Hybrid Method for Improving Accuracy of Crop-Type Detection using Machine Learning

Ankit.Arun.Mahule[1], Dr.A.J.Agrawal[2]
[1]Department of Computer Science and Engineering
Shri Ramdeobaba Kamla Nehru Engineering College, Nagpur-440013, Maharashtra, India
ankitmahule2@gmail.com
[2]Department of Computer Science and Engineering
Shri Ramdeobaba Kamla Nehru Engineering College, Nagpur-440013, Maharashtra, India
agrawalaj@rknec.edu

ABSTRACT

Crop type prediction using sensors requires a large amount of training data from different sensors. These sensors include but are not limited to temperature, humidity, wind direction, wind speed, gas sensors, etc. Gathering accurate data from these sensors is a relatively easy task, but the prediction of crop type from the gathered data requires the knowledge and implementation of high-end classification algorithms. In this work, we analyze different classification algorithms and compare their performances to evaluate the best algorithms suited for the task of crop classification. Furthermore, we also analyze the effect of these algorithms on different crop-prediction applications and recommend which techniques are best suited for which kind of crop. Finally, we suggest some recommendations to the existing algorithms to make them more effective in terms of prediction and response time. This work also suggests a novel algorithm for crop-type detection, and fuses the reviewed algorithms for obtaining a better accuracy in crop-prediction systems.

Key words: Accuracy, Classification, Crop, Prediction, Recommend

1. INTRODUCTION

Government and agricultural managers require statistics on the spatial distribution and location of cultivated vegetation for making plans functions. Groups can more accurately plan the import and export of food products based on such records. Although some ministries of agriculture and meals protection yearly commission their workforce to map exceptional crop types, these ground surveys are costly and but cover only a pattern of farms. Crop identification and classification is a multi-domain image processing problem, in which image segmentation or division is one of the most useful tasks while performing classification. The division procedure permits us to isolate the picture into noteworthy parts agreeing on a specific paradigm. Grouping calculations like k-implies fluffy c-implies and in this way the Iterative

Self-Organizing Information Analysis Technique (ISODATA) calculation are utilized effectively for division issues, anyway, these strategies characteristically, don't think about the logical data for a pixel, what is important to get a genuine division. An extremely viable methodology for including the highlights of the pixel neighbourhood is that the Bayesian estimation nears the Markov Random Field (MRF) [2-9]. With this methodology, a name field is registered accepting that reliance exists between all likelihood circulations of the pixels having a place with an equal neighbourhood. This supposition is dictated by considering Markov Chain prior dissemination. Gauss Markov Measure Field (GMMF) [4] is one of the models that mix Bayesian estimation in with Markov Random Field and it's utilized in various game plan endeavours [5-8, 11-13]. One of the most difficulties for GMMF, concerning all procedures reinforced the mix of Bayesian estimation and MRF for picture division, is that the likelihood count. On account of 1D or 3D include spaces the probability is frequently registered upheld the relating standardized histograms. Nonetheless, the calculation of the probability turns into an extremely difficult issue at the point when the measure of highlights increments. Inside the instance of harvests grouping for satellite pictures the number of highlights is high, which might impede the immediate utilization of the GMMF [4]. On the contrary hand, crop arrangement might be a perplexing assignment because of the comparability of the unearthly marks among various crops. Consequently, the decision of the element space, i.e., data sources, might be a key advance during this exploration, with the goal that we will utilize the GMMF as a classifier that licenses us to incorporate logical data and to arrive at great order results. In creators considered a pixel-based picture approach to section 5 diverse land spread types in Russia. The trial work incorporated the base Euclidean separation, the case classifier, Mahalanobis separation, the most extreme probability classifier, and grouping strategies. The component space was made out of blue, green, and red groups. The least difficult presentation was accomplished by the most extreme probability. Creators in utilized three diverse vegetation lists: the Normalized Difference

Vegetation Index (NDVI), the Green Normalized Difference Vegetation Index (GNDVI), and the Normalized Difference Red Edge Index (NDRE) for crop order inside the district found in Turkey. All lists were registered mulling over the ghostly groups acquired from the Rapid Eye satellite, which is that the principal high-goals multispectral satellite framework joining the red-edge band which is touchy to vegetation chlorophyll. Four diverse element spaces were concentrated inside the examination in: the essential ghostly space contains NDVI, GNDVI, and NDRE lists, the other three components ghostly spaces are made out of just two of the three lists included inside the first space. For crop order, the help vector machine technique was utilized. Thus, to solve this issue, the next section describes some of the recent algorithms used for this purpose, and statistically analyses them to find the advantages and drawbacks of each of them. The section also proposes some improvements in them and suggests which algorithms must be used for what kind of application.

