ijRASET

INTERNATIONAL JOURNAL
FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

www.ijraset.com

Call: ⓒ08813907089 | E-mail ID: ijraset@gmail.com

Raita Mitra for Crop and Pesticide Recommendation along with Disease Prediction using ML

Apoorva G O[1], Spoorthi M[2]

[1, 2]*Department of Computer Science and Engineering, SJMIT, Davanagere*

Abstract: Agriculture seems to be a key part of both a country's food security and its economic growth. Choosing which crops to grow is one of the most important parts of planning agriculture. The suggested system helps farmers choose crops that will do well in their area. For agriculture to grow, it's important to be able to make accurate predictions about which crops to grow. We've given you a machine-learning method called "Random Forests" that can predict how crop choices will change based on the current climate and biophysical changes. We have gathered a lot of information about crop selection from many different places. These numbers are used both to train and test the model. From different results, it's clear that RF is a good machine-learning algorithm for predicting crops in their current state and has a very high level of accuracy when analysing data. RF algorithm also helps to find the right fertiliser by taking into account NPK values, soil moisture, and the name of the crop. Since a long time ago, plant leaf disease has been one of the biggest threats to food security because it lowers crop yield and lowers the quality of the crop. Accurately diagnosing diseases has been a big problem, but recent advances in computer vision made possible by deep learning have made it possible to use a camera to help diagnose diseases in plat leaf. It talks about the new way to find diseases and use deep learning and convolutional neural networks. neutral networks (CNNs) has done a great job of putting plant leaf diseases into groups. Using an image dataset of plant diseases that was available to the public, a CNN was used to apply and train a number of neuron-wise and layer-wise visualisation methods. So, it was found that neutral networks can pick up on the colours and textures of lesions that are unique to each disease. This can be compared to how humans make decisions.

Keywords: Convolution Neural Network (CNN), NPK and Random Forest (RF)

I. INTRODUCTION

For a nation, one of the most important aspects of its growth revolves around its potential to produce food. For generations, the production of essential food crops has been correlated with agriculture. In reality, however, the rapid pace of population growth has, by far, been the single biggest preoccupation of our society. In doing so, the scope of agriculture has been greatly undermined, particularly in terms of land use and fertility. Given that the area of land under cultivation in this era of urbanization and globalization is unlikely to increase, the focus will have to be on making the most of what there is. In agriculture, crop cultivar prediction is a key factor. Although recent research has opened up statistical information on agriculture, few studies have investigated crop prediction based on historical data. However, owing to the unbridled use of fertilizers comprising nitrogen, potassium, and micronutrients, crop cultivation prediction is a challenge. In general, agro-climatic input parameters such as soil texture, rainfall, and temperature influence crop production. Input parameters for agriculture vary from region to region, and it is daunting to collect such information over large tracts of land. The vast datasets obtained can be used for crop prediction on a massive scale. Owing to the nature of the problems involved, there is a need to develop new machine learning methods for farming arable land and making the most of narrow land resources. Researchers in agriculture have been testing numerous forecasting methodologies to identify the most suitable crop for specific areas of land.

Predicting suitable crop for cultivation is an essential part of agriculture, with machine learning algorithms playing a major role in such prediction in recent years. In this era of technology and data science, the agricultural sector stands to benefit greatly from properly implemented techniques. Feature selection and classification are critical machine learning techniques. Feature selection has to do with selecting the most important attributes from a dataset. It involves picking a subset of appropriate attributes from a larger set of original attributes in terms of a predefined benchmark, such as classification performance or class separability, which plays a significant role in machine learning applications.

Three feature selection methods – filter, wrapper, and embedded are used in the selection of attributes. Filter methods offer rapid execution, though wrapper methods have a better recognition rate. In this work, wrapper feature selection techniques are used to select the best attributes from the dataset, and classification to predict the most suitable crop for a particular piece of land using the selected attributes. There are three common machine learning techniques: supervised, unsupervised, and reinforcement learning. This work uses supervised learning classification techniques for prediction. The principal contribution of this work is to find the best feature selection technique, with a classification method, to predict the most suitable crop for cultivation, based on factors such as soil and environment.

