Crop Recommendation System by Artificial Neural Network

Bangaru Kamatchi S (gdekamatch77777@yahoo.com)
Vellore Institute of Technology: VIT University
R. Parvathi
vit

Research Article

Keywords: Artificial Neural Network, Crop Classification, Climatic conditions, machine learning algorithms

Posted Date: September 16th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-874525/v1

Abstract

The agriculture yield mostly depends on climate factors. Any information associated with climatic factors will help farmers in foreordained farming. Choosing a right crop at right time is most important to get proper yield. To help the farmers in decision making process a classification model is built by considering the agro climatic parameters of a crop like temperature, relative humidity, type of soil, soil pH and crop duration and a recommendation system is built based on three factors namely crop, type of crop and the districts. Predicting the districts is the novel approach in which crop pattern of 33 districts of Tamilnadu is marked and based on that classification model is built. Thorough analysis of machine learning algorithms incorporating pre-processing, data augmentation and comparison of optimizers and activation function of ANN. Log loss metric is used to validate the models. The results shows that artificial neural network is the best predictive model for classification of crops crop type and district based on agrometeorological climatic condition. The accuracy of artificial neural network model is compared with five different machine learning algorithms to analyse the performance.

I Introduction

Agricultural organizations are battling an important issue of prediction of yield. Every farmer is eager to know the amount of yield they can expect from production. Attempts to solve this problem date back to the time when farmers first began planting crops. However, agricultural productivity is very poor compared to the food demand. Scholars, producers, agricultural scientists and governments are seeking to bring more effort and processing technologies into the field.

The machine learning algorithms are now widely used in different fields of agriculture like yield prediction, soil classification, attribute selection of crops, disease identification, crop selection, pattern identification and individual crop identification. The reason that such strategies and massive datasets were linked was that computing space and computational capacity were cheap. In mining process, every piece of data is converted into information by finding their relation and occurrence of patterns which makes the process of decision making more successful.

In agriculture, decision making begins from the stage of sowing seed and continues through every crop growth stage. For example, the decision of whether a particular crop can be sowed in a particular season can be performed by the data mining techniques. The details of each crop and its problems during growth will help in analyzing the nature of the crop and avoid losses in future. Data mining is used to extract some information from the vast volume of data or data collection. As information is usually concealed in massive repositories, data mining methods can be used to uncover and turn it into usable data. This study examines the different predictive models based on data mining (machine learning) techniques for the classification of crops. The paper is divided into four sections with a brief description of related work forming the first section. It is followed by the methodology and results of the study. The final section provides the main conclusions derived from the study.

I I Related Work

Studies on classifying crops based on machine learning algorithms have been carried out more than a decade. The random forest (RF) method is most important algorithm in crop classification for pixel based images (Ok, Akar, & Gungor, 2012). The findings of the study of Ok et al. (2012) depicted that the results for every RF system parameter combination were identical; suggesting the accuracy of the RF classification algorithm. With the aid of auxiliary knowledge, a small and intelligently selected training dataset was developed in the study of Mathur and Foody (2008) which included information from training sites. This was expected to be among the most insightful for an SVM classification prior to the process of classification (Mathur & Foody, 2008). The complexity of the algorithm of FP-tree growth was found to be lesser compared to other methods as the database searches were least frequent. In addition, it was stored in the form of a data cube in order to concisely track, evaluate and optimally distribute (Khan, 2014). Important connections or ties among a wide range of data items can be found in the association rules of mining. Most companies involved in mining industry utilize large servers with vast volumes of data gathered and processed constantly (Thakkar, Kayasth, & Desai, 2014).

The distribution of farm land is characterized by strong differences over a fairly short period of time. Although these patterns are difficult to classify, they offer important details to enhance classification efficiency. Phenologic seed styles sequence patterns (PSPs) focused on a thick pile of Sentinel 1 information and detailed plant phenology knowledge (Bargiel, 2017). A domain-specific dataset called CropDeep containing vegetables and fruits closely associated with PA, to support the identification and detection of different crops that define agricultural missions has also been utilized. At present, it includes 30 types of growing vegetables and fruit obtained by IoT visual sensors, autonomous robots and greenhouse smartphones (Zheng et al., 2019).

In attribute selection, the agricultural knowledge base consists of farming knowledge, such as regional identification, region-name, environmental parameter, area, pesticides, cultivar knowledge and seed size. It also includes crop samples with appropriate field expertise and environmental criteria (Medar & Rajpurohit, 2014). To estimate the soil dataset type, analyzed A classification rule is formalized to estimate crop yield, where methods used were Naive Bayes and K-Nearest Neighbor (Paul, Vishwakarma, & Verma, 2016).

I I I Methodology

General Architecture Overview

In the research on classification of plants from the manually collected data, two-level architecture was proposed. This levels consist of supervised crop classification and crop sort, including the best algorithm for classification, optimisation and activation. These levels are supervised. Pre-processing is another critical phase in the research. Since data has been obtained from multiple sources, issues, such as numerous data units and missing values, can emerge. The next step, the regional prediction, is the central part of the work on an improved classification model of the neural network. The techniques used to maximize the amounts of data in which the data augmentation process is carried out in order to treat data that is increasing for processing are the addition of slightly updated copies of existing or newly generated data. In planning a machine, it helps reduce exercise. Figure 1. The two levels of architecture are explained separately in order to enable farmers to follow the same style in Figure 2 in relation to cultivation, crop type and similar districts.

