IoT and machine learning based framework for Precision Agriculture

Mukesh Bisht[1], Sagar Rawat[2], Dhajvir Singh Rai[3]
[1]*Computer Science, Dev Bhoomi Institute of Technology, Dehradun India*
[2]*Computer Science, Dev Bhoomi Institute of Technology, Dehradun India*
[3] *Assistant Professor Computer Science, Dev Bhoomi Institute of Technology, Dehradun India*

Abstract—As the global population is increasing exponentially, so is the requirement of food. Also, countries like India whose 1/3rd capital came from farming needs to improve the methods of farming. We propose an IoT based system, which collects various physical factors in real time such as temperature, humidity, soil moisture etc. This data can be studied to get analytical insights. We try to apply various state of the art machine learning techniques on this collected data to predict the crop suitable for given factors.

Keywords—*IoT, Machine Learning, Classification, Agriculture, Artificial Neural Network, Crop Selection.*

I. INTRODUCTION

Agriculture is the main source of food for human species. As the population is increasing the sources and traditional approaches to farming needs to be upgraded. Growth in agricultural sector is necessary for the development of economic condition of the country. In India, most of the farmers produce traditional crops (those which are produced by their ancestors). But for a similar physical condition, there may be multiple possible crops that can be produced on the same land, which may results in greater profit.

IoT and ML have been emerged with high performance computing which makes it feasible to collect data and analyses it. It can be applied in various domains such as agriculture, medical etc. to build smart systems. ML techniques can be used to develop predictive model.

ANN [1] biologically-inspired programming paradigm which enables a computer to learn from observational data, ANNs are very good in learning patterns from nonlinear data.

Many researches are done in the field of agriculture which signifies the use of sensors to collect data. The collected data can be analyzed to extract some useful statistics. Also, ML can be used to develop a model to predict various Factors such as suitable crop, water requirements, estimated cost etc. which can improve the way the farming is done.

In this paper, we'll only consider the problem of predicting suitable crop for given physical conditions. Artificial Neural Network is used for classification of crop.

A. Factors Affecting Crop Production

1. Internal or Genetic factor.
2. External or Environmental factor.

In this paper, we ignore all the internal factors. We only consider climatic factors and edaphic factors (soil factors) which come under external factors. Nearly 50% of yield is attributed to influence of climatic factors affect the productivity of crop. [2]

Features Collected

1. Humidity
2. Soil Moisture
3. Soil Type
4. Month
5. Sun light
6. Weather condition
7. Temperature
8. Altitude
9. Atmospheric pressure

II. RELATED WORK

A. Agriculture Field Monitoring

Instead of observing the farms all the time, this paper proposes the design to monitor the same attributes using wireless sensor network then the farmer may also acknowledged via SMS by an expert.[3]

B. Environment Monitoring

The paper proposed a system that collects various environmental factors like temperature, humidity, illumination, voltage etc. from greenhouse and from there it transmits the data to nearest server via GPRS. The system includes a web application which can show the greenhouse status using maps. [4]

C. Farm Automation

Number of operations of farm can be automated like irrigation system, temperature controlled system for livestock and farm product.[5] this paper presents automatic house temperature control, lighting system control, automatic sprinkler system and security in farm houses. [6]

D. Crop yield Prediction

This paper present forecasting methods for evaluating crop yield estimates using time series models to predict crop yields. [7]

This paper extends the IoT based monitoring systems with machine learning algorithms to make predictions what crops are suitable for given environment.

III. Proposed Work

The IoT device collects data with sensor array and sends it to the server, an associative model use this data to estimate the models and "future" data for prediction. The system consists of two major components:

1. Collecting environmental data,
2. Applying machine learning techniques on collected data.

The first module consists of following:

- Sensors to collect data i.e. soil moisture sensor, humidity sensor, temperature sensor, barometric sensor, Altitude and luminosity sensor.
- Wi-Fi module for connectivity.
- Microcontroller as central processor.
- Power supply | Solar Panel.