The agricultural sector is going to face enormous challenges in order to feed the 9.6 billion people that the FAO predicts are going to inhabit the planet by 2050: food production must increase by 70% by2050, and this has to be achieved in spite of the limited availability of arable lands, the increasing need for fresh water and other less predictable factors, such as the impact of climate change, which, according a recent report by the UN could lead, among other things, to changes to seasonal events in the life cycle of plant and animals. Farmers in many parts of India are largely dependent on timely rainfall for harvest and subsequent profits. Uncertainty surrounding this phenomenon has also haunted them since the beginning of civilization. Over time however, this uncertainty had reduced significantly as farmers back in the day could almost accurately plant crops based on previous experience with weather conditions. This wisdom has been passed on from one generation of farmers to the other. Gradual onset of global warming and climate changes, over the last century, have slowly-yet steadily put this wisdom out of use. As for rain-fed farmers preparing for agriculture, soil-water equation is fragile and any delay in rainfall could easily mar the harvest. One way to address these issues and increase the quality and quantity of agricultural production is using sensing technology to make farms more "intelligent" and more connected through the so-called "precision agriculture" also known as "smart farming" with machine learning. Much like software, improvements in machine learning have seemingly endless possibilities. Researchers in modern agriculture are testing their theories at greater scale and helping make more accurate, real-time predictions. Modern agriculture has the potential to discover even more ways to conserve water, use nutrients and energy more efficiently, and adapt to climate change. Machine learning in agriculture allows for more accurate disease diagnosis all the while, helping eliminate wasted energy and resources from misdiagnoses.

Farmers can upload field images taken by satellites, UAVs, land-based rovers, pictures from smartphones, and use this software to diagnose and develop a management plan.

Recent ensemble classifiers such as RF are machine learning algorithms that construct a set of classifiers instead of one classifier, and then classify new data points by taking a vote of their predictions. RF classifier can be defined as the collection of all tree structured classifiers. RF is an advanced version of bagging. Instead of partitioning each node using the best split among all the variables, this splits every node using the best among a minimal set of predictors which is randomly chosen at that node. A fresh training data set is originated from the actual data set with replacement. Then, the multiple decision tree is grown using random forest algorithm. This strategy of producing 'N' decision tree with 'N' test result makes RF a better in accuracy when it is compared with other existing all algorithms.

In the economic development of a country like India, agriculture shows a vital role as it provides income and employment to the rural population and it acts as a main source of food. With the time, the need for the crop production is also increasing. In India, it devotes around 20% to the GDP of the country. Today the farmers cultivate the crops depending on the experience picked up from the past age. Because of the old techniques the farmers are not aware about the interest that happens in the current horticultural economy. This results in the misfortune to the farmers. Selection of crop is a most important aspect in agricultural planning. When the farmers know the accurate information on the best crop in their field as per season, it minimizes the loss. The rate of production of a crop relies on many factors:

1) Like weather specific parameters (e.g., rainfall, temperature, humidity etc.), soil parameters (e.g., soil moisture) and geography of a place (e.g., slope). Different datasets of these attributes are collected and then analysed. Collecting the data from right source plays an important role in building a prediction model as it effects on accuracy of the model. The process of analysing data using various analytical and logical reasoning to evaluate each component of the data plays a very important role. This type of examining is just one of the many steps that must be performed when conducting a research analysis. array can be used as one of the techniques to analyse and process the data.

2) There are several existing models for crop prediction about which farmers are un- aware, may be due to its complexity or cost-effectiveness. Hence there is a need of developing such a model that is simple, user-friendly, and cost- effective and reach desired accuracy.

3) All the existing methods are only region (location) based but, in our algorithm, region is embedded with season so that the accuracy of the prediction can be improved. Here crop selection forecast models are prepared based on crop weather studies for estimating yield much before actual harvest of the crops.

A. Problem Statement

The previous model doesn't take larger dataset to give detail information regarding greater number of crops, and it doesn't provide a better way to predict the suitable crops in the changed soil conditions along with this it also doesn't suggest suitable fertilizers for the selected crop. Even detection of Leaf Disease also cannot be done. Previously we don't have crop prediction, fertilizers and leaf detection in a single system. And also in the previous model the UI of the work is designed such that it is easily understandable by the common people.

B. Objectives

In Crops leaf plays a significant job as it gives data about the amount and nature of yield ahead of time contingent on the state of leaf. We propose the framework which should accomplish the following objectives:

1) To develop a model that specify which crop should be grown in the particular land based on the parameters like N, P, K values along with Ph and rainfall values are given as it is extremely useful for the framers in planning for harvest and sale of the grain harvest. To implement Random Forest Algorithm that recommends suitable crop for the corresponding region and for seasonal crops in our country i.e. rice, maize, chickpea, kidney beans, pigeon peas, cotton, Potato, fruits like apple, banana etc..
2) To implement Random Forest Algorithm in order to suggest the suitable fertilizers for the selected crop.
3) To implement CNN Algorithm for the prediction of crop leaf disease

