Exploring Machine Learning Approaches on Crop Pattern Recognition

Kazi Hasibul Kabir

ID: 2014-1-60-042

Sharmin Sultana

ID: 2014-1-60-070

Md. Zahiruddin Aqib

ID: 2014-1-60-047

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
East West University
Dhaka-1212, Bangladesh

April, 2018

Declaration

I, hereby, declare that the work presented in this thesis is the outcome of the investigation performed by me under the supervision of Dr. Shamim Akhter , Associate Professor, Department of Computer Science and engineering, East West University. I also declare that no part of this thesis/project has been or is being submitted elsewhere for the award of any degree or diploma.

Countersigned   Signature

. . . .

(Dr. Shamim Akhter)   (**Kazi Hasibul Kabir and 2014-1-60-042**)

Supervisor

Signature

. . . .

(**Sharmin Sultana and 2014-1-60-070**)

Signature

. . . .

(**Md. Zahiruddin Aqib and 2014-1-60-047**)

i

Abstract

Agriculture activities monitoring is important to ensure food security. Remote sensing plays a significant role for large scale continuous monitoring of cultivation activities. Time series remote sensing data were used for the generation of the cropping pattern. Classification algorithms are used to classify crop patterns and mapped agriculture land used. Some conventional classification methods including support vector machine (SVM) and decision trees were applied for crop pattern recognition. However, in this report, we are proposing different machine learning approaches such as Naive Bayes (NB), Deep Neural Network (DNN) and Random Forest (RF) classification to improve and find a better solution for crop pattern recognition.

Acknowledgments

As it is true for everyone, we have also arrived at this point of achieving a goal in our life through various interactions with and help from other people. However, written words are often elusive and harbor diverse interpretations even in one's mother language. Therefore, we would not like to make efforts to find best words to express my thankfulness other than simply listing those people who have contributed to this thesis itself in an essential way. This work was carried out in the Department of Computer Science and Engineering at East West University, Bangladesh.

First of all, we would like to express my deepest gratitude to the almighty for His blessings on us. Next, our special thanks go to our supervisor, "Dr. Shamim Akhter", who gave us this opportunity, initiated us into the field of "Exploring Machine Learning Approaches on Crop Pattern Recognition", and without whom this work would not have been possible. His encouragements, visionaries and thoughtful comments and suggestions, unforgettable support at every stage of our BS.c study were simply appreciating and essential. His ability to muddle us enough to finally answer our own question correctly is something valuable what we have learned and we would try to emulate, if ever we get the opportunity.

We also would like to convey our special thanks to Dr. Yann Chemin and his team for the research paper which helped us with the dataset we used in our thesis.

We would like to thank our parents for their unending support, encouragement and prayers.

There are numerous other people too who have shown me their constant support and friendship in various ways, directly or indirectly related to our academic life. we will remember them in our heart and hope to find a more appropriate place to acknowledge them in the future.

<div align="right">

Kazi Hasibul Kabir

April, 2018

Sharmin Sultana

April, 2018

Md. Zahiruddin Aqib

April, 2018

</div>

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

Agricultural activities monitoring including crop pattern generation is inevitable for estimating the agricultural economy of a country. Crop pattern generation requires proper estimation of crop classification and the classification tools needs proper crop related information. On field survey or other traditional data collection methods demand high cost in both time and economy and includes a number of manpowers. Remote Sensing (RS) data has been proven as a more effective way to monitor crop activities as well as provide useful crop related information. Many developed and developing countries GDP depends on agriculture. Thailand is a country at the center of the Indochinese in Southeast Asia. Agriculture is one of the major sectors contributing to the GDP of Thailand. In 2017, the agricultural contribution to the GDP of this country is 13% which is (million $) 51949 [1]. For such countries, monitoring agricultural activities is very important. Besides the changing pattern of the crops due to environmental degradation like heavy rainfall, flood, drought etc. leads to find a better crop monitoring system. However, differentiation of the pattern of the crops is a hard task because of the variation of irrigation time, temperature, weather, moisture and quality of the soil. So, crop classification craves a reliable system which can recognize or classify a particular crop with variations. Researchers applied different algorithms to classify crop pattern, such as, K-means algorithms, Support Vector Machine, Decision Trees etc [2]. We have investigated different machine learning approaches including Artificial Neural Network (ANN), Deep Neural Network (DNN), Naive Bayes (NB) and Random Forest (RF) classifiers to trace the best

approach on accuracy domain for crop pattern recognition. ANN is trained to feed the patterns into the network. Later, it is able to recognize once the assignation is done. NB classifiers are the probabilistic classifiers. This algorithm is based on the Bayes theorem where a dynamic self-reliance assumption between the features is made. DNN is different from the single hidden layer ANN. DNN supports many hidden layers and each layer consists of some nodes through which the data passes in a multistep process of pattern recognition. So, the empirical term of DNN means more than one hidden layer and each layer of nodes trains on a distinct set of features based on the previous layer's output [3]. RF is a supervised classification algorithm. It creates a forest by creating some random trees. However, creating the forest is a complex task and is not same as constructing the decision tree with information gain or gain index approach. The above mentioned machine learning approaches are implemented and tested, and then a comparative analysis is done to find the optimal one for the crop pattern recognition.