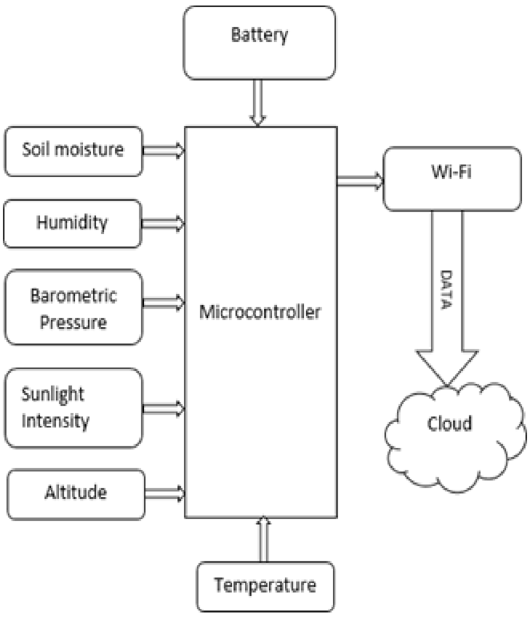

Figure 1: Architecture of the device.

The architecture of the device can be understood with the help of above diagram.

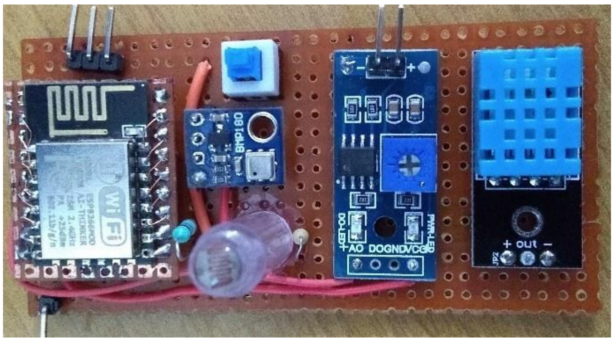

Figure 2: Embedded Circuit with Sensors.

The second module consists of:

- Data pre-processor: Data pre- processing is an integral step In ML as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn.
- ML model to make predictions

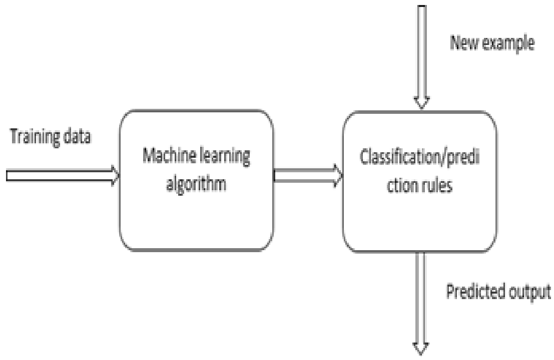

Figure 3: Model based learning.

The model is trained by considering 9 features, the feature vector is given to the model to predict suitable crop

IV. Method Overview

The objective of the proposed work is to predict the crop label using ML techniques, this includes following phases:

A. Dataset

The data is collected from 50 distinct randomly selected fields of selected crops in Dehradun district.

Input data consist of 10 parameters namely: season, soil type, humidity, altitude, weather condition, sunlight, atmospheric pressure and the crop label (ground truth). For simplicity we have only considered 3 agricultural crops: paddy, wheat and sugar-cane. The dataset contains 1/3rd data point for each individual crop. The dataset is divided into two groups: 75% for training the model which includes validation set as well and 25% for testing the model. The main purpose of using the testing dataset is to test the generalization ability of a trained model. Each crop class contains equal number of data points in both training and test dataset to avoid imbalance. please do not alter them. You may note peculiarities. For example, the head margin in this template measures proportionately more than is customary. This measurement and others are deliberate, using specifications that anticipate your paper as one part of the entire proceedings, and not as an independent document. Please do not revise any of the current designations.

1. Sample:

TABLE I.

Feature	Type	Range
Temperature	Numeric	-100-200
Weather	Categorical	clear, clouds, scattered, clouds, broken clouds, shower, rain,
Month	Categorical	Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct Nov, Dec
Soil Type	Categorical	Loam, Clay, Sand, Silt, Peat, Chalk
Humidity	Numeric	0-100
Soil Moisture	Numeric	0-100
LUX	Numeric	> 0
Pressure	Numeric	>0
Altitude	Numeric	c

2. Preprocessing:

humidity	Soil moisture	Soil type	sunlight	Weather condition	Temp (°C)	altitude	pressure	month	Crop label
0.74	0.08	alluvial	365.81	sunny	24	655	1007	November	wheat
0.41	0.41	loam	612.56	sunny	35	667	993	June	paddy
0.91	0.78	clay-loam	599.98	passing clouds	26	667	996	June	paddy
0.67	0.46	alluvium	654.89	sunny	11	657	1015	February	sugar-cane
0.49	0.69	loam	587.90	Passing clouds	34	671	996	June	paddy
0.62	0.45	alluvium	600.54	sunny	21	657	1015	march	sugar-cane
0.32	0.21	alluvium	488.34	partly sunny	17	657	1015	February	sugar-cane

Figure 4.