Study area and materials

The state of Tamil Nadu in India was selected as the study area. The most demanding and difficult job is to gather data in agriculture. Comprehensive analysis results, case studies, the state's official agriculture website, and other websites were gathered for the report. The study contained 106 specimens of vegetable crops and 61 other crops, including cereals, millets, pulses, fiber oilseeds, sugar and drilled crops. This study was focused on 106 samples of plants. But seasonal crops were not included in the current report. The classification model attributes selected are given in Table 1. Agro climate parameters that have a high effect on the crop forecast. The crop model is obtained and graded based on crop seeding for 33 districts in Tamilnadu. The data collected is published in mendeley data http://dx.doi.org/10.17632/zyyb98msjc.1

Table 1

Attributes of crop

Attribute of crop	Features
Type of Crop	Cereals, Millets, Pulses, Oil seeds, Fibre crops, Sugar crops, Forage crops, Cole crops, Vegetables, Root & Tuber, Green & Leafy, Bulb vegetables, Minor vegetables and Other crops
Type of Soil	Alluvia, Loamy, Sandy, Clayey, Black, Red, Sandy loamy, Black cotton soil, Clay loamy, Well-drained loamy, Heavy cotton, Silt loams, Well-drained sandy, Lateritic, Friable, Well-drained
Soil Ph	Low and High
Duration of crop	Maximum and Minimum
Temperature	Maximum and Minimum
Relative humidity	Low and High
Districts	29 distincts of Tamilnadu

Feature Selection

The goal of the collection of attributes is to check for a valuable collection of comparable attributes if all the attributes are used, the classification outcomes would be the case. Moreover, a smaller collection of attributes generate less complex patterns which humans can readily grasp and even imagine. Random forest algorithm is used for selecting the features. Random forests are a mixture of tree predictors, which allows each tree to be individually and uniformly distributed by the value of a random variable sampled over all trees in the wood. The forest generalization error converges with as. to a cap that increases the number of trees in the forest. The error of generalization of tree classifiers forests depends on the power and interaction of the individual trees in the forest. The use of a random range of functionality to separate per node creates error rates that are more stable with regard to noise than Adaboost. Here the attributes with score less than 1.5 is not considered but all the attributes having low score hence all attributes are selected to perform classification which is shown in figure 2.

Pseudocode

```
DecisionTreeObj = DecisionTree(MutatedDataset)

for feature_score in DecisionTreeObj.feature_scores:

   if (feature_score < np.std(DecisionTreeObj.feature_scores) * 1.5):

      MutatedDataset.drop(feature)
```

Pre-processing and Data Augmentation

The quality of input data can be improved through data pre-processing, also as data preparation, which in turn affects performance and analytical efficiency of the results. In this step, the data was converted into the similar format to enhance features. The following processes were performed: The unit of temperature for some crops were in Fahrenheit and some crops were in Celsius; hence, everything was converted to the same unit. For experimentation purpose, Data augmentation process is carried out based on minimum and maximum values, each crop data row was duplicated into fifteen rows with 0.1 increase and decrease in their values for analysis.For missing values in a data set, an average value was used based on available data.

Pseudocode

Input: 106 crops with their actual values

Output: other than categorical values all other parameters are increased by 0.1 based on minimum and maximum value and thereby creating 15 samples for each and every crop

---

```
for row in range(len(dataset))
    for delta in range(0,1.5,0.1):
        MutatedDataset = MutatedDataset.append(dataset.iloc[row] + delta)
for delta in range(0,-1.5,-0.1):
        MutatedDataset = MutatedDataset.append(dataset.iloc[row] + delta)
return (x_train, x_test, y_train, y_test)
```

Supervised Classification

The predictive properties of the models were analysed by classifying the crops based on soil and agrometeorological conditions and type of crop. Agriculture depends on the climate and soil conditions. Further, crops usually fall under the general categories such as cereals, pulses, millets, etc. Therefore, classification based on agrometeorological conditions and type of crop helps in recommending new and hybrid variety of crops to farmers. Various machine learning algorithms are used to classify the crops to determine the best model. From literature survey, six different algorithms have been considered for experimentation which were baseline model, decision tree, linear model, random forest, XGBoost and neural network model. Baseline model acts as a reference point for comparison with other models. They help in exact assessment of the properties of other machine learning methods. This model can be framed using regression error curves, mean and median of data, etc. (Whigham et al., 2015). Decision tree is a useful data mining and machine learning method for complicated data which can be alphabetical, numerical and nominal. A node is created based on the “information gain” approach determined by the attributes (Somvanhi et al., 2016). Linear model employs regressors from new and existing independent variables. This model, however, requires maintaining a balance between bias-variance and overfitting of data to obtain optimum results (Wilson & Sahinidis, 2017). XGBoost is one of the prominent data mining tools that incorporates features of many related techniques such as CPU multithreading for parallelism, processing of scant data as in decision tree and handling of huge amount of data at faster speeds as processed by block technology (Lu & Ma, 2020). Random forest is a method of “ensemble learning” based on decision tree model of machine learning. In this method, a predictor of random sample is used prior to segmentation of node; thereby decreasing the bias. The advantages of this model are introduction of two random elements, analysis of higher dimensions of data and quicker training periods (Lu & Ma, 2020). Artificial Neural Network (ANN), based on human brain’s biological neural processes, learns to recognize the patterns or relationships in the data by observing a large number of input and output examples through training.

The following tables depict the comparison of the classification models based on agrometeorological conditions and type of crop.   To build a model choosing best algorithm is very important and to choose the algorithm here 5 different classification is built and best algorithm is chosen based on the minimal error value. The classification models is built for both individual crop and type of crop and the results are depicted in Figure 3 and 4 . The autoML function in python is used for anlayzing in which ensemble model will choose the better algorithm and build a new ensemble model but for this case the artificial neural network alone is chosen by the ensemble model which shows that ANN works better for t both the crop and crop type.The results show that the neural network model has a higher train time of 523.03 seconds and a logloss value of 0.00034 for agrometeorological conditions and 648.54 seconds and low logloss value of 0.00018 for type of crop with than other algorithms with the minimal logloss value. Therefore, it can be perceived that artificial neural network provides the best predictive model for crop classification. Optimizers vary the attributes of the neural networks to decrease losses due to errors. Optimizers adam, Nadam, SGD, adagrad, adadelta, Adamax and rmsprop were used. In addition, activation functions such as Relu, leaky Relu, PRelu, Elu were used for quicker training period of the network. Also, for both the classification processes, logloss metric has been used to validate the model.Among the models Relu and SGD provides the better results .

A learning curve displays an estimator's validation and testing score for various training samples. It is a tool to see how much more training data are open to us and if a variance error or a bias error is more common in the estimator.Here the learning curve of both individual crop and crop type is illustrated in figure 5 and 6 which shows that a better model is built.