C. Methodology

Figure 1 depicts the overall process of this work. First, the input data is pre-processed to find the missing values, eliminate redundant data, standardize the dataset, and convert target attributes into factor attributes. Essential attributes are extracted from the pre-processed data using wrapper feature selection techniques. The optimized attributes have classification techniques applied to them, prior to which the dataset is split into training and testing phases. Unknown samples from the training dataset are used to train the classification algorithm to determine the crop that is best suited for cultivation in a specific area of land. The testing dataset is used to predict the crop to be raised, using the classifier. Finally, a suitable crop is obtained and the results evaluated using different performance metrics. The analysis reveals the best feature selection technique with an appropriate classification method.
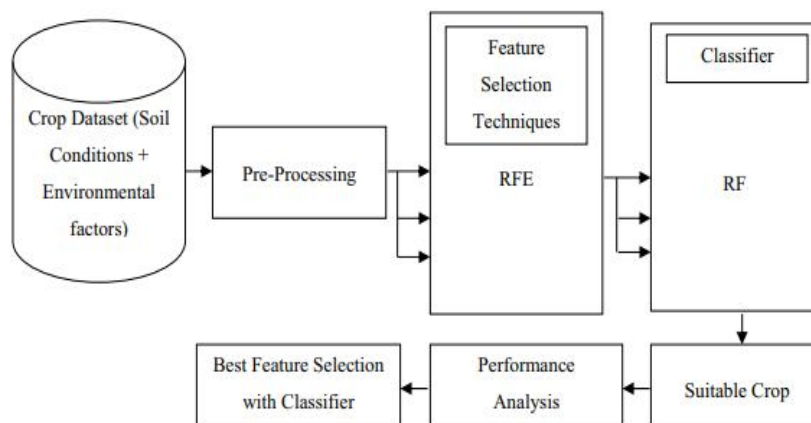

Fig. 1 Block Diagram of Proposed System

II. LITERATURE SURVEY

Clearly, a farmer is the best decision maker in the selection and cultivation of crops. Today, however, cultivar prediction is done manually in laboratories, and farmers need the help of experts to determine the most suitable crops for a specific piece of land. The experts collect soil samples from a particular portion of land and test them in the laboratory, following which they offer suggestions on the ideal crop/s to be raised. Prediction takes time, and selecting the most suitable crops is a complex task in agriculture.

Manual prediction has largely failed, owing to climatic changes and environmental factors that affect crop cultivation. Accurate predictions of suitable crops for cultivation improve production levels. Crop prediction attributes are defined by multiple factors such as genotype, climate and the interactions between the two. Accurate crop prediction needs a fundamental understanding of the functional relationship between cultivation and interactive factors like the genotype and climate. Further, it requires both detailed datasets and efficient algorithms to examine these relationships. Justified by these facts, machine learning techniques are used in this study to predict the most suitable crop for a specific stretch of land, and this technique is ideal for considering factors like the soil and environmental conditions. A number of related studies are discussed in this review:

Sanmay Das discussed the pros and cons of the filter and wrapper methods, and implemented a new hybrid feature selection approach using the boosting technique. The experiments were carried out using real-world datasets from the University of California, Irvine (UCI) repository. The results proved that the proposed method is much faster than the wrapper method [1]. Huan Liu and Lei Yu reviewed the existing feature selection algorithm for classification and clustering techniques. Subsequently, an intermediate step on a unifying platform was proposed in their work [2]. Al Maruf et al. demonstrated the superiority of the gapped k-mer composition and reverse complement features of the k-mer composition over other compositions. The Support Vector Machine (SVM) with the Radial Basis Function (RBF) kernel was used as a classification algorithm. Compared to other approaches, the iRSpot-SF performs considerably better than the Matthews, with a correlation coefficient and sensitivity of 69.41% and 84.57% and it has 84.58% accuracy [3]. Jana Novovicova et al. proposed a feature selection method with no search procedure, and one best suited for multimodal data [4]. Isabelle Guyon and Andre Elisseeff, also briefly discussed a feature selection method based on the filter and wrapper approaches and, in addition, defined feature ranking and multivariate feature selection [5]. Jia-You Hsieh et al. in their study, discussed Rice Blast Disease (RBD). The Recursive Feature Elimination (RFE) algorithm with the Auto-Sklearn was used to select key features impacting RBD. The aim of their work was to build a model as a warning mechanism for RBD [6]. Ron Kohavi and George H. John compared the wrapper and induction methods without feature subset selection, and proceeded thereafter to compare them to Relief, a filter method with feature subset selection. The strengths and weaknesses of the wrapper approach were discussed, and a series of improved designs shown [7]. Isabelle Guyon et al. implemented a Support Vector Machine (SVM) technique based on the Recursive Feature Elimination (RFE) for gene selection. Of the different methods used to select features, the RFE is a newly-developed method that selects features for small sample classification problems [8]. Marc Sebban and Richard Nock, analysed the filter model with information gain and a statistical test. A hybrid model was implemented using a minimum spanning tree that was replaced by the first nearest neighbor [9]. Lei Yu and Huan Liu, proposed a correlation filter method termed the fast correlation-based filter. Their technique was verified by two different classification algorithms in terms of real-world data, with and without feature selection [10]. Petr Somol et al. proposed a flexible hybrid sequential floating search algorithm based on the principles of the filter and wrapper methods. The advantage of the proposed method was its flexibility in terms of a trade-off between the quality of the results versus computational time, as well as enabling the wrapper-based feature selection approach to deal with problems of higher dimensionality. Experiments were carried out using the WAVEFORM dataset from the UCI repository and the SPEECH dataset from British Telecom [11]. Salappa et al. analysed the performance and efficiency of an array of feature selection algorithms with classification methods. Their experimental analysis was carried out on 15 datasets from the UCI repository. The results show that most Feature Selection Algorithms (FSAs) significantly reduce data dimensionality without impacting the performance of the resultant models [12]. Kursa et al .implemented Boruta, an all-relevant feature selection method which gathers every feature that is critical to the outcome in certain circumstances. By contrast, most traditional feature selection algorithms follow a minimally optimal method in which they rely on a small subset of features that yield a minimal error on a selected classifier [13]. Marcano Cedeno et al. proposed a feature selection method based on sequential forward selection and the feed forward neural network to find the prediction error as a criterion for selection [14]. Zahra Karimi et al .implemented a feature ranking method using a hybrid filter feature selection scheme for intrusion detection in a standard dataset. The experimental results show that the proposed technique offers higher accuracy than other methods [15]. Surabhi Chouhan et al. proposed a hybrid combination method of applying the Particle Swarm Optimization – Support Vector Machine (PSO-SVM) to select features from a dataset. Assorted benchmark datasets were tested with this technique [16]. David Heckman et al. described that the harnessing of natural variability in photosynthesizing ability as a way to improve yields, through a functional phylogenetic analysis for large-scale genetic screening is a laborious task. The potential for leaf reflectance spectroscopy to estimate photosynthetic efficiency specifications in Brassica oleracea and Zea mays, a C3 and a C4 seed, respectively, were analysed, the findings show that phenotyping leaf reflectance is an effective method to enhance the photosynthetic ability of crops [17]. Maya Gopal and Bhargavi proposed a wrapper feature selection method featuring Boruta that extracts features from a dataset for crop prediction. The technique improves prediction performance and provides effective predictors.

