International Journal of Pure and Applied Mathematics
Volume 118 No. 18 2018, 2805-2809
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue
ijpam.eu

Identification of Crop Disease by Predictive Analysis in Hadoop Environment

Ms. G.Shobana
Assistant Professor,
Department of CSE
Kumaraguru College of Technology,
Coimbatore, India.
shobana.g.cse@kct.ac.in

Ms. M.Suguna
Assistant Professor-II
Department of CSE
Kumaraguru College of Technology,
Coimbatore, India
suguna.m.cse@kct.ac.in

D. Yamunathangam
Assistant Professor,
Department of CSE
Kumaraguru College of Technology,
Coimbatore, India.
yamunathangam.d.cse@kct.ac.in

***Abstract*-Agriculture is one of the important sources for survival of life. For the past few decades, new technologies are used to improve the better productivity of the crops. Many researchers have developed monitoring and automation system for different functionalities of farming. To yield high profit in agriculture, growth of crops has to be monitored by IOT enabled devices as result massive unstructured data are generated in regular interval of time. In the proposed system, predictive analyses in Hadoop environment are used to identify the disease of crop and growth of the crops. Based on prediction, type of pesticides and the nutrition required for fields are identified. The people can easily make decision from historical data with agricultural analysis to yield high productivity.**

Keywords: unstructured data; hadoop; mapreduce; sensors.

I. INTRODUCTION

The Internet of Things and big data is a huge opportunity for farmers to monitor their crops and increase productivity**.** The various parameters have to identify for improving the agriculture. First is location of field depending upon the characteristics of the location only crops can be cultivated. Second is the Soil, pH value of soil, other properties like Phosphate, Nitrate and Potassium in the soil are factor which will determine the soil quality and type of crop that can be cultivate. Third, weather and climate are highly localized. The main things are that after cultivated the crops, growth of crops has to be monitored.

To monitor the growth of the plants, we have to use sensors for sensing the temperature, the moisture content in the soil. The growth of plant is monitored by using sensors like infrared sensor, color sensors, temperature, humidity sensors etc. The data absorbed from sensors can be evaluated in real-time for signaling the vital values that play an important role in production decision-making. All the data that are obtained from various sources are stored in the cloud for analysis [1]

As the data are sensed in the real time, the huge amount of data is obtained as result from various resources. In order to handle the large amount data from resource, we are moving to the big data. Big data analytics is used to process the large amount of unstructured data and convert those data into useful information. The Big data has V's that includes high-volume, high-velocity and high-variety and value, from this different type of information from various resources the data are processed to obtain the decision from the data.

The massive amounts of data are to be analyzed for identifying the disease in the agricultural field This paper implements the pattern matching algorithm for big data using Hadoop platform to deal with large amount of datasets which helps farmers to better understand their crop status to take accurate decisions .

The paper is organized as follows. Section 2 presents related work of existing agricultures systems and their algorithm, Hadoop environment. Section 3 presents architecture of the proposed system. Section 4 presents predictive algorithm and finally Section 5 conclusion of the works

II. LITERATURE SURVEY

Every year, over 25% of the worlds annual output of agricultural goods are ruined or damaged by pets and disease. The crop disease can be caused by fungi, bacteria, virus, etc. Fungal disease tend to produce mycotoxins, many of which are powerful carcinogens, contaminating the agricultural products so that they are unsafe for human begins. Some of the fungal disease includes northern blight, common rust, Fusarium, aspergillus. To increase crop yield,it is important to monitor growth of plants and also treatment of plant disease by detecting disease accurately and early.

S.A. Ramesh Kumar et al has identified some of symptoms of diseases occurred in paddy crops. The images of crops are captured to detect diseased leaf and stem, find color and texture of affected area etc. Based on analysis it helps to identify the disease of paddy crops [1]. Digital image of the crops contain finite number of elements, each of which can have particular location and value. The images of crops are represented as a 2D function f(x, y), where x and y are spatial coordinates, and the amplitude of at any pair of coordinates (x, y) is called as grey level of the image. First segmentation is done which is based on edge detection. It take the image which is RGB model as input image and then feature values will be calculated and then perform classifier on the data. Crops image are mapped with historical dataset for identifying disease of the crops [4].

T. Rumpf et al (2010) had proposed methods for precision crop protection to detect plant diseases early and automatically. Decision tree algorithm, artificial neural network with support vector machines were used to identify the classification between healthy leaves and leaves with disease symptoms in sugar cane fields [2]. In decision tree algorithm, data are split based on the relevant features, the next optimal data splitting feature is determined in the tree.