1.0.1 Objective

To recognize a particular crop by implementing different algorithms in crop pattern recognition system and find which algorithm performs best by comparing the accuracy. To fulfill the main objective there are some sub objectives which we need to consider.

- Analyzing the image data for pattern recognition

- Learning different tools to apply the algorithms

- Implementing different types of algorithms

- Analyzing the result

Chapter 2

Literature Review

Researchers are practicing different crop classification methods to improve the results and to provide a better crop monitoring system. Paper [4] is the most recent work related to crop classification which is done by using RF and Object-based classification (OBIA) scheme for produce land use maps of a smallholder agricultural zone. They use optimization techniques to reduce the number of input variables in the data processing to reduce the processing time by maintaining a smart accuracy level (91.7%). In paper [5], the crop classification is done using RF over time series Landsat 7 ETM+ data. They have used a time series of medium spatial resolution enhanced vegetation index (EVI) to produce variables for the RF classifier. Susceptibility of training data, number of variables and accuracy are the evaluation area of their study. They achieved 81% accuracy level. In paper [6], the author implemented different supervised machine learning approaches for crop classification from Multi-Temporal RS image including Softmax Regression-92.06% accuracy, Support Vector Machines (SVMs)-92.21%, Neural Network (NN)-92.06%, and Convolution Neural Network (CNN)-92.64%. In paper [7], the classification of crops is being done using Fuzzy C-Means (FCM) segmentation and texture, color features. In this paper the author used ten different JPEG images from the field. After the segmentation, they have applied the color and texture feature to get the accurate color texture feature of each crop. Thereafter they used Euclidean distance algorithm to identify the crops. ANN classified the crops of rice, sugarcane and maize successfully in [8] with around 70%-75% accuracy. Back propagation is executed for training the data, delta rule is applied to

adjust weights. With the enough training data, the ANN can easily encountered the data and recognize them. All of the above mentioned papers applied different machine learning approaches; however, none of them have applied DNN in crop pattern recognition. DNN is not yet been exposed in crop pattern generation or classification. Thus, we choose and apply this new machine learning algorithm along with RF and NB and do a comparative analysis among their performance.

Chapter 3
Study Area and Reference Data

3.1 Study Area

The target area of the study is Suphanburi, located in the Central Plain of Thailand between latitude $14°2'43.72''$ and $15°5'7.72''N$, longitudes $99°17'21.56''$ and $100°17'21.56''E$. Different vegetables and crops like maize, sugarcane etc are produced in this province. Rice is the main crop, evolved a few times in a year [8].

3.2 Reference Data

The experimental data is collected from a report [8]. Suphanburi 2001-2003 of remote sensing Modis data is used in the study. Moderate Resolution Imaging Spectroradiometer (Modis) data is used with three spatial resolutions: 250, 500, and 1000m. About 44 standards of Modis data products are used which include oceanography, biology and atmospheric science. Users who have proper x-band receiving system may capture the regional data. Different libraries are implemented to process NDVI data [9] [10] [11]. Modis 8-days NDVI and Modis 32-days NDVI data are calculated through the ENVI format of the images. There are several steps for performing 32-day fusion data. Firstly, re-project the images to lat/long geographic projection system. Secondly, only the study area has been taken as a subset from the image. It covers latitudes $14°2'43.72''$ and $15°5'7.72''N$, longitudes $99°17'21.56''$ and $100°17'21.56''E$. From each dataset NDVI is calculated. Lastly, eliminate the clouding pixels using LMF tool. The NDVI image data

is also validated by Advanced Very High Resolution Radiometer (AVHRR) data received from NOAA. Some noises are shown in signatures of NDVI image. Quality Controls layer were introduced to discard the tainted pixels and enhance the sample by taking more data. But the high level of noise still remains in the signature. Local Maximum Fitting algorithm [12] [13] is introduced to NDVI data which creates a noise/cloud free data. After that the Artificial Neural Network (ANN) is applied into the data. The following figure 3.1 show the land area of Suphan Buri where the different crops are identified.
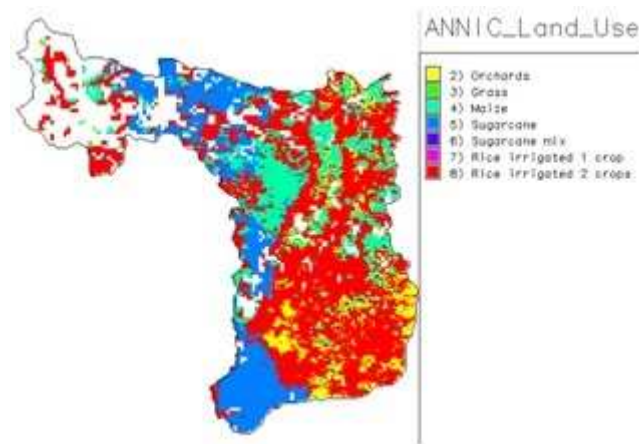