A learning curve is a link between the success of an instructor in a job and the amount of attempts or the time taken to carry out a task.

$$Y=aX^{b} \quad (1)$$

Where:
Y is the average time over the measured duration
a represents the time to complete the task the first time
X represents the total amount of attempts completed
b represents the slope of the function

District Prediction

The prediction of districts is the novel approach in which considering all the parameters of the crop and the district pattern of crop .The presence of a crop in a districts is marked as 1 and other than as 0.The district labels are factorized into the feature vectors for which one hot encoder is used in python ,then an artificial neural network model is created to classify and predict the districts in which grid search method is used to search the hyperparameters to build a strong classification model and also districts are represented in the categorical format to get output for all 33 districts. The pseudocode in which neural network model is built is as follows

Pseudocode

Create an ANN for district recommendation

Input layer (Crop, Crop Type and Filtered features using Decision Tree for important features),

Hidden layer (3, to optimize weights) and

Output layer (Represents No. of Districts as categorical output)

X = [Crop, Crop_Type, Selected_Features]

Y = [District_1..District_33]

nn = sequential()

nn.add(input_layer)

nn.add(layer_1)

nn.add(layer_2)

nn.add(layer_3)

nn.add(output_layer)

nn.compile()

nn.fit(X,Y)

Use Grid search method to search for best hyperparameters to train the network

GridSearchObj = GridSearch(nn)

suggested_parameters = GridSearchObj.best_performance

Utilize the obtained parameters to build final network and predict

nn.compile(suggested_parameters)

nn.fit(X,Y)

district_recommendation = nn.predict(test)

Crop Recommendation System

The crop recommendation system which is depicted in figure 2 is an artificial neural network model in which it has the combination of individual crop ,crop type and districts . The neural network structure is also known as its 'architecture' or 'topology.' The number of layers consists of primary units. It also provides a weight change system for interconchanging. The selection of the structure influences the outcomes. It is the most important aspect of the neural network implementation. The predictive strength of the neural network increases by adding 1 or more hidden layers to the input and output layers and units in this layer. Yet as minimal as possible a variety of hidden layers. This means that the neural network does not store all learning knowledge, but can generalize it to prevent overfits. while building of these classification models the epoch loss is consistently getting reduced and the epoch accuracy is getting increased which is shown in the figure 7,8 and 9 and also

I V Conclusion

As crop growth is basically based on the climatic and soil conditions, it is vital to consider the climatic factors of a crop for crop recommendation. Mostly, failure in agriculture, other than natural calamities, is due to poor choice of crops or faulty prediction of weather. This can be avoided by using this recommendation system. In this paper the agrometeorological condition of a crop is taken into consideration for building better classification model. Artificial neural network method which produces the minimal error value when compared to all other classification algorithm and also different optimizer and activation function in ANN model is analysed .

Declarations

1. Availability of data

Data availability

Using dataset in this paperAdaboost

2. Declaration statements

The optimizer and activation function in ANN modelthe authors declare that they have no consent.

3. Funding

No funding

4. Conflicts of interest

No conflict of interest

5. Code availability

MATLAB

6. Authors' contributions

Based on the classification results produced the ANN is taken to build the crop recommendation model.

References

1. seethapathy, & AGRO-METEOROLOGICAL DATA OF INDIAN STATE TAMIL NADU"., R, Parvathi. (2020). *Mendeley Data*. V1, doi: 10.17632/zyyb98msjc.1
2. Ahamed, A. T. M. S., et al. (2015). 'Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh', *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings*. doi: 10.1109/SNPD.2015.7176185.
3. Bargiel, D. (2017). 'A new method for crop classification combining time series of radar images and crop phenology information'. *Remote Sensing of Environment* (198 vol., pp. 369–383). Elsevier Inc.. doi: 10.1016/j.rse.2017.06.022.
4. Hao, P., et al. (2015). 'Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA'. *Remote Sensing*, 7(5), 5347–5369. doi: 10.3390/rs70505347.
5. Khan, F. (2014). 'Association Rule Mining in the field of Agriculture: A'. *International Journal of Scientific and Research Publications*, 4(7), 1–4
6. Mathur, A., & Foody, G. M. (2008). 'Crop classification by support vector machine with intelligently selected training data for an operational application'. *International Journal of Remote Sensing*, 29(8), 2227–2240. doi: 10.1080/01431160701395203.
7. Medar, R. A., & Rajpurohit, V. S. (2014). 'A survey on Data Mining Techniques for Crop Yield Prediction'. *International Journal of Advance Research in Computer Science and Management Studies*, 2(9), 59–64
8. Ok, A. O., Akar, O., & Gungor, O. (2012). 'Evaluation of random forest method for agricultural crop classification'. *European Journal of Remote Sensing*, 45(1), 421–432. doi: 10.5721/EuJRS20124535.

9. Paul, M., Vishwakarma, S. K., & Verma, A. (2016). 'Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach', *Proceedings – 2015 International Conference on Computational Intelligence and Communication* Networks, *CICN 2015*, pp. 766–771. doi: 10.1109/CICN.2015.156.
10. Thakkar, R. G., Kayasth, M., & Desai, H. (2014). 'Rule Based and Association Rule Mining On Agriculture Dataset'. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 6381–6384
11. Zheng, Y. Y., et al. (2019). 'Cropdeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture'. *Sensors (Switzerland)*, 19(5), doi: 10.3390/s19051058.
12. Ahamed, A. T. M. S., et al. (2015). 'Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh', *2015 IEEE/ACIS 16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings*. doi: 10.1109/SNPD.2015.7176185.
13. Bargiel, D. (2017). 'A new method for crop classification combining time series of radar images and crop phenology information'. *Remote Sensing of Environment* (198 vol., pp. 369–383). Elsevier Inc.. doi: 10.1016/j.rse.2017.06.022.
14. Hao, P., et al. (2015). 'Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA'. *Remote Sensing*, 7(5), 5347–5369. doi: 10.3390/rs70505347.
15. Khan, F. (2014). 'Association Rule Mining in the field of Agriculture: A'. *International Journal of Scientific and Research Publications*, 4(7), 1–4
16. Mathur, A., & Foody, G. M. (2008). 'Crop classification by support vector machine with intelligently selected training data for an operational application'. *International Journal of Remote Sensing*, 29(8), 2227–2240. doi: 10.1080/01431160701395203.
17. Medar, R. A., & Rajpurohit, V. S. (2014). 'A survey on Data Mining Techniques for Crop Yield Prediction'. *International Journal of Advance Research in Computer Science and Management Studies*, 2(9), 59–64
18. Ok, A. O., Akar, O., & Gungor, O. (2012). 'Evaluation of random forest method for agricultural crop classification'. *European Journal of Remote Sensing*, 45(1), 421–432. doi: 10.5721/EuJRS20124535.
19. Paul, M., Vishwakarma, S. K., & Verma, A. (2016). 'Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach', *Proceedings – 2015 International Conference on Computational Intelligence and Communication* Networks, *CICN 2015*, pp. 766–771. doi: 10.1109/CICN.2015.156.
20. Thakkar, R. G., Kayasth, M., & Desai, H. (2014). 'Rule Based and Association Rule Mining On Agriculture Dataset'. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 6381–6384
21. Zheng, Y. Y., et al. (2019). 'Cropdeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture'. *Sensors (Switzerland)*, 19(5), doi: 10.3390/s19051058.
22. Ahamed, A. T. M. S., et al. (2015). 'Applying data mining techniques to predict annual yield of major crops and recommend planting different crops in different districts in Bangladesh', *2015 IEEE/ACIS*

16th International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing, SNPD 2015 - Proceedings. doi: 10.1109/SNPD.2015.7176185.