In automatic agriculture monitoring system many sensors are used to monitor the crop field. Ziang Zhou et al had stated low power sensor network measures temperature, humidity and light intensity through wireless sensor nodes equipped with different sensors. Collected data by the sensors and the data are send to storage area.The base station will analyze data which is received. Communication used for transmitting the data is ZigBee technology. The data are stored in the database for future use. [3]

Kaur et al stated that for analysis massive amount of unstructured data the hadoop framework is efficient. Hadoop framework is an open source data processing framework that supports processing of large chunks of distributed data using simple programming models. The Hadoop cluster is a set of machines networked together in one location. Data storage and processing of all data occur within this cloud" of machines. User can submit jobs to Hadoop from his desktop machine in remote location from the Hadoop cluster [5].The main components of Hadoop systems are MapReduce and Hadoop Distributed File System (HDFS). HDFS means distributed file system management for large datasets of sizes of gigabytes and petabytes. MapReduce framework divides huge dataset into smaller independent units and processes these unstructured data in parallel.

HDFS has a master and slave architecture. The main components of an HDFS cluster are a single NameNode and a master server . File system management and access control to files can be done by master server. In addition, in the cluster, each node will have one DataNode.

The Map Reduce software framework, as a programming model was introduced by Google in 2004 and later adopted by Apache Hadoop. It spilts the large chunks of data and then perform Map and Reduce phases. MapReduce is a processing large datasets in parallel using lots of computer running in a cluster. We can extend the mapper class with our own instruction for handling various input in a specific manner. During map master node instructs worker nodes to process local input data and Hadoop performs shuffle process. Thus master node collects the results from all reducers and compilers to answer overall query [5].

Hrishikesh et al had proposed method for feature extraction from disease affect leaf, Hue Saturation Intensity (HSI) color space representation can be obtained from first RGB images of leaves. Diseased parts of the leaves are extracted and labeled as connected components. Labeling of Connected component can be done by an algorithmic application of graph theory, where unique labeling based on tests can be done on individual subsets.It actually collects diseased parts of the leaves as same or near valued pixels of a region. Here another binary image can be produced by removing connected components which have less relevance (objects with less than 30 pixels). Some of the features are extracted from the samples which includes 8 color features and size of disease size of diseased spot , distances of diseased spots[6]

Sharada P. Mohanty et al had proposed the system for deep convolution neural network from which they identify 14 crop species and 26 diseases. The trained model achieves an accuracy of 99.35% on a held-out test set, demonstrating the feasibility of this approach. Diagnosis of crop disease by smart phone can be done by training deep learning models on publicly available image [7].Computer based technological improvements in agricultural systems emerged with new dimensions with the help of ICT (Information and Communication Technologies). But these systems are not able to fulfill the needs of today's generation due to various factors like large amount of data to be processed, processing speed, storage space of data, reliability, availability, scalability etc.

The main aim is to develop the system for predicting the disease in the agriculture field. All existing work concrete only on particular diseases of crop, only on particular crop growth condition, whereas in proposed system help us to identify the various disease in different crops in the agricultural fields. The large amounts of information about the crops are stored in distributed file system to avoid the loss of data.

III. PROPOSED SYSTEM

The system monitors the agriculture fields using sensors and actuators. The data collected from the fields are unstructured data. The architecture of proposed system is shown the figure 3.1.

3.1 Data collection

The information is collected from the various sensors like infrared sensors, color sensors, humidity sensors. The height, width, diameter of stem, flower, color of leaves, steams, moisture content of the soil are the some parameter identified for monitoring for the growth of the plant. The agriculture fields are sensed in regular interval of time which contains unstructured data in form of values and images. Historical report can be get from agriculture department related to individual crops of specific geographically area.

3.2 Data preprocessing and storage

The unstructured data collected from sensors, historical data contains images, values, noisy etc. Real world data is often incomplete and inconsistent, clean data for the processing to identify the pattern in the growth of the plant. The massive unstructured data are collected and stored in single unit called as HDFS system. The system creates group of nodes and coordinates work among cluster of nodes**.** The information about growth of crops and historical data about disease of crops are stored. In HDFS system data are stored redundantly across the server to avoid the losses of data. Data are split according to features which is independent of each other for parallel execution

