International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 june 2021, pp: 1444-1449 www.ijaem.net issn: 2395-5252

Bees Echelon

Himanshu Verma[1]
[1]B.Tech Scholar, Department of IT , Maharaja Agrasen Institute Of Technology, Delhi

Submitted: 01-06-2021 Revised: 14-06-2021 Accepted: 16-06-2021

ABSTRACT : Bees are present all over the world in different size , color, height , features , breeding way and many more things . So a large number of trials were made to classify them , stated bumble bee or honey bee , many researches are already present and were done on different characteristic like size , color etc. But absolutely the analysis there are either that focus on qualitative or quantitative , but to fight it out , thus researchers came up with discovery which could be both qualitative and chemical analysis made to classify them. And making use of machine learning algorithm to classify them gives a booster . Now the classification would be made easy as they are accurate and are quicker and are easy to implement. Thousands of photographs needs to be collected and stored for data set. And by making use of various machine learning algorithms and getting details about the bees which could use by researchers in additionally categorizing of bees . Handling of images were done in order to be applied to the algorithms and have feature extraction done. We have taken extra efforts to reduce the space efficiency there are lots of photographs(data set) which contributes to space and the image also contain unused variety of things like unnecessary background details which makes difficult for the algorithm to point out the bees in the photograph which were collected .

Keywords :- K-Nearest Neighbor , Logistic Regression , Decision Tree , Histogram of oriented gradients.

I. INTRODUCTION

The principal goal of the project is to classify the bees whether the bee is bumble bee or honey bee . Honey bee can be seen in our day to day life , normally bees which we see around us is honey bee which are from Apini tribe . They produce and reserve enormous bulk of honey and these bees build enormous nests by manufacturing of wax . Western honey bee is recognized for growing honey and crop pollination and they are there on earth except antarctica , they have long tongue just like straw which helps them to suck nectar from blossom . Bumble bee is latest , they are huge , blurry insects with tiny , short wings. Bumble bee is bigger than honey bee but they are not capable of producing honey as honey bee do , but bumble bee are very supreme pollinators (food producing), it's a flying bug which can be spotted at higher places , but mainly in South USA , female bumble bee have a practice of biting ,but it doesn't focuses on humans and animals. So it was totally strenous to point these bees , so opulence of researches were in terrible channel the recognise these bees into bumble bee and honey bee.

As many researches tried a lot to classify the bees based on either qualitative or quantitative but no one was able to do it on both of them Like many researcher considered about the wing , they were doing chemical analysis and plenty quantitative analysis , they composed enough of images and studied it deeply and managed to find different pleasing features while doing this analysis . Many of them made use of different technologies which are present in market just like cloud computing and machine learning so on and so forth. In this regards to classify them I collected images and made classes for me so that it would be easy for me to identify them correctly . Took help of google to find different amount of images in which bees are present , also took from github and used red green blue to grey , it makes the image in two dimensional array , but still occupying a lots of space and size , so made use of histogram of oriented images to convert it into one dimensional array reducing the space and size while doing so and did run different machine learning algorithm to recognise which is honey bee or bumble bee , in person advantage :- helps farmer to find out which is honey bee or bumble bee , increasing their production and reducing their time while doing so.

II. RELATED WORK

There were many researches done regarding this topic , like Haiquan Wang[4] used the feature subset selection, along with the parameters of classifier notably impact the classification accuracy. For obtaining optimal performance the artificial bee colony algorithm proposed to reduce the working time of the feature

subset and the parameters of support vector , and while improving the improving performance of ABC algorithm, scout bee phase and initialization are improved. To assess the suggested approach, the simulation was carried out based on datasets from the UCI database. The efficacy of the proposed method is confirmed by simulation results.

Irene LG Newton [5] founded that microbial ecologists now daily make use of next-generation progressing methods to gauge microbial diversity in the environment. One tool laboriously make use of by numerous groups is the Naïve Bayesian Classifierenlarged by the Ribosomal Database Project (RDP-NBC). However, the steadiness and relianceof classifications provided by the RDP-NBCis dependent on the training set utilized.

