International Conference on Recent Trends in Data Science and its Applications
DOI: rp-9788770040723.155

Agro–Advisory System Using Machine Learning Algorithms

Ishan Gupta
Department of Data Science and Business Systems
SRM Institute of Science and Technology,
Kattankulathur
g.ishan091@gmail.com

Dr.V.Kavitha,
Professor, Department of Data Science and Business Systems
SRM Institute of Science and Technology,
Kattankulathur
kavithav2@srmist.edu.in

***Abstract*—The agro-advisory system is a computerized toolthat gives personalized recommendations on the best cropsto grow in a particular location. This method makes crop recommendations based on scientific evidence while takinginto consideration a number of variables such as soil type,climate,rainfall,temperature,andotherenvironmentalconditions.Thetechniqueisintendedtohelpfarmers,agriculturaladvisors,andresearcherschoosethebestcropsto cultivate, resulting in increased agricultural production and sustainability. Modules of the system included at a gathering, soil analysis,cropdatabases,andmachinelearning.Severalapplicationsexistforthesysteminagriculture,agriculturalconsulting,researchandenvironmentalmanagement.Finally,anagro-advisorysystem is a powerful tool that can help farmers and otherstakeholdersmakeinformedcropselectionandmanagement choices, resulting in improved agriculturalproductivityandsustainability.Thepaperpresentsarecommendation system that aims to solve the problem ofsuggestingappropriatecropsbasedonsite-specificparameters. Theproposedsystemutilizesanensemblemodelwithmajorityvoting,incorporatingDecisionTree,K-NearestNeighbor, Logistic Regression, Random Forest, SVM, and Naïve Bayes as learners,tooffercrop recommendations that are both highly accurate and efficient.**

***Keywords*—Advisorysystem,MachineLearning,Decision Tree, K-Nearest Neighbor, Logistic Regression,SVM,RandomForest, Naive Bayes.**

I. Introduction

Agro-advisory systems are becoming increasingly essentialforagricultureandfoodproduction.Withtheglobalpopulationpredictedtoreach10billionby2050,farmersandagriculturalsystemsmustbeabletoproducemorefoodwhileusingfewer resources. A agro-advisorysystemcanassistproducers in optimising crop yields, reducing environmentalimpact,and increasingprofitability.

This paper aims to explore the idea of agro-advisory systemsandthepotentialadvantagestheyoffertocontemporaryagriculture.Initially,wewillexaminetheunderlyingtechnologybehindagro-advisorysystems,whichincludestheimplementationofmachinelearningalgorithmsandpredictiveanalytics.Wewillthenlookatthevariousapplicationsofagro-advisorysystems,suchasyieldprediction,pestanddiseasecontrol,andnutrientmanagement. Finally, we will go over the potential benefitsofagro-advisorysystemsaswellasthechallengesthatmaysurface when they are implemented. Agro-advisory systemtechnologyiscomplicatedanddiversified.Largedatasets canbe analysed using predictive analytics to discover patternsthat can be used to influence crop selection and managementdecisions.Machinelearningalgorithmscanbeused tomodelthe environment and develop decision-making models thatcan be used to determine the best crop combination for aparticular environment. Satellite and drone technology canalso be used to provide information about soil conditions,crop health, and potential pest and disease threats. Agro-advisory systems have a broad range of uses. Yield predictionmodels can be used to evaluate a crop's potential yield in agivenenvironment,allowingfarmerstochoosethemostprofitablecrops.Furthermore,pestanddiseasecontrolmodels can be used to identify pest and disease risk areas,allowingfarmerstotakepreventiveaction.Nutrientmanagementmodelscanalsobeusedtoidentifyareasofsoildepletionandrecommendfertilisersorothersoilamendments.Therearenumerouspotentialbenefitstoimplementing agro-advisory systems. Farmers can increaseyields while reducingenvironmental effect byoptimisingcropselectionandmanagementoptions.Agro-advisorysystems can also provide important data to inform policychoices and promote sustainable agriculture. Finally, agro-advisorytoolscanassistfarmersinenhancingprofitabilitybylowering input costs and increasing yields. However, therearesomeproblemsthatmustbeaddressedbeforecroprecommendation algorithms can be implemented. To begin,the accuracy of the models is contingent on the integrity ofthe data used. Furthermore, many farmers, particularly indevelopingnations,maybehesitanttoadoptnewtechnology.Finally,greatercollaborationbetweendatascientistsandfarmers is necessary to guarantee that models are tailored tofarmers' specific needs. Finally, agro-advisory systems havethe potential to revolutionise contemporary agriculture andincrease food production. Agro-advisory systems, which usepredictive analytics, machine learning algorithms, satelliteanddronetechnology,canhelpfarmersoptimisecropselection and management options and increase profitability.However, some challenges must still be overcome beforethesesystemscan be successfully implemented.