23. Bargiel, D. (2017). 'A new method for crop classification combining time series of radar images and crop phenology information'. *Remote Sensing of Environment* (198 vol., pp. 369–383). Elsevier Inc.. doi: 10.1016/j.rse.2017.06.022.
24. Hao, P., et al. (2015). 'Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA'. *Remote Sensing*, 7(5), 5347–5369. doi: 10.3390/rs70505347.
25. Khan, F. (2014). 'Association Rule Mining in the field of Agriculture: A'. *International Journal of Scientific and Research Publications*, 4(7), 1–4
26. Mathur, A., & Foody, G. M. (2008). 'Crop classification by support vector machine with intelligently selected training data for an operational application'. *International Journal of Remote Sensing*, 29(8), 2227–2240. doi: 10.1080/01431160701395203.
27. Medar, R. A., & Rajpurohit, V. S. (2014). 'A survey on Data Mining Techniques for Crop Yield Prediction'. *International Journal of Advance Research in Computer Science and Management Studies*, 2(9), 59–64
28. Ok, A. O., Akar, O., & Gungor, O. (2012). 'Evaluation of random forest method for agricultural crop classification'. *European Journal of Remote Sensing*, 45(1), 421–432. doi: 10.5721/EuJRS20124535.
29. Paul, M., Vishwakarma, S. K., & Verma, A. (2016). 'Analysis of Soil Behaviour and Prediction of Crop Yield Using Data Mining Approach', *Proceedings – 2015 International Conference on Computational Intelligence and Communication* Networks, *CICN 2015*, pp. 766–771. doi: 10.1109/CICN.2015.156.
30. Thakkar, R. G., Kayasth, M., & Desai, H. (2014). 'Rule Based and Association Rule Mining On Agriculture Dataset'. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(11), 6381–6384
31. Zheng, Y. Y., et al. (2019). 'Cropdeep: The crop vision dataset for deep-learning-based classification and detection in precision agriculture'. *Sensors (Switzerland)*, 19(5), doi: 10.3390/s19051058.