Jianhua Jiang [7] researched about supply chain network (SCN) management, multi-objective Pareto optimization means the network can meet the demand for both minimal cost and minimal lead-time in SCN. As they need to take care of cost and time so they did not search for multi-objective pareto optimal solution(POS) in SCN . Moreover with the broad application of the internet, so SCN applications can reach to the internet . So the outcome is that the complexity of the SCN grows at a very high rate with number of suppliers growing. It's really a hugeprovocation to find the global multi-objective POS inside a restricted time in SCN management. In order to overcome this problem and to find the solution for it , first, this paper proposes an artificial bee colony (ABC) optimization algorithm with two improvements: (1) a novel solution framework designed to extend the application field of the SCN based on complex network; (2) theacceleration of search speed by adopting naive Bayes classifier.

Loulin Huang , Ngaire H. Hart [2] came up with a concept that rather than identifying individual bees manually , images of ground nests should be thoroughly studied because the number of nest tells us the population of bee , those insects which might be flying near the nest it gives an concept of the population of bee in this area. They made use of WEKA software for calculating accuracies and performance of the classifier.

Jerzy DembskiJulianSzymański [1] came up with a new idea they were using a video stream to detect bee by using neural network concept . They made use of assorted color models for the input format and did the comparison which was needed for the bee detection in a feedforward convolutional architecture . So the detection process was built of a neural binary classifier that in which video streams were recorded and was surveyed and then it found out ROI windows in the frames and was able to process whether it consisted of bee or not.

III. METHODOLOGY

As we have collected thousands of photographs taken by google , github and some of them clicked by me and many resources from the web which are used as a dataset in this reasearch so these may need a proper methodology to be followed for their storage as the space complexity is there we have to take care of that too. So there were unnecessary details in the photographs so , it was getting difficult to extract the bees out of it so we ran different algorithm for taking out essential part from the picture as a lot of space was needed , by extracting essential features out of it we reduced a lot of space and huge amount of time too which model would take to train them , so we did saved time. And now these extracted images need to be stored so we stored them in database and analysed them and thought of giving different classes to the bees so that we could make our work easier and more recognizable to other person , so taking zero for honey bee and taking one for bumble bee but these all needs to be done after the training and testing , so first of all trained all the dataset and hence picking out these classes , applying validation algorithm on the dataset and providing result with accuracies.

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 june 2021, pp: 1444-1449 www.ijaem.net issn: 2395-5252

1. Collection of photographs :-

Collected images from various ways like google , github and more sites too used them as database.

2. **Extracting Essential Features :-**

As there were thousands of pictures so it was very much difficult to extract features , and adding to this problem bees were concentrated in small part of the picture so it become worst , so by using machine learning algorithm we made this work easier , converted image which was in three d form to one d form three channel (red , green, blue) to one channel(gray) but the image wasn't clearly visible so need to perform HOG which improved quality but still a few problem were there .

3. **Storing in database :-**

So extracted essential features and also reduced the dimensions (removing unnecessary details from the images) and needed to be stored in database which would be needed for classification and validation , did by converting them into CSV files , by taking image number and stored in sorted manner.

4. **Dividing data into training set and testing set :-**

As we have to do both training and testing on the same dataset , so we thought of dividing them in different parts and deployed the models on them.

5. **Classification and validation using KNN :-**

K-nearest neighbor classification , it is a parametric method in which classification and regression is possible , talking about the knn classification , the output can be a class membership it is measured by profusion vote of neighbors . If k==1 , so it means item is totally assigned to nearest neighbor class.

Talking about knn regression , the property value of the object is determined from the output and it is equal to the average value of k nearest neighbor.

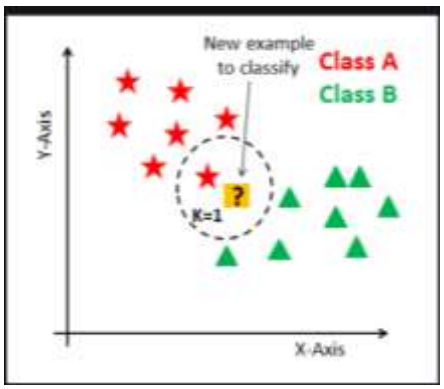


Fig.1 - KNN

While validating , we did split in three parts , one for training data , one for testing data and one for cross validating data and these are used as follows :- for finding the nearest neighbor we made use of training data , for finding simplest value of 'k' we made use of cross validation and finally we did run our model on testing data.