The data preprocessing is the technique of presenting the data into the required format for the better understanding and the better results but it may differ from situation to situation and also in terms of data to data. Different data requires different kinds of approach for the pre-processing such as for numerical data the preprocessing techniques can be feature scaling, whereas in case of the categorical data the preprocessing techniques varies such as label encoding, one hot encoding etc.

In this dataset, three types of pre-processing techniques have been used such as:

1. Feature Scaling since the range of values of raw data varies widely, the data is normalized for better accuracy of classifier. Numeric features are mean normalized,

$$x' = \frac{x - \text{average}(x)}{\max(x) - \min(x)}$$

Mean normalization

2. **Label Encoding:** Label Encoding is used to convert categorical data into numerical data. Each category is encoded to some unique integer value. For e.g. male or female can be encoded to 0 or 1 respectively.

3. **One Hot Encoding:** One hot encoding is used to binarize categorical data by creating dummy variable. As this form of representation is better understandable by the ML algorithms.

3. Artificial Neural Network

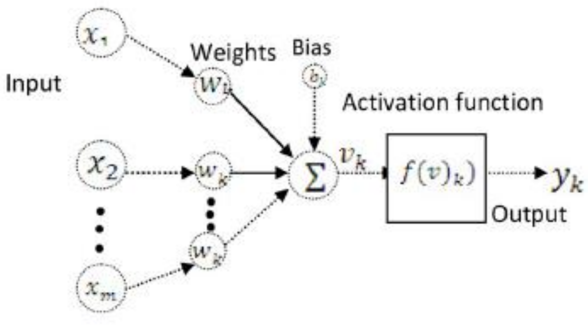

Figure 5

ANN consists of following major parts:

1. Input Layer: It contains nodes which corresponds to different features, the input feature vector R9
2. Hidden Layer: Only 1 hidden layer is used. It has 10 neurons.
3. Output Layer: R3 independent probabilities for corresponding crops for considered crops.

Activation Function: In hidden layers ReLu (Rectified Linear Units) is used. The biggest advantage of ReLu is indeed non-saturation of its gradient, which greatly accelerates the convergence of stochastic gradient descent compared to the sigmoid

ReLu:

$$f(x) = max(0, x) = \begin{cases} x_i, & if\ x_i \geq 0 \\ 0, & if\ x_i < 0 \end{cases}$$

Sigmoid:

$$f(x) = \left(\frac{1}{(1 + exp^{-x})} \right).$$

In output layer, sigmoid is used because its output is between 0 to 1, which is interpreted as independent probabilities of ground truth classes. [8]

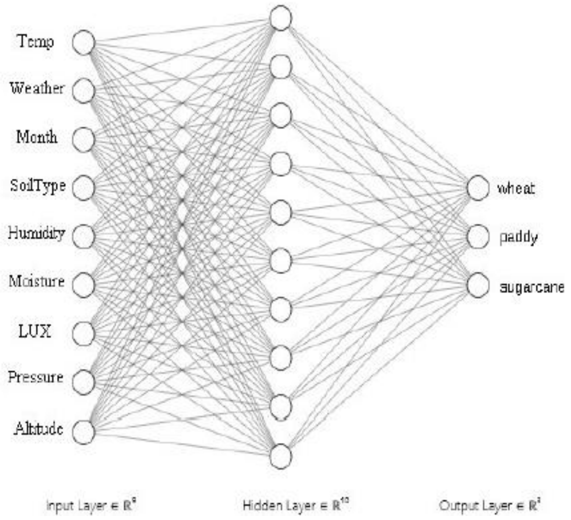

Figure 6

We trained our model using stochastic gradient descent [9] with a batch size of 128 examples. As the training dataset is small, we start with a large learning rate of 0.1 then gradually decrease the rate to 0.01 [10]. The epoch size is set to 10.