In Boruta, the Z score has the most accurate measure, since it takes into consideration the variability of the mean loss of accuracy among trees in a forest [18]. Aileen Bahl et al. developed a random forest (RF) model with the RFE for improved prediction accuracy [19]. Maya Gopal and Bhargavi analysed the performance of machine learning (ML) algorithms with a variety of feature selection techniques for crop selection prediction. The results showed that the random forest provides higher accuracy than other ML algorithms [20]. Maya Gopal and Bhargavi proposed sequential forward selection, which is a special sequential feature selection process. It is a greedy search algorithm that attempts to find the 'optimal' feature subset by iteratively selecting features based on the performance of the classifiers [21].

III. SYSTEM DESIGN

System design thought as the application of theory of the systems for the development of the project. System design defines the architecture, data flow, use case, class, sequence and activity diagrams of the project development.

A. System Architecture

The below architecture diagram in figure 2 illustrates how the system is built and is the basic construction of the software7method. Creations of such structures and documentation of these structures is the main responsible of software architecture.
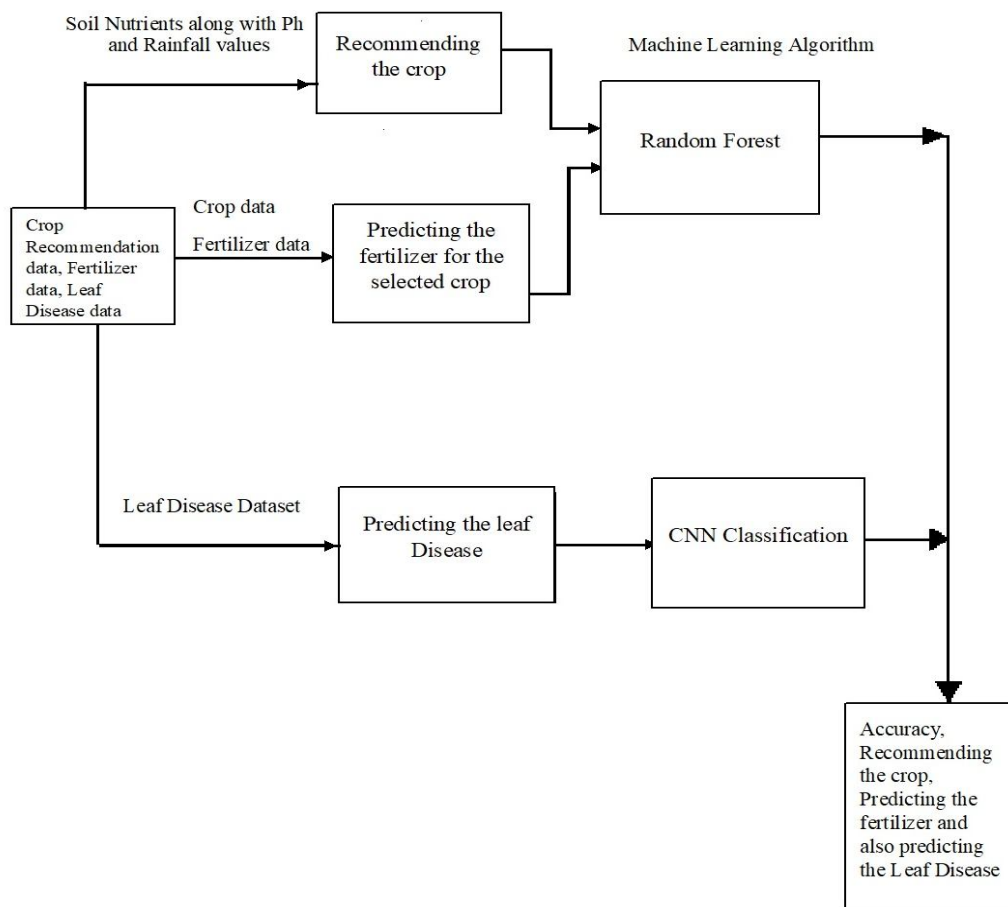