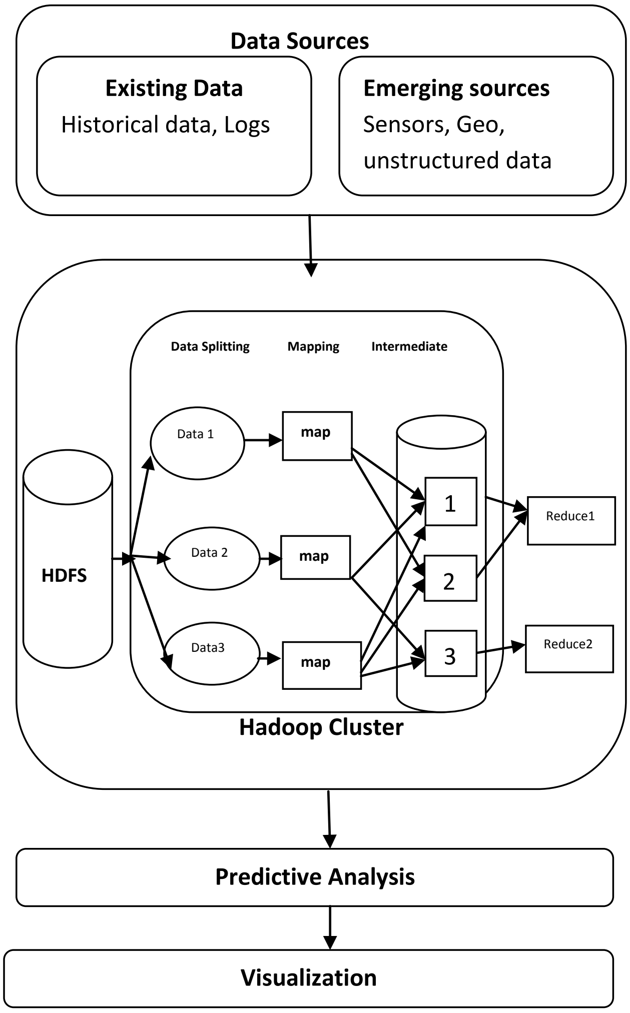

Figure 3.1 Architecture of Proposed System

IV.PREDICTIVE ANALYSIS

The system helps to identify the growth of the plants with various parameters like nitrogen, phosphorus, potassium, magnesium, sulfur, calcium. The hadoop system has to main function as Map function and Reduce function. The map function splits data into smaller independent subtasks and processes the data in parallel and combines the result. The Reduce function collects the output data from map function and process final result to the query

4.1 Pattern matching Algorithm

The captured images of leaves are given as input to the hadoop ecosystem. In the HDFS file system contains the information about the historical images, some rules about the disease occurrence are identified in the crops are stored. It is useful for detection of diseases. Input data are given to Job tracker. Job tracker will split the jobs into independent job and form key, value pairs for parallel processing. Each individual job is assigned to the task trackers. Task trackers will be performing the tasks independently. Task tracker identifies the irregular pattern from the captured images. Irregularities in the images can be spots, change in texture, color etc. Rules are already defined for disease based on the information from the report of agricultural department. For the identification certain disease, mapping calculates the similarity of the rules for each data. Based on the information we can able to identify how much symptoms match with the symptoms of disease. The deviation can be identified by comparing pixel value of the original images with the existing data for different types of crops .When Compared with these rules**,** matching rate > =α. α which be defined by expert as the threshold value**.** If the data matches with data bases, information will provided to user.

In mapping phase, Job trackers split the task trackers, it will perform the task independently. The intermediate results are obtained and data shuffling will take places. The pattern matching algorithm is running parallel and result is obtained.

The result are visualized that will determine whether plants are affected by disease or not. The types of disease, color of the crops can be identified. Based on information pesticide can be given to field to improve the growth of the plants. Since agriculture fields are monitored by using sensors, the manual works of farmers are reduced and early identification of the disease in the crops.

V.CONCLUSION

This paper provides a path in diagnosing the disease by the identification of presence of disease in plant by predictive analysis. The approach is well suitable for different crop types and diseases. The result will be expected to improve considerably with more training data. Early prediction disease in the agriculture field helps the farmer for identifying the severity of the detected disease as the result it will increase the productivity of the crops.

REFERENCE

[1]. Ashwani Kumar Kushwaha, Sweta Bhattachrya, "Crop yield prediction using Agro Algorithm in Hadoop", in International Journal of Computer Science and Information Technology & Security, 2015

[2]. T. Rumpf , K. Mahlein, "Early detection and classification of plant diseases with Support Vector Machines based on hyperspectral reflectance", 2010

[3]. Ziang Zhou, Kun Xu, "Design of Agricultural Internet of Things Monitoring System Based on ZigBee", 2016

[4]. Sujeet Varshney, Tarun Dalal, "Plant Disease Prediction using Image Processing Techniques", 2016

[5]. Dr. Doreswamy, Ibrahim Gad , B.R. Manjunatha, "Big Data Aggregation Using Hadoop and Map Reduce Technique For Weather Forecasting", International Journal of Latest Trends in Engineering and Technology, 2016

[6]. Hrishikesh, P. Kanjalkar, S.S.Lokhande , "Feature Extraction of Leaf Diseases", International Journal of Advanced Research in Computer Engineering & Technology, 2014

[7]. Sharada P. Mohanty, David P. Hughes, "Using Deep Learning for Image-Based Plant Disease Detection", 2016

[8]. Kaur, Anureet. "Big Data: A Review of Challenges, Tools and Techniques. IJSRSET,

[9]. G. Mansingh, H. Reichgelt, & K. M. O. Bryson, "CPEST: An expert system for the manage ment of pests and diseases in the Jamaican coffee industry", Expert Syst. Appl., vol. 32, No. 1, pp. 184-192, 2007.

[10]. W. A. DerwinSuhartono, M. Lestari, & M. Yasin, "Expert System in Detecting Coffee Plant Diseases", Int. J. Electr.Energy, vol. 1, No. 3, pp. 156-162, 2013.