2. LITERATURE REVIEW

Calculations that order a given arrangement of readings into a gathering of yield types are called crop characterization calculations. These calculations take a shot at a harvest based preparing set to return up with a model or a gathering of decides that arrange a given information sensor information into one among a gathering of discrete yield types. The k-implies [5] calculation has been utilized for grouping purposes. There are many grouping calculations, seemingly superior to this, and consequently the method of reasoning behind picking this is because the code overall has been composed during a multi-strung, bunch prepared way, and k-implies is that the main calculation in Microsoft R supporting such a component.

Conventional enhancement calculations, as stochastic slope plunge (SGD), upgrade the experimental misfortune work straightforwardly. The SDCA [6] picks an exceptional methodology that upgrades the twin issue. The twin misfortune work is parameterized by per-model loads. In each cycle, when a preparation model from the preparation informational index is examined, the relating model weight is balanced all together that the twin misfortune work is advanced concerning this model. No learning rate is required by SDCA to settle on a choice advance size as is required by different inclination drop techniques.

Meanwhile neural systems are commonly known to be utilized in profound learning and demonstrating complex issues like picture acknowledgment, they're additionally handily adjusted to re lapse issues. Any class of measurable models is regularly seen as a neural system if they utilize versatile loads and may inexact non-direct elements of their data sources. Neural system relapse is especially fit for issues where a progressively customary relapse model can't fit an answer.

Quick Logistic Regression [7] is taken care of by the rxLogisticRegression () calculation which is utilized to anticipate the value of an all-out factor from its relationship to at any rate at least one free factors accepted to have a strategic circulation. The contrast between this model and subsequently the first (Logistic Regression) is this is frequently commonly quicker and takes into account a more extensive customization.

The advancement system used for rxLogisticRegression is that the constrained memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS). Both the L-BFGS and standard BFGS calculations utilize semi Newtonian techniques to assess the computationally concentrated Hessian framework inside the condition utilized by Newton's strategy to ascertain steps. In any case, the L-BFGS [8] guess utilizes just a constrained measure of memory to process ensuing advance heading, so as that it's particularly fitted to issues with an outsized number of factors. The memory Size parameter determines the measure of past positions and slopes to store to be utilized inside the calculation of the resulting step.

This student can utilize flexible net regularization: a straight blend of L1 (tether) and L2 (edge) regularizations. Regularization might be a strategy that will render a not well-presented issue progressively tractable by forcing imperatives that give data to enhance the information which forestalls over fitting by punishing models with extraordinary coefficient esteems [9]. This will improve the speculation of the model learned by choosing the ideal unpredictability inside the predisposition fluctuation exchange off. Regularization works by adding the punishment that is identified with coefficient esteems to the mistake of the speculation. An exact model with outrageous coefficient esteems would be punished more, yet a less precise model with increasingly traditionalist qualities would be punished less. L1 and L2 regularization have various impacts and uses that are corresponding in specific regards. l1Weight: are regularly applied to inadequate models, when working with high-dimensional information. It pulls little loads related highlights that are generally irrelevant towards 0. l2Weight: is ideal for information that is not inadequate. It pulls huge loads towards zero. The resulting segment portrays the proposed work.

3. PROPOSED WORK

The block diagram for the proposed system is shown in Figure 1. From the diagram we can observe that the input goes into different classification layers, and an output is obtained from each of the layers. A simple machine learning-based selection system is then used to find out the most matching class. This class is shown at the output as the predicted crop class.
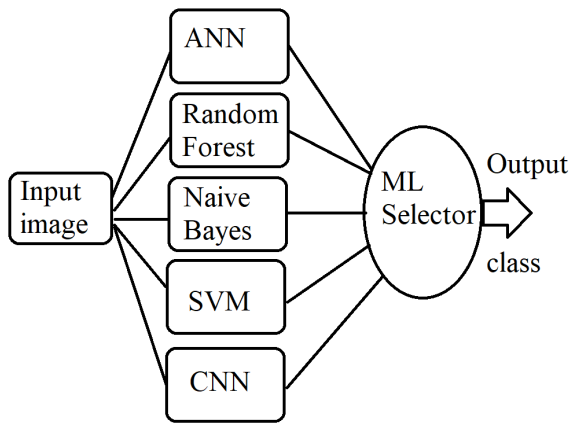
Figure 1: Block diagram of the proposed hybrid classifier