Figure 3.1: ANN classified land use map

3.2.1 NDVI (Normalized Difference Vegetation Index)

Normalized difference vegetation index is a remote sensing method that allows one to display greenness of vegetation. It compares reflectivity of Near-Infrared wavelength and Red wavelength band. The formula of NDVI is, NDVI= NIR-RED/NIR+RED Green leaves strongly reflect NIR (0.7 to 1.1). So, the higher the NIR reflection is the healthier the vegetation. The calculation of NDVI shows a result in a range of +1 to -1. If NIR = 0.5 and RED = 0.8 then NDVI = 0.71 (near to 1) So, it is rainforest. If NIR = 0.4 and

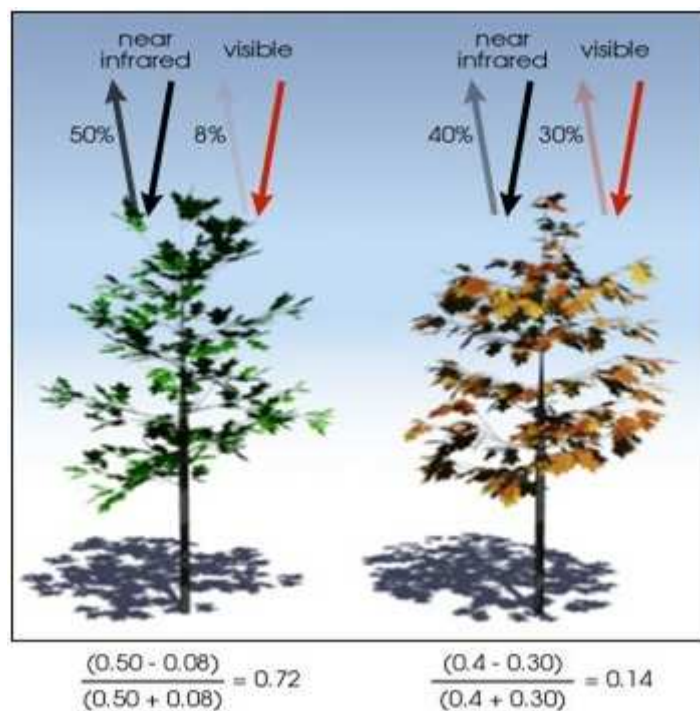RED = 0.30 then NDVI = 0.14 (near to 0) So, it is stressed leave [11].

$$\frac{(0.50 - 0.08)}{(0.50 + 0.08)} = 0.72 \qquad \frac{(0.4 - 0.30)}{(0.4 + 0.30)} = 0.14$$

Figure 3.2: Normalized Difference Vegetation Index [11]

3.2.2 Advanced Very High Resolution Radiometer (AVHRR)

The AVHRR is a broad-band, four or five channel scanner, sensing in the visible, near-infrared and thermal infrared portion of electromagnetic spectrum that can be used for remotely determining cloud cover and the surface temperature. Note that the term surface can mean the surface of the Earth, the upper surface of cloud or the surface of body water. This scanning radiometer uses 6 detectors that collect different band of radiation wavelengths.

The first AVHRR was a 4-channel radiometer, first carried on TIROS-N. This was

subsequently improved to a 5-channel instrument (AVHRR/2) that was initially carried on NOAA-7. The latest instrument version is AVHRR/3, with 6 channels and first carried on NOAA-15 launched in May 1998 [11].

Chapter 4

Proposed Method

In this section we are going to explain the implementation of the machine learning approaches including Naive Bayes(NB), Deep Neural Network (DNN) and Random Forest (RF).

4.1 Naive Bayes

Bayesian approach is a supervised learning technique and uses statistical model for classification. This classification is used as a probabilistic learning method like naive Bayes classifier [14]. In a learning problem, naive Bayes classifier requires a number of parameters linearly in the number of variables (features/predictors). It creates a model which assigns class labels to problem instances, represented as vectors of feature value [14]. In this model, class labels are drawn from finite set. NB classifiers assume that given the class variable the value of a particular feature is independent based on the value of any other feature. For example, if rice has some features like white, round etc. NB considers each of these features independently to find out the probability that it is rice. Bayes theorem general equation [15] is,

$$P(C_k|x) = \frac{P(C_k)P(x|C_k)}{P(x)} \tag{4.1}$$

$P(C_k|x1, x2, ..., xn)$ represents each of K possible outcomes for classes C_k . In Eq. 4.1, the numerator is only significant because the denominator doesn't rely on C [15]. This

numerator is equivalent to the joint probability model which can be written as [15],

$$P(C_k|x_1, x_2, ..., x_n) = \prod i = 1^n P(X_i|C_k) \tag{4.2}$$

and after that we can assume each feature xi is conditionally independent of every other feature x_j for $j! = i$ given the class variable C. So, under this assumption, the conditional distribution over class variable C is,