II. Related Work

Pudumalar, S., et al. [1] Many businesses, including finance,retail, medical, and agriculture, rely on data mining. Datamining is used in agriculture to analyze both biotic and abioticaspects.AgricultureisanimportantindustryinIndia,contributing significantly to the economy and employmentgeneration.FarmersinIndiafrequentlyconfrontthe

DOI: rp-9788770040723.155

problemof choosing the incorrect crop based on soil needs, resultinginaconsiderabledecreaseinproduction.Precision farming is a contemporary agricultural technology that addresses this issue by selecting the optimal crop for each area based on data from studies on soil quality, soil types, and crop production statistics.This method improves crop selection accuracy andyield. The authors present a recommendation system based onparticularsitefeaturesthatutilizesanensemblemodelwithamajority voting method and learners such as KNN, RandomForest, Naive Bayes, and CHAID.

Kumar,Avinash,ChittaranjanPradhanandSobhangiSarkar.[2]Croplossesinagriculturearefrequentlycausedbytheimproperselectionofcropsforacertainfield.Farmersfrequentlyfailtounderstandcroprequirementssuchasminerals,soilmoisture,andothersoilrequirements,resultinginsufferingandfinanciallosses.Anothertypicalissuethatfarmersfaceisthediscoveryofpestsanddiseasesthatcanaffectcrops,oftenatlatestageswhencontrolisdifficult.OurstudytacklestheseconcernsbyofferingaRecommendationSystemwhichforecaststhebestcropforthefarmeranddetectsprobablepestswhilerecommendingtreatmentmethods.Inthis research,theyimplemented theSupportVectorMachine(SVM),theDecisionTreeandtheLogisticRegression algorithms and found that the SVM classificationmodel surpasses the other algorithms in terms of accuracy.Bandara,Pradeepa,etal.[3] Astheamountofavailablefarmlandislimited,automatingdifferentaspectsofagriculturehasbecomecrucial,withorwithouthumanintervention.InSriLanka,despitehavingaccesstomanualagriculturalmethods,thereisnosystemavailabletoidentifyenvironmentalfactorsandrecommendtheidealcropvarietyforfarming.Toaddressthis,thepaperproposesarecommendationsystemthatusesanintegratedmodelcombiningArduinomicrocontrollersforcollectingenvironmentalfactorsMLtechniquessuchasSVMand NaiveBayes,unsupervisedmachinelearningalgorithmssuchas NaturalLanguageProcessingandK-MeansClusteringtosuggestacropthatfitstheselectedland'ssite-specificparameters.Theproposedapproachtacklesthesignificantchallengeofselectingthemostappropriatecropvarietytocultivateinalimitedareabecauseenvironmentalelementssuchassoilconditions,temperature,andwaterlevelsareunpredictable and continually changing. Thecroprecommendationsystemutilizesenvironmentalparametersandfeedsthemthroughitstrainedsub-modelstopredictthemostsuitablecroptypefor thechosenlocation.

Rajak, Rohit Kumar, et al. [4] The proposed paper presents acroprecommendationsystemthatutilizesArduinomicrocontrollerstocollectenvironmentalfactors,andmachine learning techniques such as Nave Bayes and SVMalongwithunsupervisedmachinelearningalgorithmsK-Means Clustering and NLP, to recommend the most suitablecropforaspecificpieceoflandwithprecisionandeffectiveness.Duetounknownenvironmentalcircumstances,the system solves the challenge of determining the best cropto produce on residential or farmed grounds. The systemforecasts the best crop for the given area by analysing theobtaineddata with trainedsub-models.