Fig. 2 Architecture Diagram of Proposed System

The above figure 2 represents that through the system, the user can able to take decision about consulting which crop to deploy and also user can able to take decision about using of preferred fertilizer for the recommended crop by proving the exact information about the soil nutrient and crop. We created a dataset for crop recommendation and fertilizer prediction along with datasets for leaf disease prediction from various sources. Using this information, the system can provide accurate prediction of a particular crop and fertilizer based on Random Forest algorithm. This system also gives accurate information about predicting the leaf disease using CNN.

In leaf disease prediction there is a dataset which consist of all the different plant leaf diseases which we have taken into account. The module is trained repetitively to attain the maximum accuracy. If a new image is given to the model it's features get compared with the features that are already trained in the dataset. It then provides the appropriate result.

B. Dataflow Diagram

Data flow diagram also referred as bubble graph. This diagram is useful for representing the system for all degree of constructions. The figure is differentiated into parts which show maximizing data path & practical aspect. The below figure 3 shows the dataflow of the proposed system. As in the Figure 3, firstly we need to collect data with respect to fertilizers like by considering Potassium, Nitrogen and Phosphorus and next step we need to build the prediction model and later we need to consider the predicted crop with respect to this fertilizer for the required crop is suggested
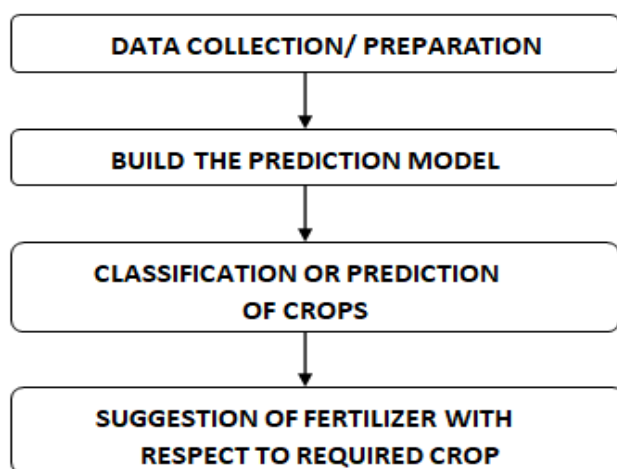
Fig. 3 Workflow for Fertilizers Selection

C. Sequence Diagram

A sequence diagram is a type of UML (Unified Modeling Language) diagram that depicts the interactions between objects or components in a system or process. It shows the order in which messages are exchanged between the objects, along with the time sequence of those messages. Below figure 4 and 5 shows the sequence diagram of crop and fertilizer recommendation.
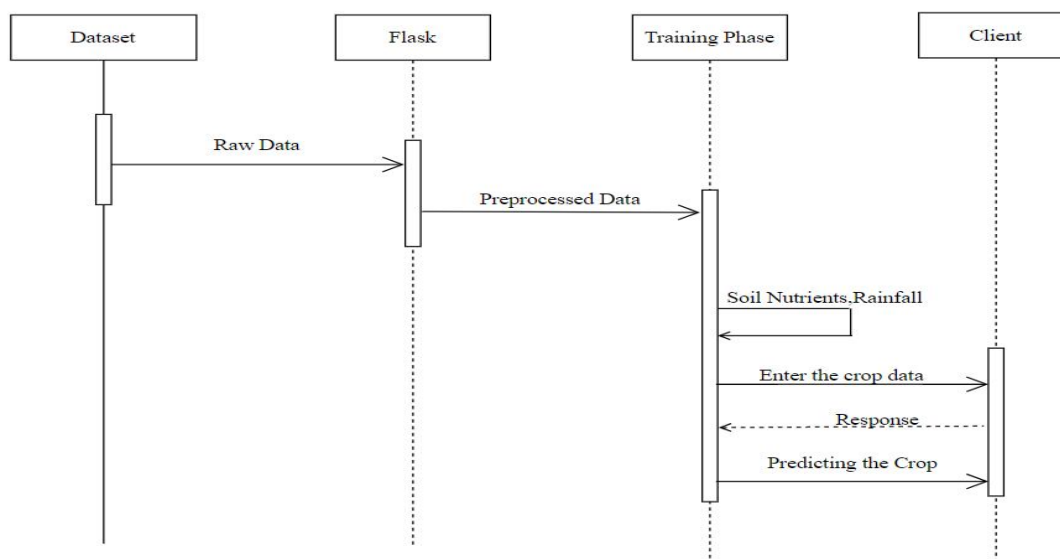