$$P(C_k|x_1, x_2, ..., x_n) = 1/zP(C_k)\prod_{i=1}^{n} P(x_i|C_k) \tag{4.3}$$

Where, $Z = \sum P(C_k)P(X|C_k)$ is the evidence which is called scaling factor [15]. This scaling factor is dependent on the input variable. Nave Bayes classifier merges the model with a decision rule. One of the rules is picking the most probable value which is known as MAP decision rule [15]. So, if we replace Eq.(3) with class label $\bar{y} = C_k$ [15],

$$\bar{y} = 1/zP(C_k)\prod_{i=1}^{n} P(x_i|C_k) \tag{4.4}$$

Figure 4.1 shows the architectural model of naive Bayes. This algorithm is a probabilistic algorithm which is take decision form the test data. At first, naive Bayes classifier algorithm learns from the training data and creates a data model. Then the model is checked by the testing data and takes the best decision. To run the NB model, there are some defined parameters. First of all, for evaluating the model, dataset is being split in training dataset (70%) and test dataset (30%). The optimization was processed by naivebayes function which was found in library caret package available in the R software. In the naivebayes function, factor variable is declared and training dataset is defined.

4.2 Deep Neural Network (DNN)

Deep learning or deep neural network is a recent topic of machine learning which models the non-linear representations of data. In previous, it is hard to train artificial neural
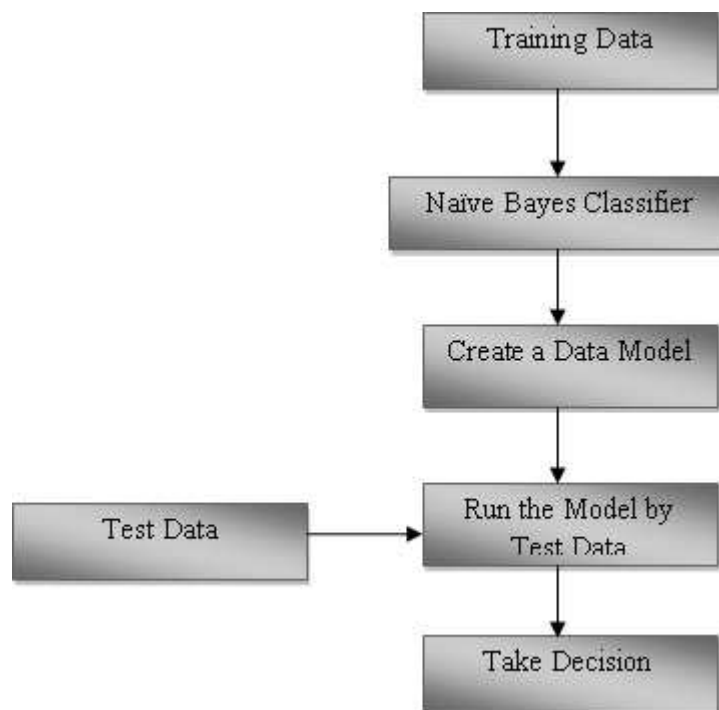
Figure 4.1: Architectural model of Naive Bayes

network because of vanishing gradients and over fitting problems [16]. In deep learning both of these problems has been solved by different activation functions and dropout regularization and a huge amount of data. In deep neural network, multiple layers are being used. Each successive layer uses the output from the previous layer as input. There are many methods to implement deep neural network. Softmax regression and Rectifier Linear Unit (ReLU) are two of them. Softmax regression is a logistic regression which can handle multiple classes. The equation of softmax regression is,

$$\sigma(z)_j = e^{(}z_j)/(\sum_{(} k = 1)^K e^{(}z_j)) \tag{4.5}$$

Where, $j = 1, 2, ..., K$ The output of softmax regression can be used as a probabilistic distribution over K different possible outcomes [17]. ReLU is an activation function which can be defined as, $f(x) = max(0, x)$ where, x is the input of the neuron. To evaluate the deep neural network model, keras (create tensor flow environment) library will be needed for R software [17]. In the model three layers are created. First layer is the input layer with 136 neurons, second and third layers are the hidden layers with 68 neurons and 32 neurons respectively, and the final layer is the output layer with 8 neurons. In the model, ReLU activation function is set for input layer and hidden layer. For calculating the prediction output, softmax regression is set to the model. There are 8 class variables so that the output neuron is eight (8). By optimizing 136 input patterns in the ReLU function, 68 new patterns are created for hidden layer neuron. After that, it predicts the crop by calculating the softmax regression in the output layer. This method is repeated 200 times ($epoch = 200$). After the iterations, the model calculates the average accuracy based on output prediction. As mentioned above the neurons of input layer, hidden layer (two) and output layer in our model are 136, 68, 32 and 8 respectively. Figure shows the sample architectural view of our DNN.