Figures

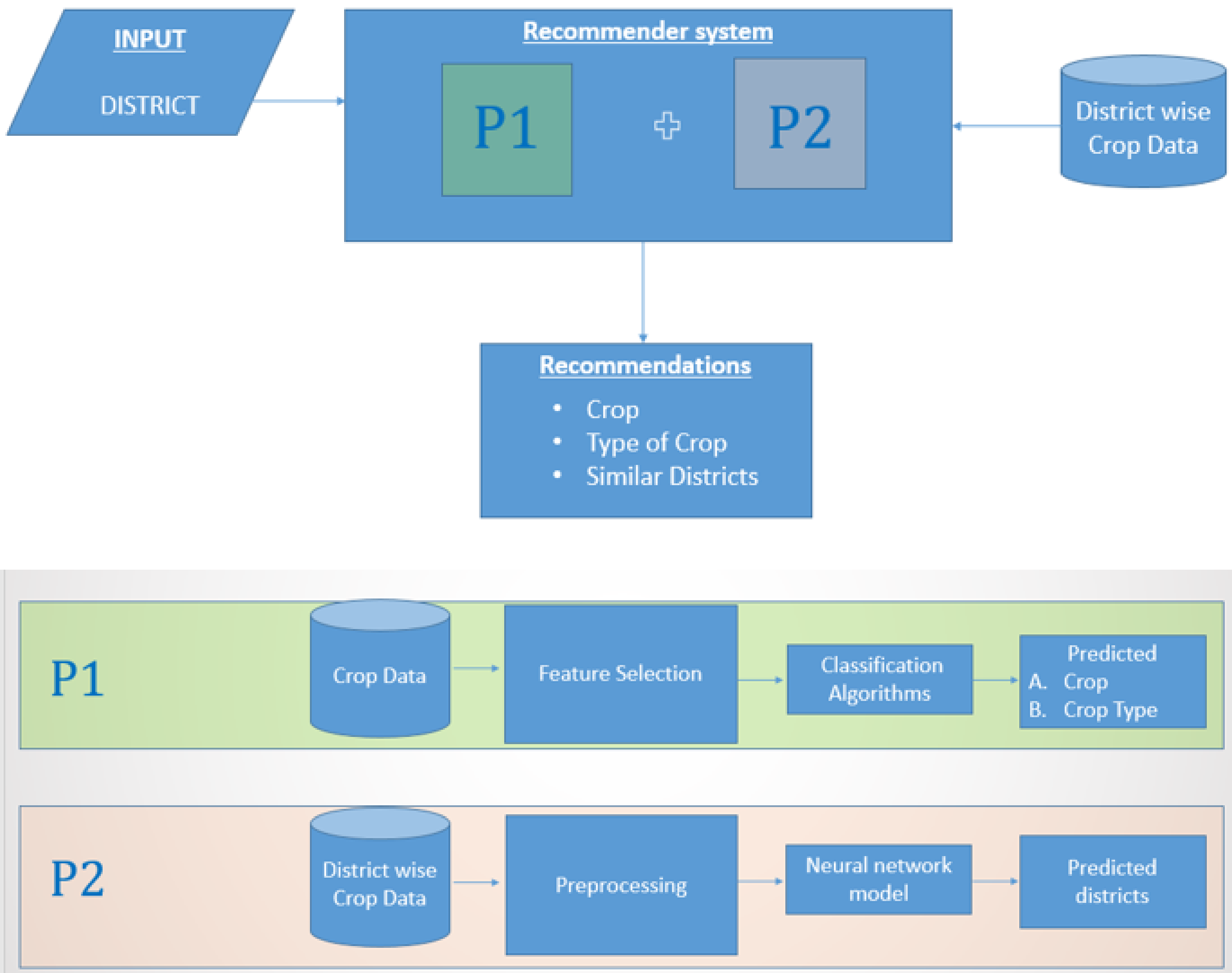

Figure 1

Crop recommendation system

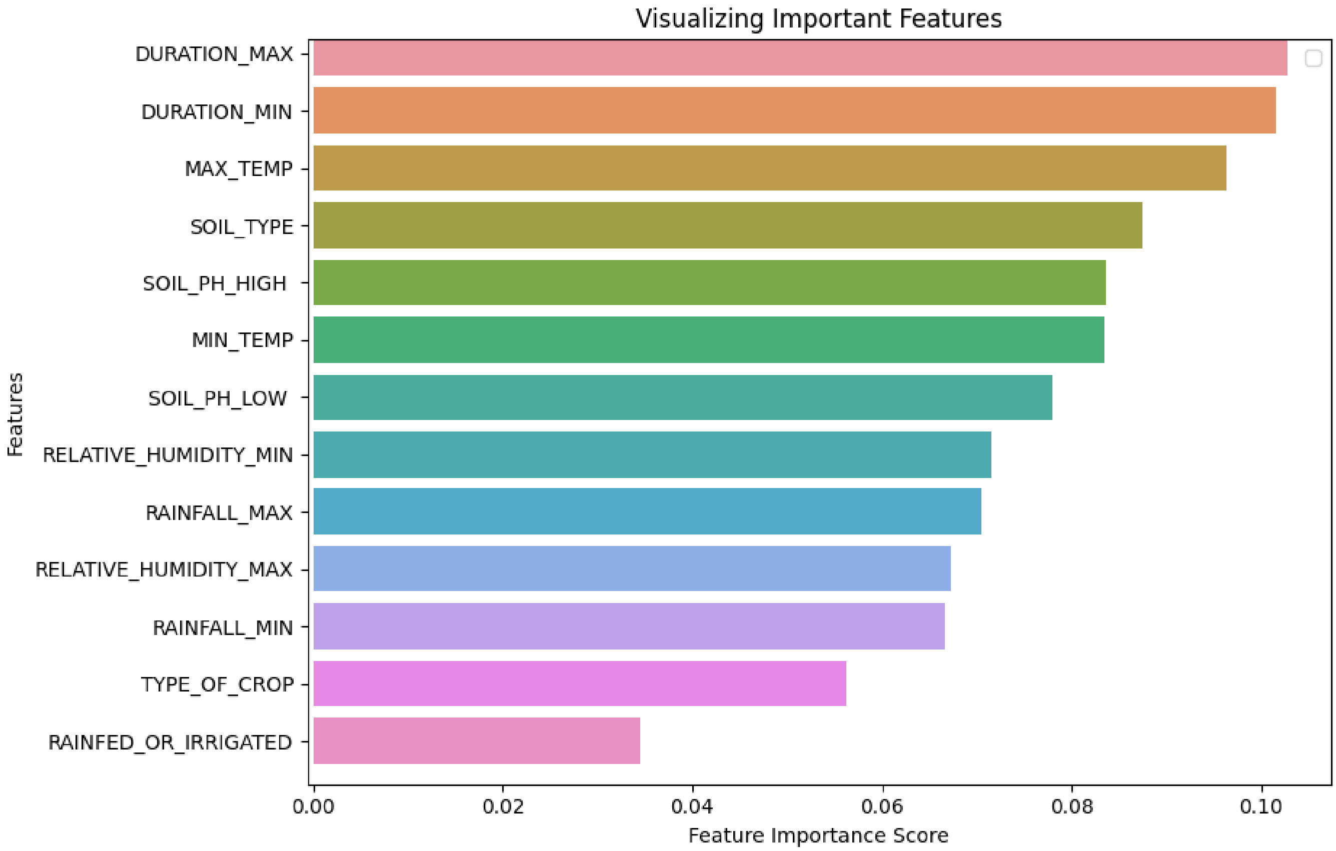

Figure 2

Feature selection using random forest

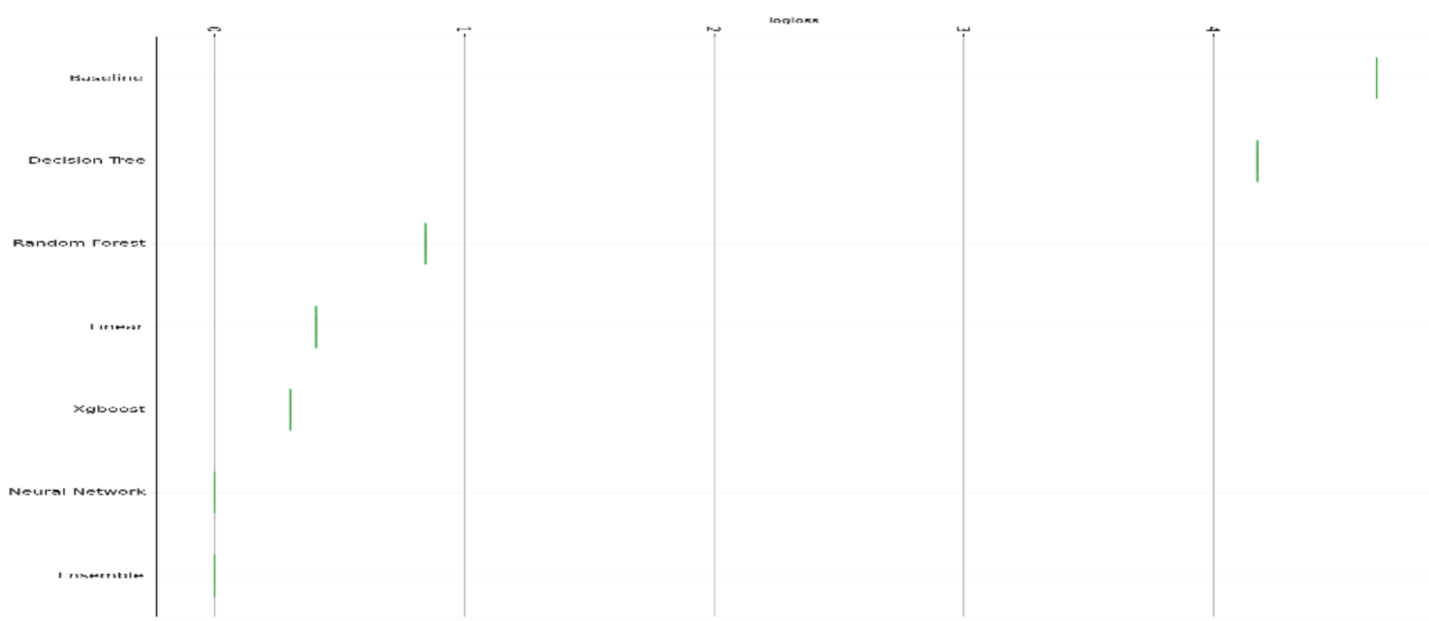