III. Background Techniques

A. DatasetCollection

Thedatasetfilecontainsimportantsoilcharacteristicsthatarerequired to determine the suitable crop for cultivation. Inaddition, various online resources that provide general cropstatistics were consulted. The dataset which we used to trainour model that contains 2200 data. The following productsare considered in this algorithm: apple, blackgram, banana,coconut,cotton,chickpea,coffee,grapes,jute,lentil,kidneybeans,maize,mungbean,muskmelon,mango,mothbeans, orange,pigeonpeas,pomegranate,papaya,watermelon, and rice. As well as soil parameters such asphosphate(P),potassium(K),nitrogen(N)andpH.Temperature,rainfall,andhumidityareallweatherconditions. (showninTable 1andFig.1)

TABLE1. Soil And Weather Conditions

	N	P	K	temperature	humidity	ph	rainfall	label
0	90	42	43	20.879744	82.002744	6.502985	202.935536	rice
1	85	58	41	21.770462	80.319644	7.038096	226.655537	rice
2	60	55	44	23.004459	82.320763	7.840207	263.964248	rice
3	74	35	40	26.491096	80.158363	6.980401	242.864034	rice
4	78	42	42	20.130175	81.604873	7.628473	262.717340	rice

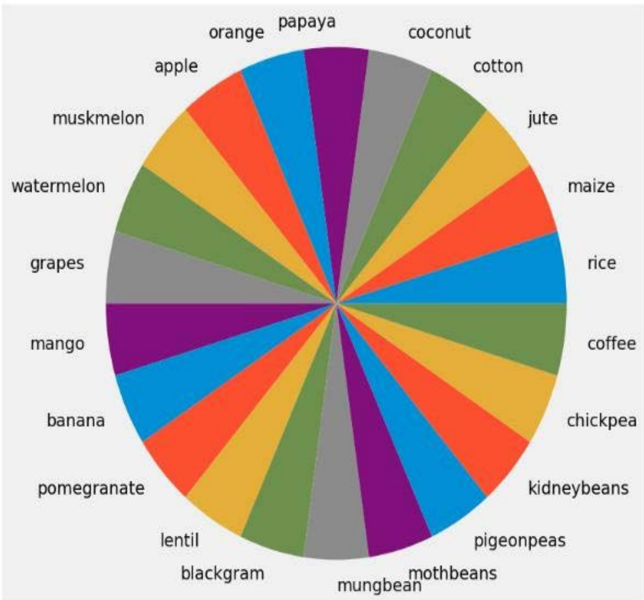

Fig.1.crop presentindataset

We have added two more datasets for states and soil kinds.ThefirstincludesAgriCommodityPricesandTradingInformation,Weather(Historical)DatathatAffectsAgriculture, Land Usage Statistics Per Crop Land for EachCrop, Crop Yield Information, Agri Inputs Data, Crop PestandDiseaseInformation,RetailandWholesalePricesforAll AgriCommodities,andallfactorsandsourcesthataffect

DOI: rp-9788770040723.155

AgricultureandAgriCommodityPrices(showninFig.2).Thesecondoneorganisessoilkindsby state.

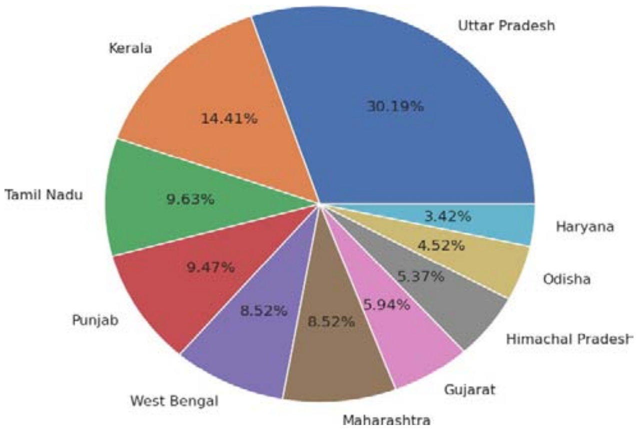

Fig.2. States productions percentage

B. *Learners Used in the Model:*

1) *KNN:*

It is a technique that may be applied to both classification and regression [5]. It is a simple solution that maintains all existing instances and categorizes new cases using a similarity metric. The technique uses a distance measure that is Manhattan distance or Euclidean distance to determine the "closeness" of the sample group and classify it accordingly.