In the figure, the first layer is the input layer with six neurons and the second layer is the first hidden layer with 4 neurons. These neurons are fully connected with one another.
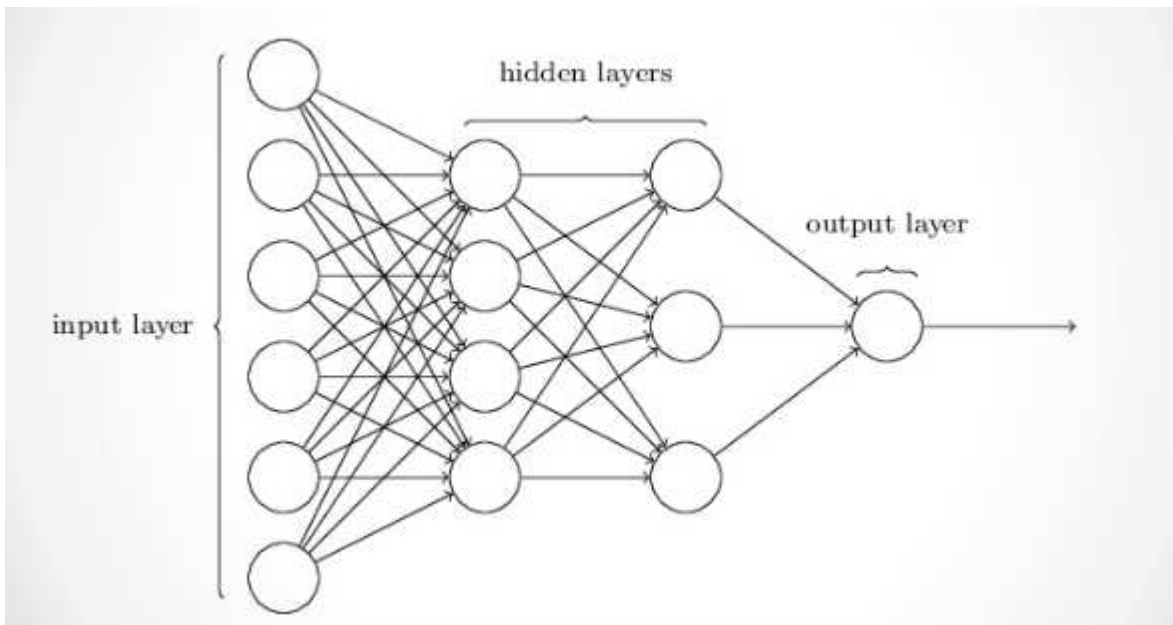
Figure 4.2: Deep Neural Network

The third layer is another hidden layer with three neurons which is also fully connected with the previous layers neuron. And the final layer is the output layer with one neuron. Every layer except the input layer uses output of the every layer as input. Hidden layer evaluates ReLU function to calculate the output of the previous layer. On the other hand, output layer uses softmax function for taking the decision of which pattern it is.

4.3 Random Forest (RF)

For classification and regression random forest algorithm is ensemble learning method which consists of a group of tree-based classifiers. It generates an infinitude random decision trees which then aggregated to compute a classification. For crop pattern classification, random forest is found more efficient. Because it takes a few user-defined parameters and return accuracy level which are comparable or better than Decision Tree [18], Support Vector Machine [19] and Neural Network [20]. To build a forest classification trees (300 trees), RF algorithm randomly samples the dataset with replacement. It uses a bootstrap method to enhance the ability of a classification tree. Bootstrap sample is constructed by each tree to find out the Out-of-Bag (OOB) error. Observations of the main dataset that are left out from the bootstrap sample are declared as Out-of-Bag observations [21]. For measuring prediction error, OOB error is the procedure of RF algorithm. OOB error is calculated using the one-third data from the constructed trees and reciprocates to the rate misclassified variables. OOB error can improve the prediction performance by calculating predictions on the observations which are not used in the learning process.

To run the model of RF algorithm, there are some parameters need to be defined. The main parameters of this model are ntree and mtry. Prediction performance of model will be higher if the value of mtry is increased [22]. So for the model optimization of the parameters mtry and ntree are needed to be considered. First of all, for evaluating

the model, dataset is being split in training dataset (70%) and test dataset (30%). This splitting is repeated ten (10) times to estimate prediction performance. Each time the training and testing data are used to find out the accurate accuracy by averaging all predictions and estimate each variables importance in the classification.

In the model, there are 300 classification trees and mry is 8. To run this model, class variable, training data and testing data is declared in randomForest function. For R software library randomForest has packages like randomForest which first create tree and forest and then train that forest by training dataset (70%). After training the model, it tests the forest by testing dataset and predicts the crop by evaluating the pattern. Figure 4.3 shows the architectural view random forest algorithm. Random
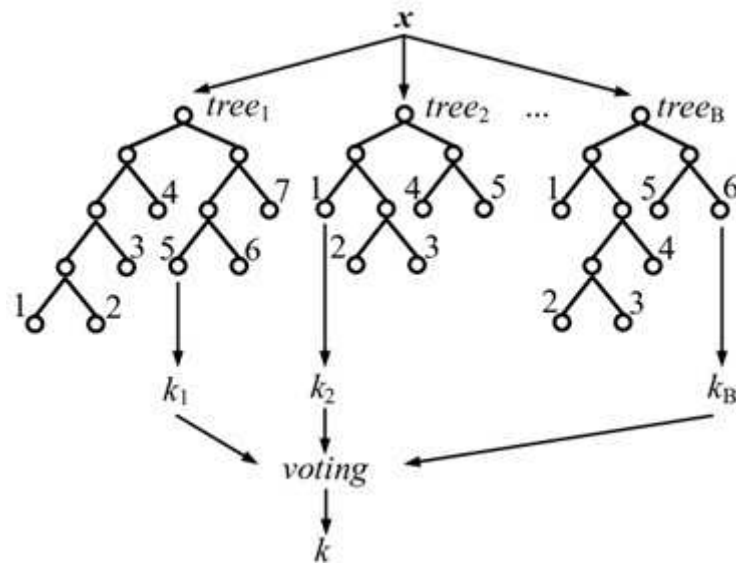