6. Logistic Regression :-

Logistic Regression is a classification algorithm. The nomenclature is done like this because of the function used in this method which is logistic

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 june 2021, pp: 1444-1449 www.ijaem.net issn: 2395-5252

function , it is also known as sigmoid function. It's S shaped figure makes it unique , the curve's shape, it can take any real valued number and our research used to map in 0 and 1 honey bee and bumble bee repectively.

1/(1+e^-value)
Where e is the base of the natural
log (Euler's number or the EXP()
function),valueis theactual numerical value thatyou want totransform.

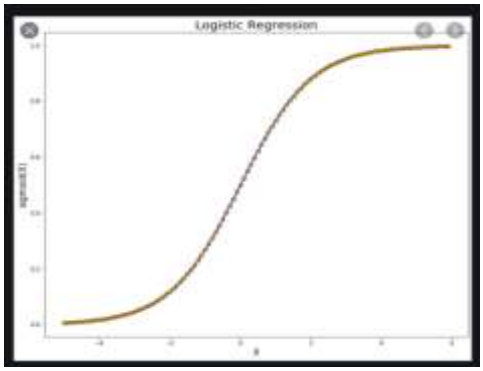


Fig.2 Logistic Regression

7. **Decision Tree:-**

A decision tree is a alternative help tool thatis a tree like structure in which there are choices and their consequences which contain useful resource prices and application and also include chance event result.They are a nonparametric managed getting to know method used for both category and regression
tasks.

8. **Accuracy of the algorithm :-**

Deploying models and doing classification and validation we got different accuracy of different algorithm .

IV. RESULTS

Classification and finding out which bee it is , it was getting a big task for farmers and time consuming too , so by making use of new technology we can make this part feasible and in more better way. Tried to cover as many types of bees possible by collecting huge amount of images and to handle this huge data we did red green blue to gray format so that we can remove the unwanted part from the pictures and also did histogram of oriented gradients . But there was memory issue like it was difficult to store images in three d format , so we made use of functions to convert them into one d and saved a lot of space in this process. After converting them in one d format , there were more issues now , while loading the images from the converted format images were not recognizable , were blur or had some more issues , so now principal component analysis needs to be done , which uses a method of reconstructing the data collinearly which makes information within the data to be contained in small number of features known as components.

Receiver Operator Characterstics (ROC)curve is used as evaluation matrix in binary classification problem. It is curve which is drawn with true positive rate on y axis and false positive rate on x axis , it is being drawn on different value of threshold which helps in separating signal from noice.
Area under curve(AUC) denotes the amount of ability classifier has to find difference between classes and this is the summary of the Roc curve.

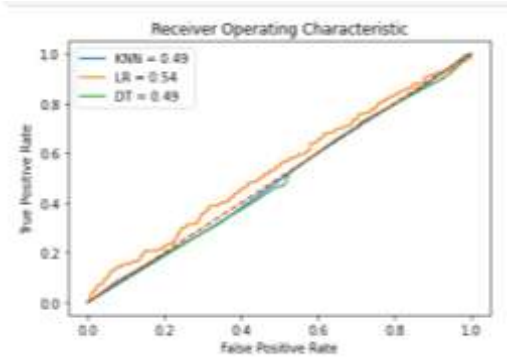
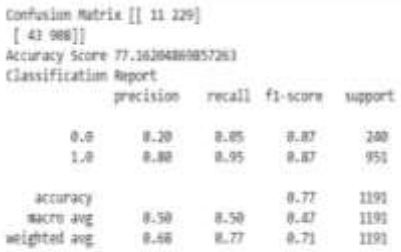


Fig. 3 ROC- AUC curve

Then data was trained and tested on different algorithm and their confusion matrix was obtained . Confusion matrix is an M x M matrix which is used for evaluating the performance in classification model . It compare predicted value with particular target values . It gives a good view of how the algorithms are performing.
For KNN



For Logistic Regression

International journal of advances in engineering and management (IJAEM)
Volume 3, issue 6 june 2021, pp: 1444-1449 www.ijaem.net issn: 2395-5252

```
Confusion Matrix [[ 56 184]
 [183 768]]
Accuracy Score 69.1855583543241
Classification Report
              precision    recall  f1-score   support

         0.0       0.23      0.23      0.23       240
         1.0       0.81      0.81      0.81       951

    accuracy                           0.69      1191
   macro avg       0.52      0.52      0.52      1191
weighted avg       0.69      0.69      0.69      1191
```