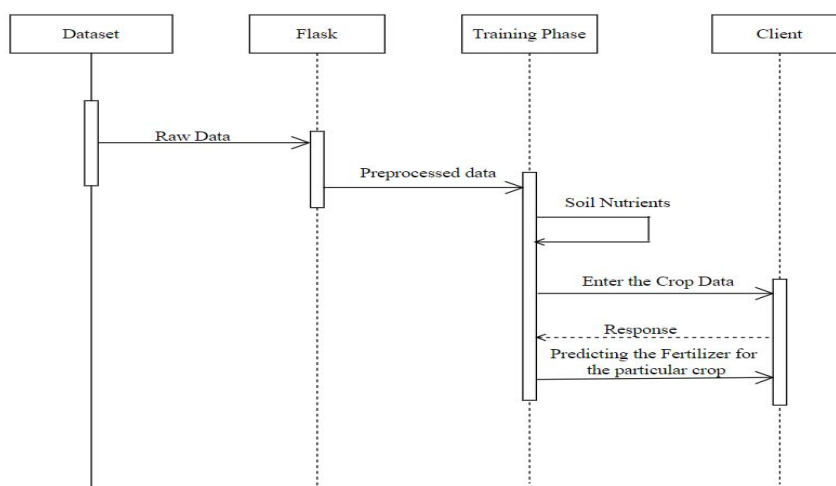Fig. 4 Sequence Diagram of Crop Recommendation

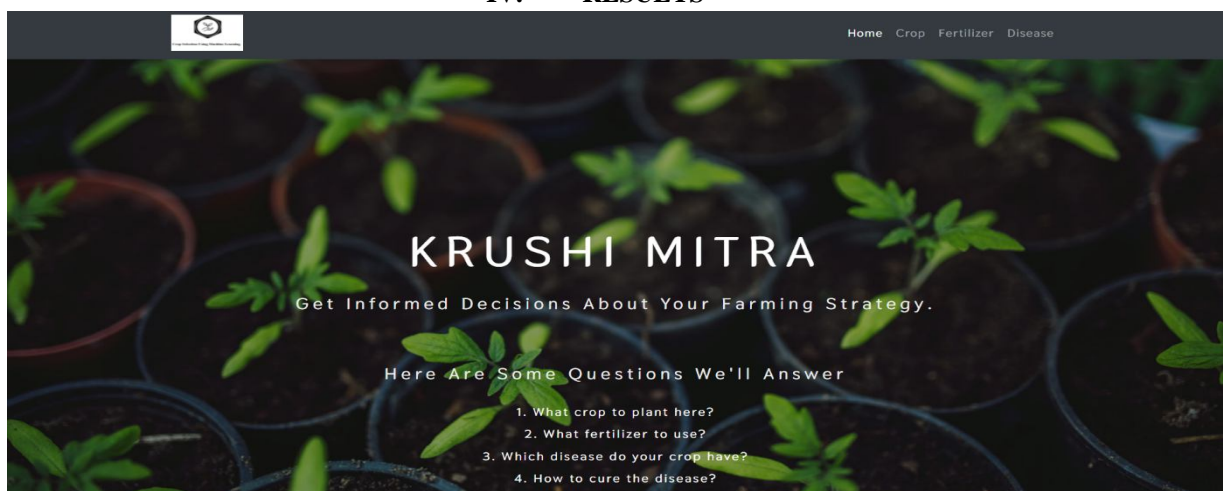Fig. 5 Sequence Diagram of Fertilizer Recommendation