Figure 4.3: Random Forest

forest algorithm creates many small trees and by combining those trees, it creates a forest. In the figure, this algorithm creates b number of trees by x dataset. Every tree has two children and some of the child extends by creating another child. Random forest

takes value from every tree $(k1, k2, ..., kb)$ and after voting these values, it predict the actual value.

These three algorithms calculate the accuracy, Kappa, Sensitivity, Specificity (These are probabilistic statistical measure) by their prediction value. The equations of Kappa, Sensitivity, and Specificity's are discussed below.

- Cohen's Kappa or Kappa is a statistical measure for categorical items. The Kappa statistical value is from 0 to 1 [23]. Most statistical software like R software can calculate Kappa. The formula of Kappa is,

$$Kappa = \frac{P_0 - P_e}{1 - P_e} \tag{4.6}$$

Where, $P0$ = Actual probability and Pe = Random probability.

- Two other statistical measure performances are sensitivity and specificity [24]. These two are binary classification test. Sensitivity is known as true possibility rate that measures the probability of positives which are correctly identified by the model [24]. On the other hand, specificity is also known as true negative rate that measures the probability of negativity which are correctly identified [24]. The formulas of these two statistical measures are,

$$Sensitivity = \frac{TP}{(TP + FN)} \tag{4.7}$$

$$Specificity = \frac{TN}{(TN + FP)} \tag{4.8}$$

Where, TP = True Positive, FP = False Positive, TN = True Negative, FN = False Negative

Chapter 5

Result and Analysis

5.1 Classification Accuracy

We have carried out the implementation in R programming language using RStudio. All the programs were run in 64-bit Intel core i5 2.5 GHz machine. We used medium spatial resolution data for our experiment. Our data is collected from Remote Sensing Images. At the beginning of the implementation we divide the whole data set into two parts. One is the training data set which holds 70% of the full data and the other part is test data which is the 30% of the full data. Several libraries and packages have been used in RStudio to implement these algorithms. All the algorithms show different result which is basically statistical measurement like the overall accuracy, Kappa, sensitivity and specificity.

According to Table 5.2 the accuracy of naive Bayes is 60.53% which is the lowest accuracy. In naive Bayes, the accuracy is calculated based on evidence Z, where class/feature variable is dependent on the input variable X_i .

Table 5.2 presents the accuracy of deep neural network (DNN) 84.89%. Two hidden layers, an input layer and an output layer's are used in our DNN implementation. For our model two hidden layers provide better accuracy than increasing the hidden layer numbers. If we increase the number of hidden layers, the accuracy decreased. For example, we tried to use three hidden layers and it gave an accuracy of 82.75%, for four hidden layers the accuracy is 80.25%. This scenario varies from problem to problem. In

input layer ReLU (Rectifier Linear Unit) function is used to find the probability of desire output. A rectified linear unit has output 0 if the input is less than 0, and raw output otherwise. That is, if the input is greater than 0, the output is equal to the input. The function of ReLU can be expressed as,

$$F(x) = max(x, 0) \tag{5.1}$$

ReLU activation is the simplest non-linear activation function.

Random Forest (RF) algorithm is also applied on the limited data sets and acquires 94.74% accuracy. In RF model we create 300 classification trees which are called ntree. Eight (8) prediction/class variables through mtry are implemented. Then the classification trees are created by using the training data. After that we have created the Random forest with these classification trees. Thereafter, the test features are applied and used the rules of each randomly decision tree to predict the outcome and store them. Then the votes are calculated for each predicted target and consider the high voted prediction target as the final prediction.

To measure the crop identification performance, the Kappa (Cohen's Kappa) co-efficient is calculated (in table 5.3) based on accurate accuracy and random accuracy. In NB algorithm, the Kappa value is 62.01% which falls into the substantial probability (Table 5.1). In DNN, the Kappa accuracy is 82.19% which is nearly perfect probability and in RF, the Kappa probability is 92.88% which also falls into nearly perfect probability.

The sensitivity probability reflects the correctness of the assumption and the specificity reflects the probability to recognize a correct crop. Figure 5.1, 5.2 and Figure 5.3 shows the corresponding sensitivity and specificity of the crops for Nave Bayes, Deep Neural Network, and Random Forest machine learning approaches.