Figure 3

Comparison of classification model for individual crop

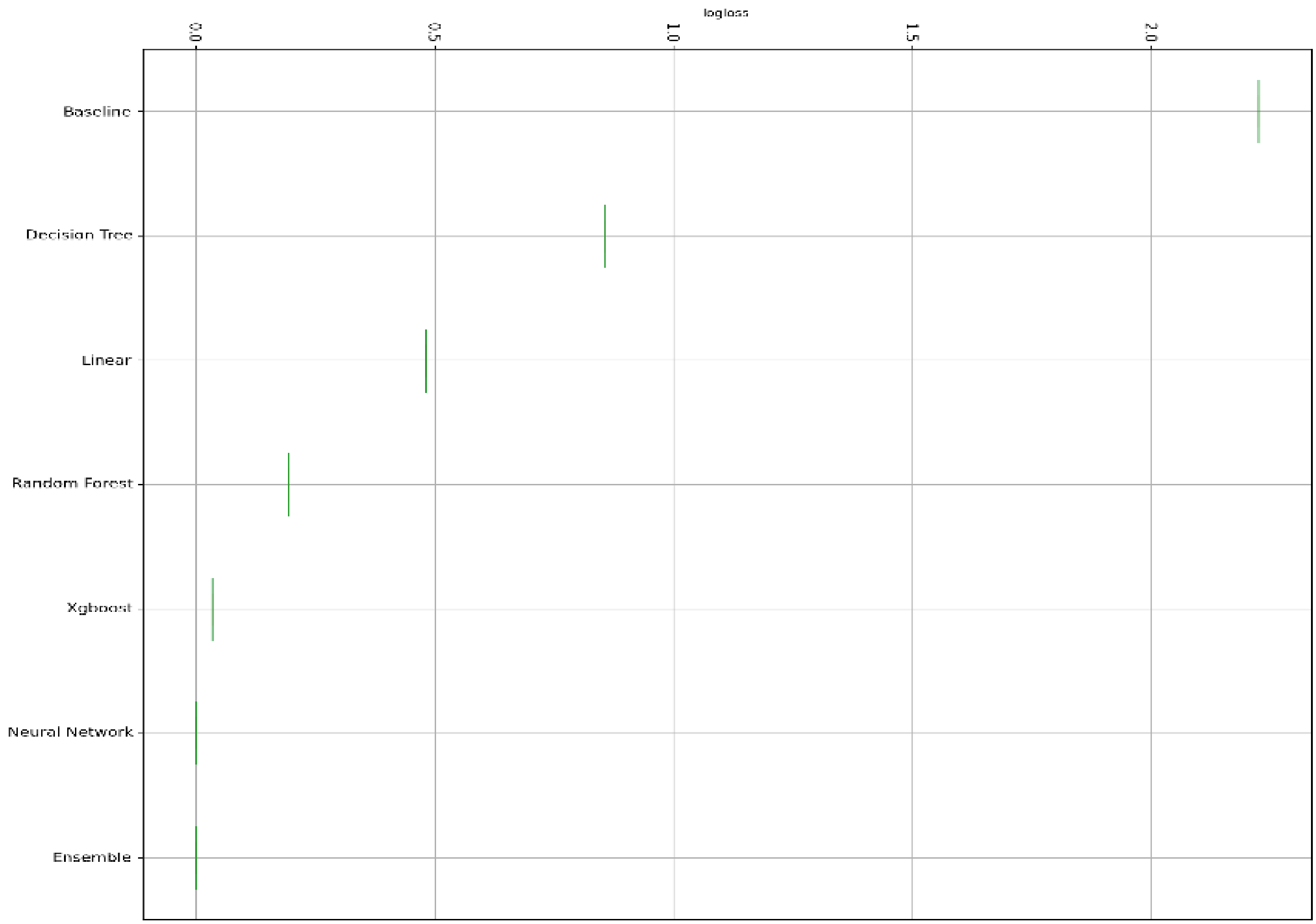

Figure 4

Comparison of Classification models for Type of Crop

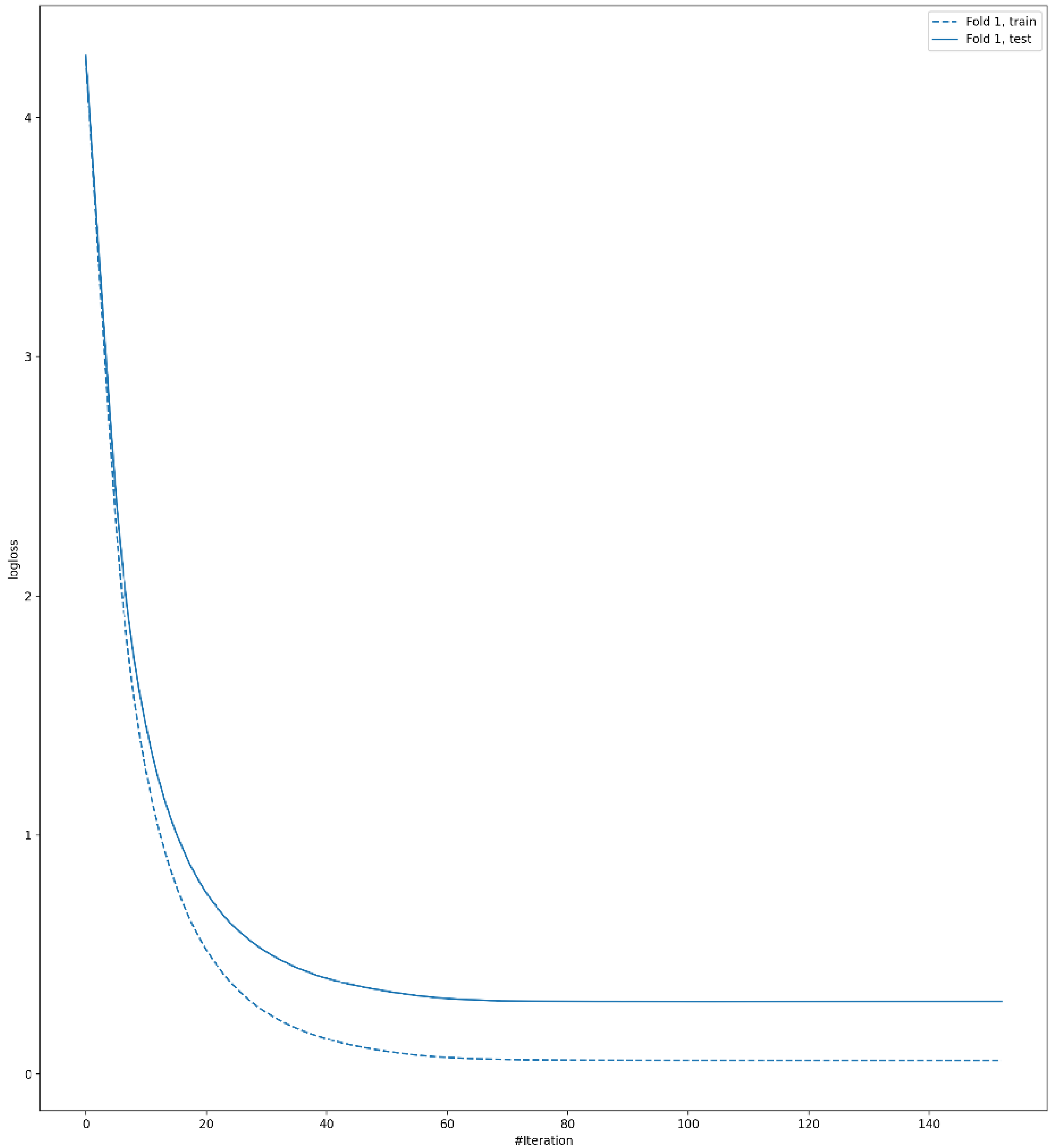

Figure 5