For Decision Tree

```
Confusion Matrix [[ 20 220]
 [104 847]]
Accuracy Score 72.79596977329975
Classification Report
              precision    recall  f1-score   support

         0.0       0.16      0.08      0.11       240
         1.0       0.79      0.89      0.84       951

    accuracy                           0.73      1191
   macro avg       0.48      0.49      0.47      1191
weighted avg       0.67      0.73      0.69      1191
```

V. CONCLUSION

Through this research we have identified different types of bees based on their size , type, color and categorized them into bumble bee and honey bee . While performing the algorithms we found different feature of bees and explained the methodology. By collecting more images we can improve accuracy of our algorithms.

REFERENCES

[1]. Jerzy DembskiJulianSzymański.(2019).Bees Detection on Images: Study of Different Color Models for Neural Networks.ICDCIT 2019, 11319 ,295-308.10.1007/978-3-030-05366-6_25

[2]. Ngaire H. Hart , Loulin Huang.(2014).Monitoring populations of solitary bees using image processing techniques. International Journal of Computer Applications in Technology,50(50).45-50.10.1504/IJCAT.2014.063907

[3]. 3.Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario and Tim Landgraf.(2018).Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence.Front. Robot. AI.https://doi.org/10.3389/frobt.2018.00035

[4]. 4.Haiquan Wang , Hongnian Yu , Qian Zhang , Shuang Cang , Wudai Liao , FanbingZhus.(2016).Parameters optimization of classifier and feature selection based on improved artificial bee colony algorithm.IEEE.10.1109/ICAMechS.2016.7813454

[5]. 5. Irene LG Newton &GuusRoeselers.(2012).The effect of training set on the classification of honey bee gut microbiota using the Naïve Bayesian Classifier.BMC Microbiology.

[6]. 6. S. Jain ,V. Bajaj,A.Kumar.(2016)Efficient algorithm for classification of electrocardiogram beats based on artificial bee colony-based least-squares support vector machines classifier.Electronics Letters,52(14).1198 – 1200.

[7]. 7.Jianhua Jiang, Di Wu, Yujun Chen, Dianjia Yu, Limin Wang &Keqin Li.(2019)Fast artificial bee colony algorithm with complex network and naive bayes classifier for supply chain network management.Metrics,23.13321–13337 .https://doi.org/10.1007/s00500-019-03874-y

[8]. Selim Dilmac , Mehmet Korurek.(2013).A new ECG arrhythmia clustering method based on Modified Artificial Bee Colony algorithm, comparison with GA and PSO classifiers.IEEE,19-21.10.1109/INISTA.2013.6577616

[9]. Schroder S .(2002)The New Key to Bees: Automated Identification by Image Analysis of Wings.ECOL INFORM

[10]. 10.Abhilakshya Agarwal,Rahul Pradha.(2020).Bees Classifier Using Soft Computing Approaches.International Conference on InnovativeComputingandCommunications. 737-748.https://doi.org/10.1007/978-981-15-5148-2_64

[11]. Yongcun Cao , Saisai Ji and Yong Lu.(2020).An improved support vector machine classifier based on artificial bee colony algorithm. Journal of Physics Conference Series1550:042073.10.1088/1742-6596/1550/4/042073

[12]. ShunmugapriyaPalanisamy , Kanmani S.(2012). Classifier Ensemble Design using Artificial Bee Colony based Feature Selection.IJCSI.9(3).

[13]. Pinar Tapkan, LaleOzbakir , SinemKulluk ,AdilBaykasoglu.(2016).A cost-sensitive classification algorithm: BEE-Miner.ELSEIVER,95.https://doi.org/10.1016/j.knosys.2015.12.010

[14]. Manuel Chica , Pascual Campoy.(2012).Discernment of bee pollen loads using computer vision and one-class

classification technique.Journal of Food Engineering.112(1-2).50-59.https://doi.org/10.1016/j.jfoodeng.2012.03.028

[15]. 15. Haldi Rao, Lichuan Gu , Juanjuan Feng.(2019). selection based on artificial bee colony and gradient boosting decision tree. Appied Soft Computing.74.634-642.https://doi.org/10.1016/j.asoc.2018.10.036