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier()

knn.fit(x_train,y_train)

predicted_values = knn.predict(x_test)

x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('K Nearest Neighbours')
```

2) *SVM Algorithm*

There are several machine learning algorithms that can be used to classify structured and unstructured data [6]. Classification is assigning objects to certain groups, and the major goal of classification problems is to determine which group or class new data belongs to. One widely used supervised ML model for regression and classification is the Support Vector Machine, although it is mainly used for classification tasks. SVMs separate training data points into groups with the smallest possible margin. To enable nonlinear classification, the kernel trick is a technique used by SVMs that implicitly translates inputs into high-dimensional feature spaces. Overall, the SVM is a powerful tool for solving classification problems, particularly in scenarios where the relationship between input features and output classes is nonlinear. Its ability to perform well on both unstructured and structured data has made it a popular choice in the field of machine learning.

In our model, we used SVM as follows:

(I) importing the SVC module from the sklearn.svmClass;

(II) creating an SVM classification object.

(III) Finally, we fitted our data.

```
from sklearn.svm import SVC
```

```
# data normalization with sklearn
from sklearn.preprocessing import MinMaxScaler
# fit scaler on training data
norm = MinMaxScaler().fit(x_train)
x_train_norm = norm.transform(x_train)
# transform testing dataabs
x_test_norm = norm.transform(x_test)
SVM = SVC(kernel='poly', degree=3, C=1)
SVM.fit(x_train_norm,y_train)
predicted_values = SVM.predict(x_test_norm)
x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('SVM')
```

3) *LogisticRegression:*

It is a statistical model that's commonly used to analyzerelationshipsbetweenabinarydependentvariableandoneormoreindependentvariables,whichisalsoknownastheLogit model [7]. The model employs a logistic function to estimatethe probability of a certain outcome or event, based on thevalues of the independent variables. In its most basic form,the model is used to model a binary dependent variable.However, there are more sophisticated versions of the modelthatcanhandlemorethantwocategoriesornon-linearrelationships.

Logistic Regression is widely used in various fields, such asfinance,epidemiology,socialsciences,andmarketing,toanalyzeandpredictoutcomesbasedoncertainvariables. Itisalso commonly used in machine learning, where it is used asaclassificationalgorithmtopredicttheprobabilityofacertain class.

Inour model,weusedLogisticRegressionasfollows:

(I) ImportingLogisticRegressionmodulefromsklearn.linearClass.

(II) CreatingLogistic Regression object.

(III) Finally,wesuitourfacts.

```
from sklearn.linear_model import LogisticRegression

LogReg = LogisticRegression(random_state=2)

LogReg.fit(x_train,y_train)

predicted_values = LogReg.predict(x_test)

x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('Logistic Regression')
```

4) *RandomForest:*

DOI: rp-9788770040723.155

It'sanensemblelearningmethodthatcanbeusedforarangeofproblems,includingregressionandclassification.Theapproach involves creating numerous decision trees duringtraining, and the mode or mean prediction of the individualtrees is used as the final prediction to prevent overfitting tothe training set. Tin Kam Ho pioneered the notion of randomforestsbyusingEugeneKleinberg's"stochasticdiscrimination"techniquetoclassificationviatherandomsubspacemethod.Later, AdeleCutlerandLeoBreimandesigned the "Random Forests" algorithm, which combinesBreiman's "bagging" approach with random feature selectionto generate a set of decision trees with controlled variance.Whilerandomforestswereinitiallydesignedforthree-dimensional data, recent studies have shown that they canalso be usedfor arbitraryobjects byleveragingpairwisesimilaritiesbetweenitems.Overall,aRandomForestconsistsofdecisiontreesthatprovideaclassificationfortheattributesofanewobject,andtheforest'sfinal prediction isbasedonamajorityvotefromthe trees.

Inourmodel,weusedRandomForestasfollows:

(I) ImportingRandomForest modulefromsklearn.ensembleClass.

(II) CreatingRandomForestobject.

(III) Finally,wesuitourfacts.

```
from sklearn.ensemble import
RandomForestClassifier

RF = RandomForestClassifier(n_estimators=20,
random_state=0)
RF.fit(x_train,y_train)

predicted_values = RF.predict(x_test)

x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('RF')
```

5) DecisionTree:

Itis awidely-usedsupervisedlearning technique[8]forclassification and regression tasks. Its objective is to create apredictive model by learning decision rules from a given setofdata(i.e.,trainingdata)thatcananticipatethetargetvariable's'classorvalue.TheDecisionTreecomprisesdecision nodes and branches, with the leaves signifying theend outcomes. The recursive process each node acts as a testfor a distinct attribute, and each branch that stems from thenode stands for one of the feasible reactions to that test. Thisprocessisiteratedrecursivelyforeachsubtreethatoriginates from the new nodes until the final Decision Tree model isformed.

Inourmodel,weused theDecisiontreemethodasfollows:

(I) Importing package DecisionTree Classifierfromsklearn.treeClass

(II) CreatingDecisionTreeClassifierobject.

(III) Finally,wefittedourdata.