ANN learning curve for individual crop

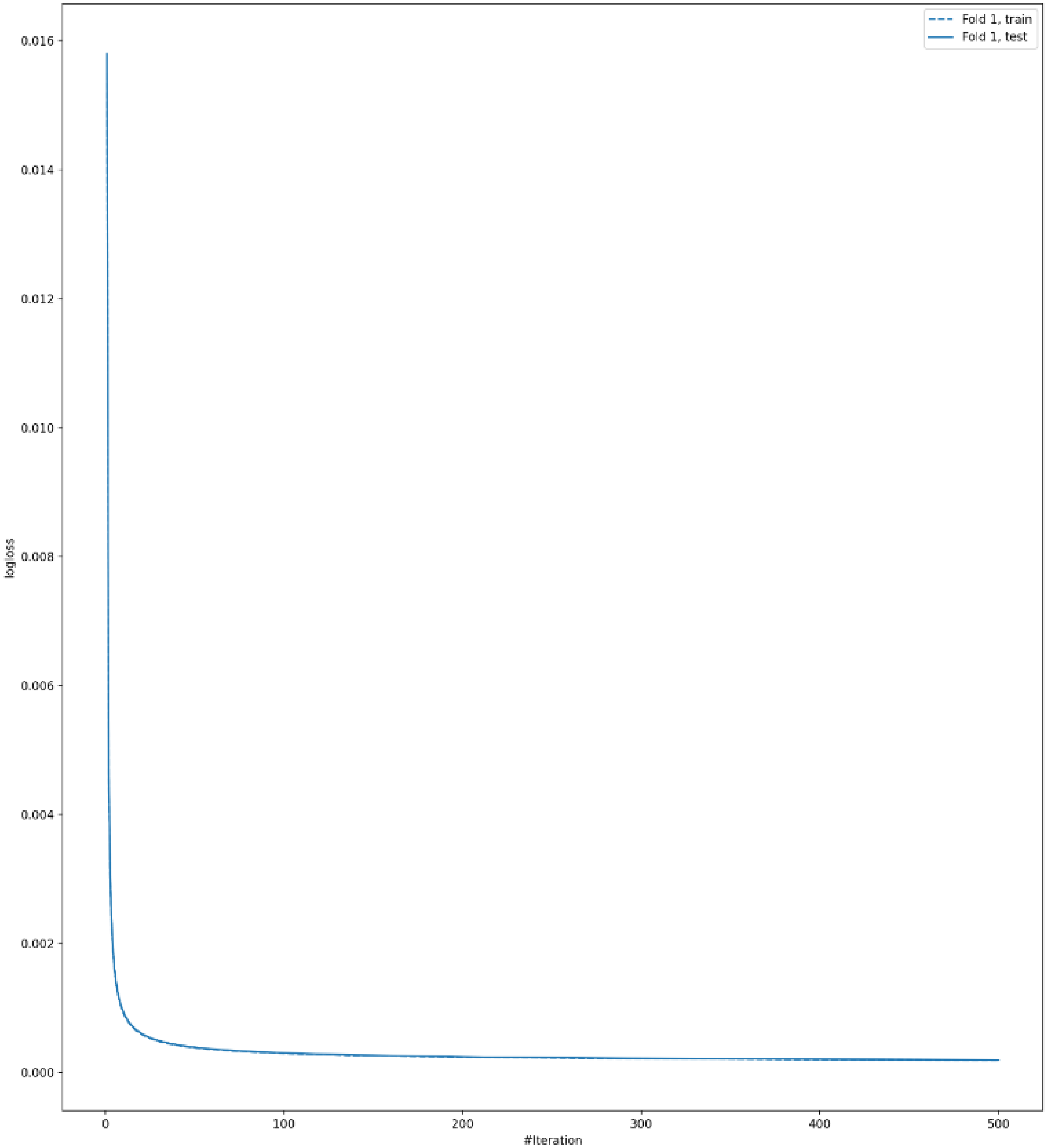

Figure 6

ANN learning curve for Typeof crop

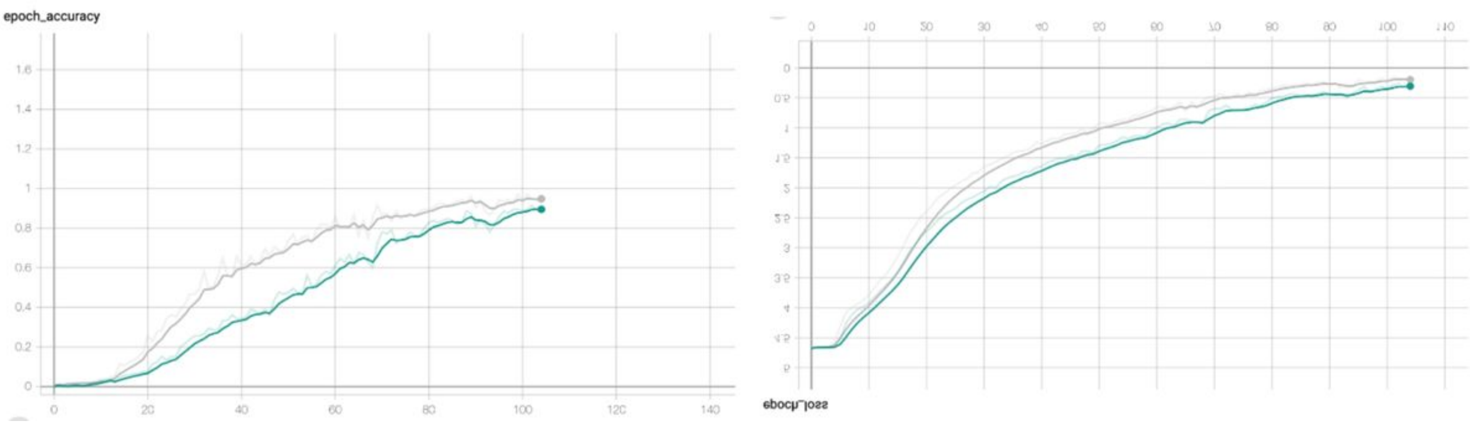

Figure 7

Epoch accuracy and loss for train and test of individual crop