Crop type classification is a very complicated process, due to the number of interdependent steps which are needed to be followed in a proper order. Due to these complexities, the selection of standard algorithms in each stage is very crucial so that the final outputs can be accurate. The following general steps are followed for any hyper spectral image classification system; these steps can be modified as per the application, in which the classifier is applied,

- Placement of the image acquisition device
- Acquisition of the image
- Image fusion
- Segmentation or clustering of the fused image
- Feature extraction and selection
- Classification and post-processing

In the proposed crop type classification application requires these steps to be followed in a proper sequence. In our work we will be utilizing the concept of classification to evaluate the Type of crop present in a particular region, which will help industries to use automatic drone imagery data for analysis of different areas. The layers which are required for processing & classifying the leaf diseases are,

• Image Capture Layer

This is the primary layer within the processing set. This layer is liable for capturing leaf images and arranging them for correct processing. This layer must confirm that each one the captured leaf images are taken at a correct angle, at a correct distance, and at a correct lightning condition to urge high accuracy during processing.

• Pre-processing Layer

Images captured by the primary layer are given for pre-processing. The pre-processing layer will remove any noise (if present) within the image, smoothen the image, and make it ready for processing. The only responsibility of this layer is to convert the image into a form that's consistent in terms of size, brightness, amplitude, etc. This layer is typically not concentrated upon during the design of the classification

network, but if properly designed it can improve the system performance by leaps and bounds.

• Segmentation Layer

The segmentation layer is that the first point of processing for the image. The pre-processed image is given to the presentation layer to extract the foremost important components from the image. The components include the regions of interest, just like the leaf regions, the infected regions, and therefore the background regions. Usually the background regions are detected with utmost accuracy by algorithms like Otsu, Saliency maps, etc. then subtracted from the input image to urge the ultimate segmented image.

• Feature Extraction Layer

The segmented regions are given for feature extraction. This step defines the extent of accuracy of the system, and must be most carefully selected. Input image pixels can't be given directly for classification because they will be very high in number, and may need inconsistent leaf positioning. Thereby this layer converts the input image into a numerical array of features that will be used for representing the image uniquely. An accurate feature extraction layer will produce similar features for images in the same class, while it'll produce dissimilar feature sets for images of various classes. Methods like grey level co-occurrence matrix, colour maps, edge maps, etc. are used for this purpose.

• Feature Selection Layer

While the feature extraction layer will try its best to supply different feature sets for various class of images, but redundancy will always persist. This redundancy affects the classification performance just in case of huge datasets. Thanks to this, there's a requirement to get rid of the redundancy from the feature sets. Algorithms like feature set variation, differential evolution, etc. are used for this purpose.

• Classification Layer

A major part of research in plant disease detection is directed towards this layer. This layer is liable for comparing the features with one another, and evaluating the simplest possible category for the query image. Algorithms like convolution neural networks, deep nets, support vector machines, etc. are used for this purpose. The proposed ML selector works by finding out the most recurring class which is identified by ANN, Random Forest, Naïve Bayes, SVM, and CNN, and provides that particular class at the output.

4. RESULTS AND ANALYSIS

In this section, we define a set of criteria for analyzing the algorithms. The set of algorithms are categorized as per the length of the input, the type of features extracted, the accuracy, and the probable application of the system under various real-time scenarios. Table 1 showcases this comparison in detail.

Table 1: Performance and analysis of different classifiers

Algorithm	Crop Types	Accuracy (%)
ID3[1]	Wheat, Bajra	85.00%
Naïve Bayes [2]	Wheat, Cotton, Bajra	86.00%
Bayesian Belief [3]	Various	79.50%
K2 [4]	Corn, Cotton, Orange	85.00%
Gradient Ascent Training [5]	Various	89.00%
Linear Regression [6]	Various	86.00%
Multi-Layer Perceptron [7,8,9]	Various	91.00%
RBF [10,11]	Various	93.00%
Conjunctive Rule Algorithm[12]	Various	30.00%
Proposed	Various	95.00%

From the above table we can statistically summarize that the performance of the proposed classifier which is better than MLP and RBF systems will produce better accuracy than any other classifier combination.