```
from sklearn.tree import DecisionTreeClassifier
DT =
DecisionTreeClassifier(criterion="entropy",random_state=2,max_depth=5)

DT.fit(x_train,y_train)

predicted_values = DT.predict(x_test)
x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('Decision Tree')
```

6) NaiveBayesClassifier:

TheNaiveBayesclassifierisapopularstatisticalclassificationtool.ItisbasedonBayes'theorem,whichstatesthat the probability of witnessing evidence or features givena hypothesis or class is proportional to the probability of thathypothesisorclass.Giventheclass,theNaiveBayesclassifier assumes that the characteristics are independent ofeach other and estimates the probability of each class bymultiplying the probabilities of each feature given that class.Becauseofthissimplifyingassumption,itiscomputationallyefficient, particularly for huge datasets. The Naive Bayesclassifierhasbeenusedeffectivelyinmanydisciplines,includingnaturallanguageprocessing,textclassification,andspam filtering. It may be applied to both binary and multi-classclassificationissues.

We used the Naive Bayes classifier in our model to predicttheclassofagivensamplebasedonitsattributes.Weutilisedthe classifier's probabilities to produce suggestions for cropkinds and fertilisers that are best suited to a specific soil andweathercircumstances.

Inourmodel,weusedNaiveBayesclassifier asfollows:

(I) ImportingNaive Bayesclassifier modulefromsklearn.naive_bayesClass.

(II) CreatingNaiveBayesclassifierobject.

(III) Finally,wesuitourfacts.

```
from sklearn.naive_bayes import GaussianNB
NaiveBayes = GaussianNB()

NaiveBayes.fit(x_train,y_train)

predicted_values = NaiveBayes.predict(x_test)
x = metrics.accuracy_score(y_test, predicted_values)
acc.append(x)
model.append('Naive Bayes')
```