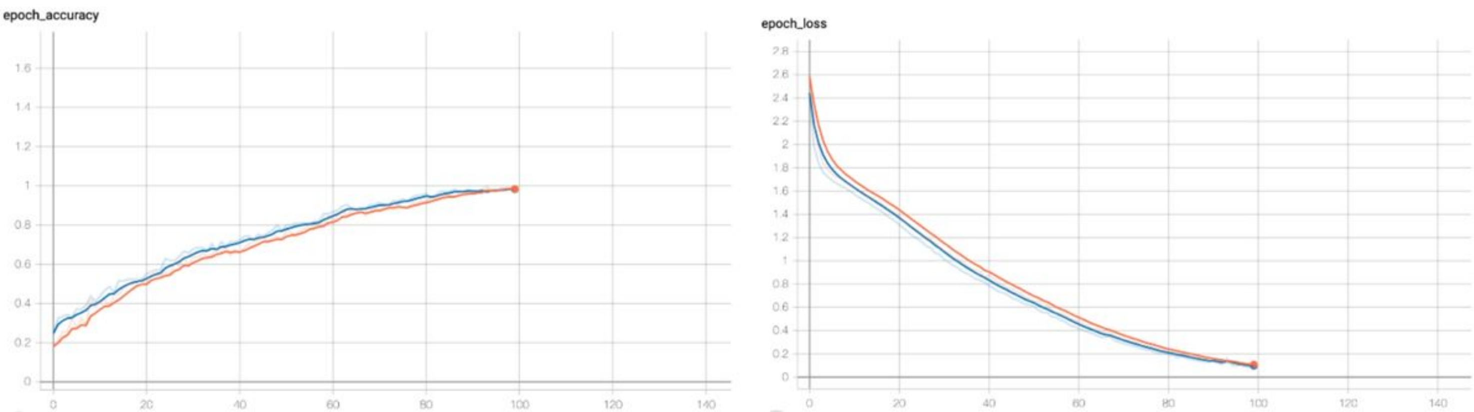

Figure 8

Epoch accuracy and loss for train and test of type of crop

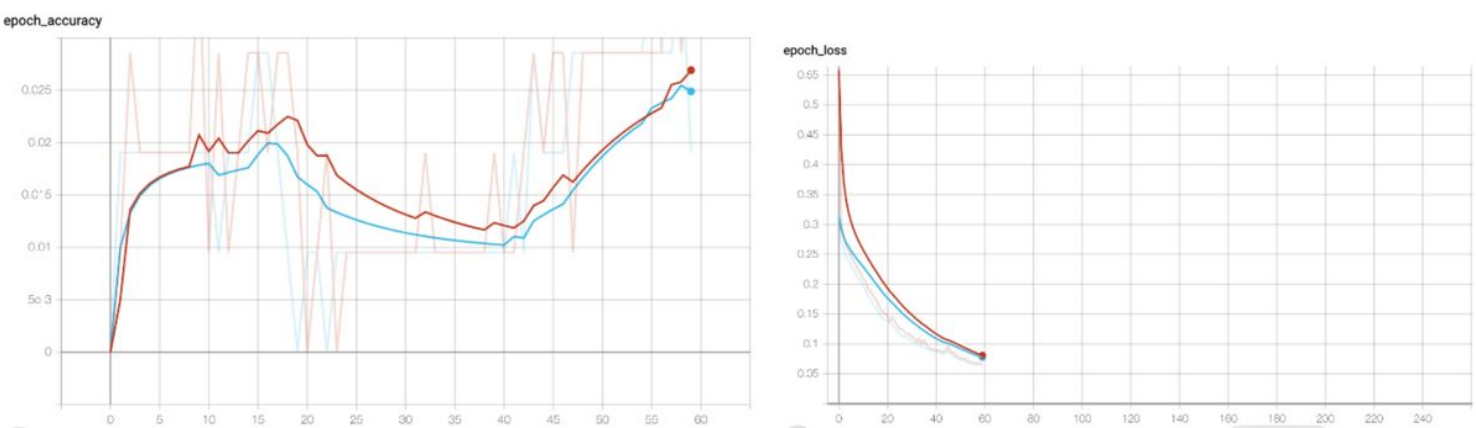

Figure 9

Epoch accuracy and loss for train and test of Districts