5. CONCLUSION AND FUTURE WORK

From the outcome we can see that the accuracy of the proposed classifier which is better than MLP and RBF is better than the usual classification systems. Moreover, the proposed classifier, MLP, and RBF have been proven to be the most effective in terms of classification, but their effectiveness can be further improved with the help of a bag of features like a bag of sensors and a bag of visual features. These features when combined with a deep learning-based classification system can produce very high accuracy.

As an extension of this research, we propose the use of artificial intelligence techniques like q-learning and reinforcement learning to further improve the performance of the system. Moreover, the usage of feature selection techniques will further improve the system accuracy.

REFERENCES

[1] M.V.R. Vivek, D.V.V.S.S. Sri Harsha, P. Sardar Maran, "A Survey on Crop Recommendation Using Machine Learning", International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-7, Issue-5C, February 2019.

[2] S.Pudumalar, E.Ramanujam, "Crop Recommendation System for Precision Agriculture", 2016 IEEE Eighth International Conference on Advanced Computing (ICoAC).

[3] Melchizedek I. Alipio, Allen Earl M. Dela Cruz, Jess David A. Doria and Rowena Maria S. Fruto, "A Smart Hydroponics Farming System Using Exact Inference in Bayesian Network", 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE 2017).

[4] Ranjith Bose, Ranjith, Suraj Prakash, Subham Kumar Singh, Dr Vishwanath, "Intelligent Approach for Classification of Grain Crop Seeds Using Machine Learning", International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 05 May-2018.

[5] P. Kanaga Priya, Dr. N. Yuvaraj, "An IoT Based Gradient Descent Approach for Precision Crop Suggestion using MLP", International Conference on Physics and Photonics Processes in Nano Sciences

[6] Iwan Syarif, Dito Hafizh Indiarto, Ira Prasetyaningrum, Tessy Badriyah, Edi Satriyanto, "Corn Pests and Diseases Prediction using linear Regression and Natural Spline Methods", 2018 International Conference on Applied Science and Technology (ICAST). https://doi.org/10.1109/iCAST1.2018.8751583

[7] Nataliia Kussul, Mykola Lavreniuk, Sergii Skakun, and Andrii Shelestov, "Deep Learning Classification of Land Cover and Crop Types Using Remote Sensing Data", IEEE Geosciences and Remote Sensing Letters.

[8] Muhd Khairulzaman Abdul Kadir, Mohd Zaki Ayob, Nadaraj Miniappan, "Wheat Yield Prediction: Artificial Neural Network based Approach", 2014 4th International Conference on Engineering Technology and Technopreneuship (ICE2T). https://doi.org/10.1109/ICE2T.2014.7006239

[9] Andreas Kamilaris, Francesc X. Prenafeta-Boldu, "Deep learning in agriculture: A survey", Computers and Electronics in Agriculture 147 (2018) 70–90. https://doi.org/10.1016/j.compag.2018.02.016

[10] Abhishek Pandeya, Sunil Kr. Jhab, and R.Prasada, "Retrieval of Crop Parameters of Spinach by Radial Basis Neural Network Approach Using Xband Scatterometer Data", ISSN 10683674, Russian Agricultural Sciences, 2010, Vol. 36, No. 4, pp. 312–315. Allerton Press, Inc., 2010. https://doi.org/10.3103/S1068367410040245

[11] G. Camps-Valls, A. J. Serrano-Lopez, L. Gomez-Chova, J. D.Martın-Guerrero, J. Calpe-Maravilla, and J. Moreno, "Regularized RBF Networks for Hyper Spectral Data Classification", Conference Paper in Lecture Notes in Computer Science September 2004. https://doi.org/10.1007/978-3-540-30126-4_53

[12] Jose M. Pena, Pedro A. Gutierrez, Cesar Hervas-Martinez, Johan Six, Richard E. Plant and Francisca Lopez-Granados, "Object-Based Image Classification

of Summer Crops with Machine Learning Methods",
Remote Sens. 2014, 6, 5019-5041.
https://doi.org/10.3390/rs6065019

[13] Shubhangi Neware, "**Fruit Grading System using k means clustering and Artificial Neural Network**", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.1, January – February 2020
https://doi.org/10.30534/ijatcse/2020/95912020

[14] H. D. Gadade, Dr. D.K.Kirange, "**Machine Learning Approach towards Tomato Leaf Disease Classification**", International Journal of Advanced Trends in Computer Science and Engineering, Volume 9, No.1, January – February 2020
https://doi.org/10.30534/ijatcse/2020/67912020