```
K Nearest Neighbours --> 0.975
SVM --> 0.9795454545454545
Logistic Regression --> 0.9522727272727273
RF --> 0.990909090909091
Decision Tree --> 0.9
Naive Bayes --> 0.990909090909091
```

DOI: rp-9788770040723.155

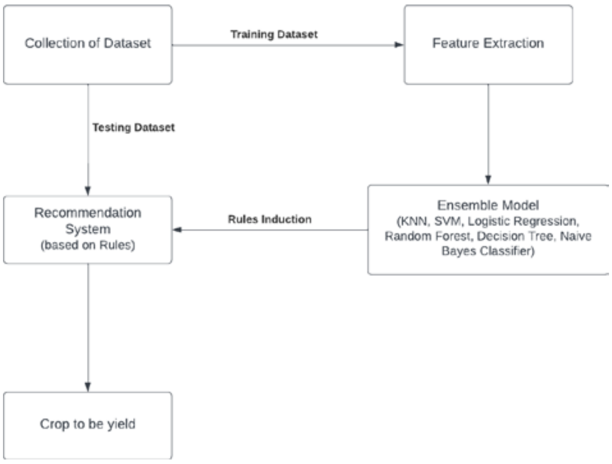

Fig.3.BlockDiagramof ProposedModel

IV. Results and Discussion

The suggested method may recommend appropriate cropsandfertilisersforacertainsoiltypeandclimaticcircumstanceswithaprecisionof99.09%bycombiningmachinelearning techniques,earthanalysis,andmeteorologicaldata.Thebestagro-advisorysystemalgorithmsweredeterminedtobeRandomForestandNaïve Bayes Classifier. This implies that the proposed techniquemight be a beneficial tool for farmers to increase crop outputand soil health by making tailored suggestions based on theirindividualagriculturalneeds.

Based on their performance on the appropriate datasets, theaccuracypercentagesofmanyalgorithmswereanalysedand presented in the Fig.4.

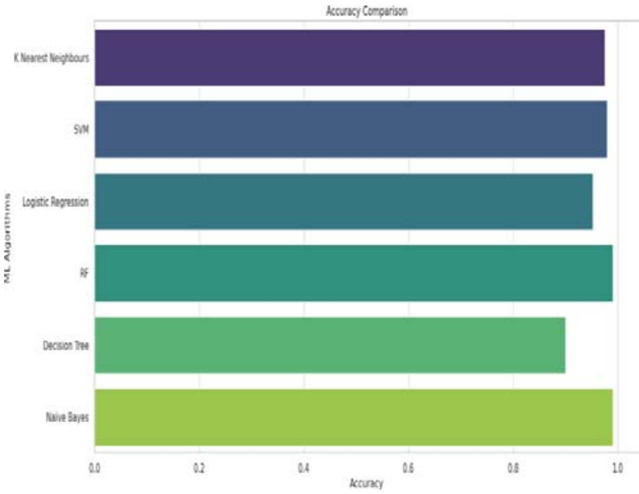

Fig.4.MLAlgorithms andAccuracy

Fig.5.MLAlgorithms andAccuracyachieved

TABLE 2. Algorithm's Accuracy & F1 score

Algorithms	Accuracy	F1score
KNN	97.5%	97.5%
SVM	97.95%	97.65%
LogReg	95.22%	95.15%
RF	99.09%	99.3%
DecisionTree	90%	98.18%
NaïveBayes	99.09%	99.05%

As a result of the comparative graph above, we can infer thatwhen dealing with data sets comparable to the ones used inRandom Forest, the Naive Bayes Classifier method worksbest,asshownby thegraphandin the table.

V. Conclusion

InacountrylikeIndia,whereagricultureissoimportant,itiscriticaltoguaranteethateverycomponentofthesector,includinginvestmentsinagriculturalseeds,ismanagedproperly. The suggested method may deliver personalisedsuggestionsforcropvarietiesandfarmingproceduresthataredmatchedtotheindividualconditionsofafieldbyutilizing dataonsoilfeaturesandweathertrends. Thiscanassistfarmers in increasing food output, improving soil health, andlowering inputcosts.

Lookingahead,ournextgoalistoimprovedatacollectingbyincluding a larger variety of variables, in order to furtheroptimisethesystem'saccuracy and efficacy.

References

[1] S. Pudumalar, E.Ramanujam, R.H.Rajashree, C. Kavya,T.Kiruthika,and J. Nisha,“Croprecommendation system for precision agriculture,” In 2016EighthInternational Conferenceon AdvancedComputing(ICoAC),IEEE, January, pp. 32-36, 2017.

[2] Pazhani. A, A. J., Gunasekaran, P., Shanmuganathan, V., Lim, S., Madasamy, K., Manoharan, R., &Verma, A. (2022).Peer–Peer Communication Using Novel Slice Handover Algorithm for 5G Wireless Networks.Journal of Sensor and Actuator Networks, 11(4), 82.

[3] Bandara,and Pradeepa,etal.,"Croprecommendationsystem," International Journal of ComputerApplications, vol.175.22,pp. 22-25, 2020.

[4] Rajak,and RohitKumar,etal."Croprecommendationsystemtomaximizecropyieldusingmachinelearningtechnique," International Research Journal of Engineeringand Technology, vol.4.12,pp. 950-953, 2017.

[5] Dhanabalan, S. S., Sitharthan, R., Madurakavi, K., Thirumurugan, A., Rajesh, M., Avaninathan, S. R., & Carrasco, M. F. (2022). Flexible compact system for wearable health monitoring applications.Computers and Electrical Engineering, 102, 108130.

[6] Gandhi, and Niketa, et al. "Rice crop yield prediction in Indiausingsupportvectormachines," 201613thInternationalJointConferenceonComputerScienceandSoftwareEngineering(JCSSE). IEEE, 2016.

[7] Huang,and Tingkai,etal."Analysisofthegrainlossinharvestbasedonlogistic regression," ProcediaComputerScience, vol.122,pp. 698-705, 2017.

[8] Singh,and Priyanka,etal.,"Decisiontreebasedmobilecrowdsourcingforagricultureadvisorysystem."2014AnnualIEEEIndiaConference(INDICON).IEEE, 2014.

[9] Karakatič, Sašo, MarjanHeričko, and ViliPodgorelec."Weighting and sampling data for individual classifiers andbaggingwithgeneticalgorithms," 20157thInternationalJoint Conference on Computational Intelligence (IJCCI),IEEE,vol. 1, 2015.