IV.   RESULTS

Fig. 6 Home Page

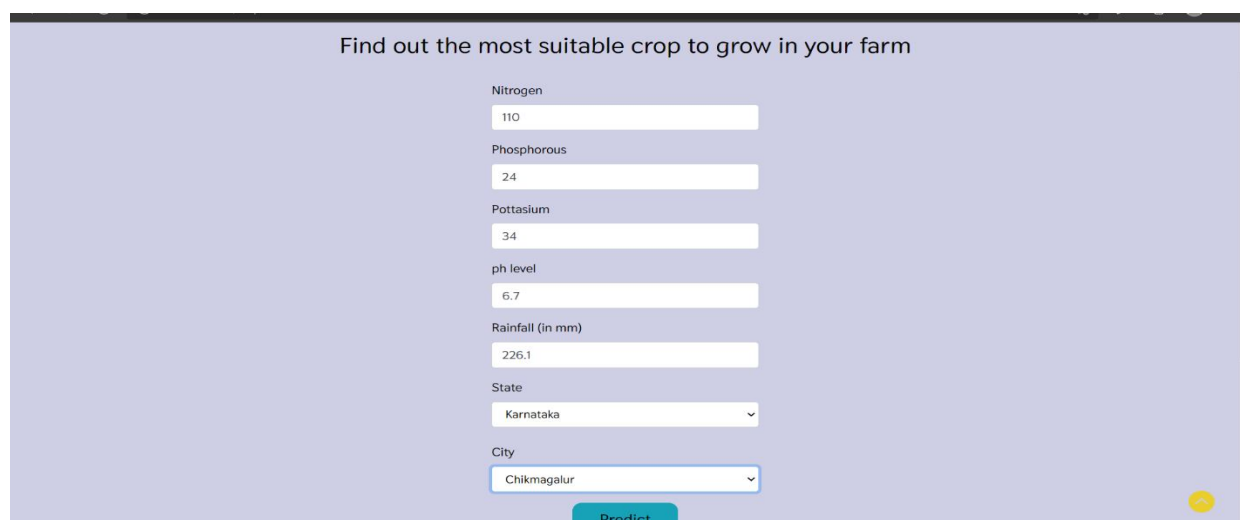Fig. 7 Input Factors for getting which crop can be grown in the particular land

Fig. 8 Crop recommended

Fig. 9 Input values to predict fertilizer for the selected crop

Fig. 10 Fertilizer suggested for the selected crop

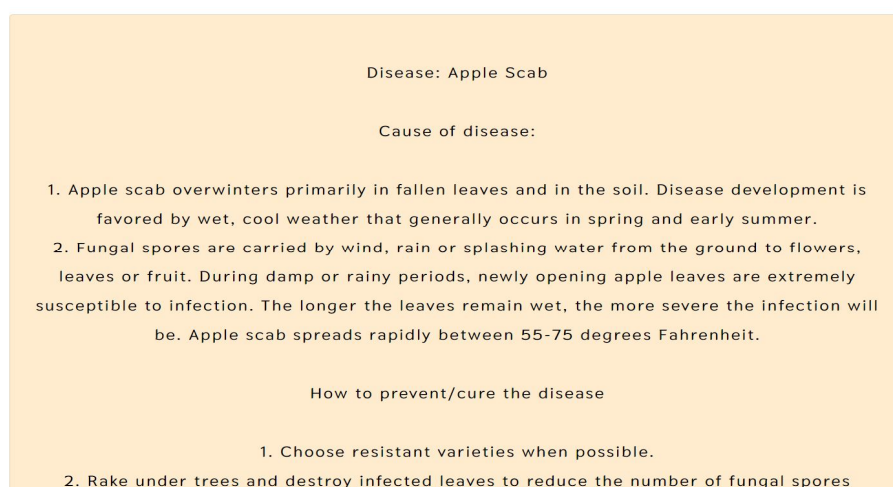Fig. 11 Uploading the leaf image for predicting the disease

Fig. 12 Disease predicted for the uploaded leaf image

Figure 11 and 12 shows the results of test image uploaded and predicted. Results contain the details of disease identified and also prefer the pesticide for the identified disease along with the description of the usage of pesticide.

V. CONCLUSION

This work is employed to gain the knowledge about the crop that can be deployed to make an efficient and useful harvesting. While the existing system, lags a bit in accuracy in prediction. The ultimate result that was foretold wasn't that accurate with the original. To overcome this drawback, we have a proposed a work replacement technique with random forest algorithm. It takes only less time for its processes and the accuracy of the prediction is high. In leaf disease prediction, it focus on how image from given dataset (trained dataset) are used predict the plant leaf diseases using CNN model. This brings some of the insights about plant leaf disease prediction. As maximum types of plant leaves will be covered under this system, farmer may get to know about the leaf which may never have been cultivated and can lists out all possible plant leaves, it helps the farmer in decision making of which crop to cultivate. Also, help the farmer to get insight into the demand and the cost of various plants in market.

REFERENCES

[1]   S. Das. Filters, wrappers and a boosting-based hybrid for feature selection. In International conference on machine learning 1 (2001), pp. 74–81.

[2]   H. Liu and Y. Lei, toward integrating feature selection algorithms for classification and clustering, IEEE Trans. Knowl. Data Eng. 17 (2005), pp. 491–502. doi:10.1109/ TKDE.2005.66.

[3]   A. Maruf, M. Abdullah, and S. Shatabda, iRSpot-SF: Prediction of recombination hotspots by incorporating sequence-based features into Chou's Pseudo components, Genomics 111 (2019), pp. 966–972. doi: 10.1016/j.ygeno.2018.06.003.