Table 5.1: KAPPA'S STATISTICAL VALUE

0	**Equivalent to chance**
0.1 to 0.20	**Slight probability**
0.21 to 0.40	**Fair probability**
0.41 to 0.60	**Moderate probability**
0.61 to 0.80	**Substantial probability**
0.81 to 0.99	**Nearly probability**
1	**Perfect probability**

Table 5.2: ACCURACY OF MACHINE LEARNING METHODS

Algorithm Name	**Accuracy %**
Naive Bayes	**60.53**
Deep Neural Network	**84.89**
Random Forest	**94.74**

Table 5.3: KAPPA CO-EFFICIENT FOR MACHINE LEARNING METHODS

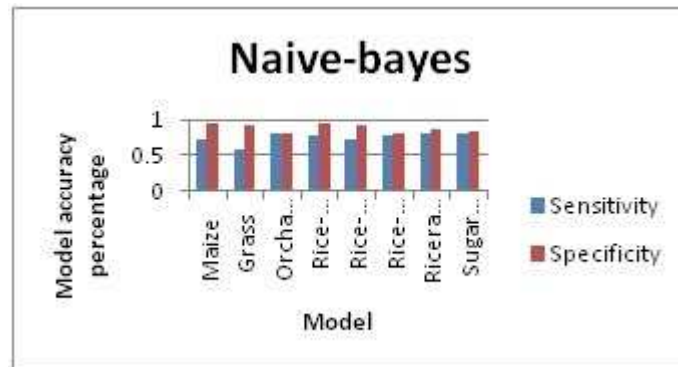| **Algorithm Name**	**Kappa %**
Naive Bayes	**62.01**
Deep Neural Network	**82.19**
Random Forest	**92.88**

Figure 5.1: Sensitivity and Specificity model of Naive Bayes
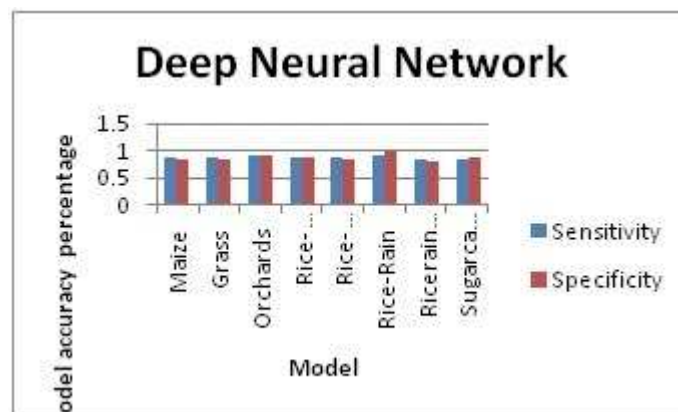

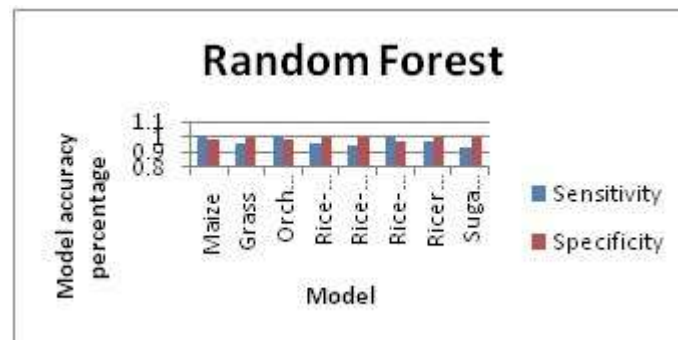Figure 5.2: Sensitivity and Specificity model of Deep Neural Network

Figure 5.3: Sensitivity and Specificity model of Random Forest

5.2 Optimized Method

The above observation, we shows that the statistical measurement of deep neural network and random forest are very imminent. If we analyze the Kappa accuracy, we can see both the accuracy falls into same range (0.81 to 0.99) which is nearly perfect probability but the overall accuracy is high in Random forest which is 94.74%. There are some features for which random forest is giving higher accuracy. These are,

- It can handle thousands of input variables without variable deletion

- It gives estimate of what variables are important in the classification

- It generates an internal unbiased estimate of the generalization error as the forest building progresses

- It has methods for balancing error in class population unbalanced data sets

- Prototypes are computed that gives information about the relation between the variables and the classification

- It offers an experimental method for detecting variable interactions

Besides random forest works better because if there are any missing values, the random forest algorithm calculates an Out-of-Bag (OOB) error and then again uses them for classification. Furthermore, although we have insufficient observation and high data dense random forest gives nearly perfect accuracy. So, from above discussion, we can say random forest can be an optimized method for crop pattern recognition.

Chapter 6

Conclusion

This study examines that RF algorithm shows the better result and can be a prominent crop pattern classification method.More robust data is used to get the appropriate classification [8]. Three types of algorithms are implemented - NB, DNN, RF.Among the algorithms, RF manifests the best performance and highest accuracy (94.74%). If we could train the algorithms with modern and larger dataset, the accuracy would be gained higher. The theory behind the algorithms implemented has been correctly handled with proper care.Data has been prepared perfectly and missing values are trained appropriately to handle the errors. DNN performs well with larger amount of dataset. It can be the reason of getting less accuracy after implementing DNN.

