, "Data mining techniques for weather prediction: A review," *International Journal on Recent and Innovation Trends in Computing and Communication*, vol. 2, no. 8, pp. 2184–2189, 2014.

[13] S. M. Paras, A. Kumar, and M. Chandra, "A feature based neural network model for weather forecasting," *International Journal of Computational Intelligence*, vol. 4, no. 3, pp. 209–216, 2009.

[14] "Critical factors to be considered for selection of crops," http://vikaspedia.in/agriculture/crop-production/critical-factors-to-be-considered-for-selection-of-crops.

[15] S. Apipattanavis, F. Bert, G. Podestá, and B. Rajagopalan, "Linking weather generators and crop models for assessment of climate forecast outcomes," *Agricultural and forest meteorology*, vol. 150, no. 2, pp. 166–174, 2010.

[16] P. Cantelaube and J.-M. Terres, "Seasonal weather forecasts for crop yield modelling in europe," *Tellus A: Dynamic Meteorology and Oceanography*, vol. 57, no. 3, pp. 476–487, 2005.

[17] E. Khosla, R. Dharavath, and R. Priya, "Crop yield prediction using aggregated rainfall-based modular artificial neural networks and support vector regression," *Environment, Development and Sustainability*, pp. 1–22, 2019.

[18] M. Paul, S. K. Vishwakarma, and A. Verma, "Analysis of soil behaviour and prediction of crop yield using data mining approach," in *2015 International Conference on Computational Intelligence and Communication Networks (CICN)*. IEEE, 2015, pp. 766–771.

[19] R. Kumar, M. Singh, P. Kumar, and J. Singh, "Crop selection method to maximize crop yield rate using machine learning technique," in *2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*. IEEE, 2015, pp. 138–145.

[20] F.-H. Tseng, H.-H. Cho, and H.-T. Wu, "Applying big data for intelligent agriculture-based crop selection analysis," *IEEE Access*, vol. 7, pp. 116 965–116 974, 2019.

[21] S. Pudumalar, E. Ramanujam, R. H. Rajashree, C. Kavya, T. Kiruthika, and J. Nisha, "Crop recommendation system for precision agriculture," in *2016 Eighth International Conference on Advanced Computing (ICoAC)*. IEEE, 2017, pp. 32–36.

[22] R. Priya, D. Ramesh, and E. Khosla, "Crop prediction on the region belts of india: A naïve bayes mapreduce precision agricultural model," in *2018 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, 2018, pp. 99–104.

[23] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.

[24] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

[25] J. J. Hopfield and D. W. Tank, ""neural" computation of decisions in optimization problems," *Biological cybernetics*, vol. 52, no. 3, pp. 141–152, 1985.

[26] [Online]. Available: https://niti.gov.in/writereaddata/files/Telangana_Presentation_1.pdf

[27] Agriculture in telangana. [Online]. Available: https://pjtsau.edu.in/index.html

[28] Tin Kam Ho, "Random decision forests," in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol. 1, Aug 1995, pp. 278–282 vol.1.