[4]   I. Kononenko, Estimating attributes: Analysis and extensions of RELIEF. European conference on machine learning 784 (1994), pp. 171–182.

[5]   J. Pavel Pudil, N.N. Choakjarernwanit, and J. Kittler, Feature selection based on the approximation of class densities by finite mixtures of special type, Pattern Recognit 28 (1995), pp. 1389–1398. doi:10.1016/0031-3203(94)00009-B.

[6]   J. Novovicová, P. Pudil, and J. Kittler, Divergence based feature selection for multimodal class densities, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1996), pp. 218–223. doi:10.1109/ 34.481557.

[7]   H. Liu and H. Motoda, Feature Selection for Knowledge Discovery and Data Mining, 1st ed., Springer, US, 1998.

[8]   M. Dash, K. Choi, P. Scheuermann, and H. Liu. Feature selection for clustering-a filter solution. IEEE International Conference on Data Mining, 2002, Proceedings., Maebashi City, Japan, pp. 115–122.

[9]   I. Guyon, S. Gunn, M. Nikravesh, and L.A. Zadeh (eds.), Feature Extraction: Foundations and Applications, 1st ed., Springer-Verlag, Berlin Heidelberg, 2008.

[10] I. Guyon and A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (2003), pp. 1157–1182.

[11] W. Paja, K. Pancerz, and P. Grochowalski, Generational feature elimination and some other ranking feature selection methods, in Advances in Feature Selection for Data and Pattern Recognition, International Publishing, Springer, Charm, Vol. 138, 2018, pp. 97–112.

[12] J.-Y. Hsieh, W. Huang, H.-T. Yang, C.-C. Lin, Y.-C. Fan, and H. Chen. Building the rice blast disease prediction model based on machine learning and neural networks., Easy chair the world of scientists (2019), pp. 1–8.

[13] R. Kohavi and G.H. John, Wrappers for feature subset selection, Artif. Intell. 97 (1997), pp. 273–324. doi:10.1016/S0004-3702(97)00043-X.

[14] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik,   Gene selection for cancer classification using support vector machines, Mach Learn 46 (2002), pp. 389–422. doi:10.1023/ A:1012487302797.

[15] M. Sebban and R. Nock, A hybrid filter/wrapper approach of feature selection using information theory, Pattern Recognit 35 (2002), pp. 835–846. doi:10.1016/S0031- 3203(01)00084-X.

[16] Y. Lei and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In Proceedings of the 20th International Conference on Machine Learning, Washington, DC, USA, 2003.

[17] P. Somol, J. Novovičová, and P. Pudil, Flexible-hybrid sequential floating search in   statistical feature selection, in Joint IAPR International Workshops on Statistical Techniques in Pattern Recognition (SPR) and Structural and Syntactic Pattern Recognition (SSPR), Dy. Yeung, J. T. Kwok, A. Fred, F. Roli, D. de Ridder eds., Vol. 4109, Springer, Berlin, Heidelberg, 2006, pp. 632–639.

[18] A. Salappa, M. Doumpos, and C. Zopounidis, Feature selection algorithms in classification problems: An experimental evaluation, Optimisation Methods and Software 22 (2007), pp. 199–212. doi:10.1080/10556780600881910.

[19] M.B. Kursa, A. Jankowski, and W.R. Rudnicki, Boruta–a system for feature selection, FundamentaInformaticae 101 (2010), pp. 271–285. doi:10.3233/FI-2010-288.

[20] A. Marcano-Cedeno, J. Quintanilla-Domínguez, M.G. Cortina-Januchs, and D. Andina. Feature selection using sequential forward selection and classification applying artificial metaplasticity neural network. In Industrial electronic conference, 36th annual conference on IEEE industrial electronics society, Glendale-USA, 2010.

[21] Z. Karimi, M.M.R. Kashani, and A. Harounabadi, Feature ranking in intrusion detection dataset using combination of filtering methods, International Journal of Computer Applications 78 (2013), pp. 21–27. doi:10.5120/13478-1164.

[22] S. Chouhan, D. Singh, and A. Singh, An improved feature selection and classification using decision tree for crop datasets, International Journal of Computer Applications 142 (2016), pp. 5–8. doi:10.5120/ijca2016909966.

[23] D. Heckmann, U. Schlüter, and A.P.M. Weber, Machine learning techniques for predicting crop photosynthetic capacity from leaf reflectance spectra, Mol Plant 10 (2017), pp. 878–890. doi: 10.1016/j.molp.2017.04.009.

[24] P.S. Maya Gopal and R. Bhargavi., Feature selection for yield prediction in Boruta algorithm, International Journal of Pure and Applied Mathematics (2018), 139–144.

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 ⊘ (24*7 Support on Whatsapp)