Our main target of this research is to train the machine with different algorithms and find out which algorithms can perform better in a certain situation. With these types of research RF can utensil. To locate and identify the rotten crops or which crop is produced well in the certain location can be helpful with our findings. RF classifies the data properly and can distinguish the main crop in the study area. This approach can be applied in the larger area where the data is sufficient enough. However, the process is utilized to gain the highest possible outcome.

Bibliography

[1] "Top 10 agricultural countries by gdp contribution 2017." [Online]. Available: https://www.mbaskool.com/fun-corner/top-brand-lists/17226-top-10-agricultural-countries-by-gdp-contribution-2017.html

[2] C. M. Bishop, "Pattern recognition and machine learning," 2018.

[3] C. V. Nicholson, "Introduction to deep neural networks (deep learning) - deeplearning4j: Open-source," 2013. [Online]. Available: https://deeplearning4j.org/neuralnet-overview

[4] V. Lebourgeois and S. Dupuy, "A combined random forest and obia classification scheme for mapping smallholder agriculture at different nomenclature levels using multisource data (simulated sentinel-2 time series, vhrs and dem)," 2018.

[5] K. Tatsumi and Y. Yamashiki, "Crop classification of upland fields using random forest of time-series landsat 7 etm+ data," 2018.

[6] R. Rustowicz, "Crop classification with multi-temporal satellite imagery," 2018.

[7] A. Patil and L. Y S, "Classification of crops using fcm segmentation and texture, color feature," 2018.

[8] "Statistics how to. (2018). cohens kappa statistic." [Online]. Available: http://www.statisticshowto.com/cohens-kappa-statistic

[9] K. G. R. N. A. S. A. Y. C. Sajia Akhter, Ipshita Sarkar and H. Kiyoshi, "Adapting the lmf temporal splining procedure from serial to mpi/linux clusters, journal of computer science 3(3): 130-133, 2007,science publication, new york, usa."

[10] Y. C. Shamim Akhter, Kento Aida, "Grass gis on high performance computing with mpi, openmp and ninf-g programming framework, proceedings of isprs 2010 , 9-12th august kyoto, japan."

[11] "Advanced very high resolution radiometer - avhrr - noaa satellite information system (noaasis); office of satellite and product operations." [Online]. Available: http://noaasis.noaa.gov/NOAASIS/ml/avhrr.html

[12] K. A. Shamim Akhter, Yann Chemin, "Porting a grass raster module to distributed computing, examples for mpi and ninf-g, osgeo journal vol2.september 2007,pp36-44 issn 1994-1897.2006."

[13] Y. C. Shamim Akhter, Kento Aida, "Vegetation index maps of asia temporally splined for consistency through a high performance and grid system, proceedings of 3rd international conference on software and data technologies (icsoft 2008), porto, portugal, 5-8 july,2008."

[14] "Naive bayes classifier." [Online]. Available: http://www.statsoft.com/textbook/naive-bayes-classifier

[15] B. A. Olshausen, "Bayesian probability theory," 2004.

[16] "Deep learning in r." [Online]. Available: https://www.r-bloggers.com/deep-learning-in-r-2

[17] D. Honda and M. Chemin, "Near real time agriculture activity monitoring using multitemporal modis earth observation satellite data."

[18] M. Pal, "Random forest classifier for remote sensing classification. int. j. remote sens. 2005, 26,," p. 217.

[19] R. D. R. Pringle, M.J.; Denham, "Identification of cropping activity in central and southern queensland, australia, with the aid of modis mod13q1 imagery. int. j. appl. earth obs. geoinf. 2012," p. 276.

[20] K. A. Shamim Akhter and Y. Chemin, "Asynchronous grid lmf processing of vegetation index for asia, international journal of geo-informatics. vol 4, no 4."

[21] C. A. Oliveira, S. and J. Pereira, "Modeling spatial patterns of fire occurrence in mediterranean europe using multiple regression and random forest," 2018.

[22] "Medium. (2018). how random forest algorithm works in machine learning." [Online]. Available: https://medium.com/@Synced/how-random-forest-algorithm-works-in-machine-learning-3c0fe15b6674

[23] "Sensitivity and specificity." [Online]. Available: https://onlinecourses.science.psu.edu/stat507/node/71

[24] S. S. R. Lawrence, R.L.; Wood, "Mapping invasive plants using hyperspectral imagery and breiman cutler classifications (randomforest). remote sens. environ. 2006, 100," p. 356.

Appendix A

List of Acronyms

RS	Remote Sensing
ANN	Artificial Neural Network
DNN	Deep Neural Network
NB	Naive Bayes
RF	Random Forest
OBIA	Object-Based Classification
EVI	Enhanced Vegetation Index
SVM	Support Vector Machine
CNN	Convolution Neural Network
FCM	Fuzzy C-Means
MODIS	Moderate Resolution Imaging Spectroradiometer
NDVI	Normalized Difference Vegetation Index
AVHRR	Advanced Very High Resolution Radiometer
NIR	Near-Infrared
ReLU	Rectifier Linear Unit
OOB	Out-of-Bag