The hyper parameters of the model is tuned using k fold validation with k = 3. After tuning, the model is then trained on the whole training dataset which gets around 13% loss on test dataset.

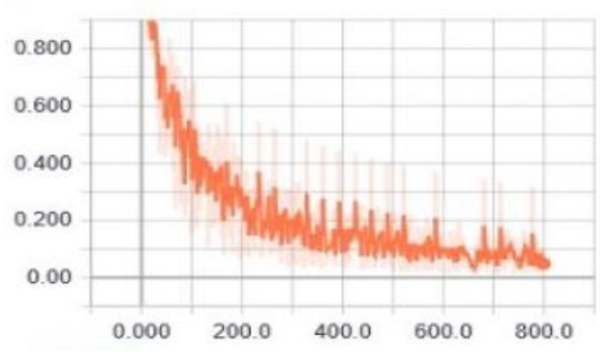

Figure 7: Learning Curve.

x axis of this curve represents steps and y axis represents error rate.

V. CONCLUSION

In India, the average income of a farmer is estimated at Rs. 77,976 per year, according to the Dalwai committee report [Jan 30, 2018]. Modern techniques can be used to improve farming industry. IoT and ML based solutions could help in proper utilization of resources.

The selection of crop is a tedious and time consuming task for a farmer. Farmer usually select crop according to their traditional knowledge and past experiences. The proposed system predicts all the suitable crops. Thus, it eliminates the human biases.

This study shows that IoT devices can be used to collect data and ML algorithms can be applied to get analytical insights. In practice, there can be lot of other practical factors which we did not include, availability of resources, cropping system options, market demand and availability, Govt. Policies etc.

VI. FUTURE WORK

Many different adaptations, tests, and experiments have been left for the future due to lack of time (i.e. the experiments with real data are usually very time consuming, requiring even days to finish a single run).

Collecting correct data is a time consuming task. So, we only apply these methods on experimental dataset. The quality of the results can be further improved by collecting more data. Also, other type of sensors could be used such as electrochemical sensors can be used to provide real time monitoring of soil nutrient data.

The following ideas could be tested:

- Market analysis could be done on predicted crops, to do risk analysis, which is also an important factor during selection of crops.
- Effect of climatic condition on certain crops could be analyzed.

References

[1] Ms. Sonali, B. Maind, Ms. Priyanka Wankar "Basic of Artificial Neural Network" Department of Information Technology Datta Meghe Institute of Engineering, Technology & Research.

[2] Dr. HR Agalave "Effect of environmental factors on productivity of crop" Principal & Head, Dept. of Botany, S.A.S. Mahavidyalaya, Mukhed, Nanded, Maharashtra, India

[3] B. BalajiBhanu, K. RaghavRao, J. V. N. Ramesh, Mohammed Ali Hussain, "Agriculture Field Monitoring and Analysis using Wireless Sensor Networks for improving Crop Production", IEEE, 2014

[4] Jianfa Xia, Zhenzhou Tang, Xiaoqiu Shi, Lei Fan, Huaizhong Li,"An environment monitoring for precise agriculture, based on wireless sensors network", IEEE,2011.

[5] Shining Li, Jin Cui, ZhigangLi,"Wireless Sensors Network forPrecise Agriculture Monitoring", 2011, China.

[6] DrishtiKanjilal, Divyata Singh, Rakhi Reddy, Prof Jimmy Mathew,"Smart Farm: Extending Automation to the Farm ", IJSTR, 2014.

[7] Askar Choudhury, James Jones, "Crop yield prediction using time series models " Illinois State University 2014

[8] Cheng Guo and Felix Berkhahn "Entity Embeddings of Categorical Variables" Neokami Inc.

[9] Chigozie Enyinna Nwankpa, Winifred Ijomah, Anthony Gachagan, and Stephen Marshall "Activation Functions: Comparison of Trends in Practice and Research for Deep Learning.

[10] Large-Scale Machine Learning with Stochastic Gradient Descent L´eon Bottou NEC Labs America, Princeton NJ 08542, USA

[11] An Empirical Study of Learning Rates in Deep Neural Network for Speech Recognition, Andrew Senior, Georg Heigold, Marc'Aurelio Ranzato, Ke Yang Google Inc., New York