

Bees Assorter

Himanshu Verma

B.Tech Scholar , Department of IT , Maharaja Agrasen Institute of Technology , Delhi ,India

To Cite this Article

Himanshu Verma, "Bees Assorter", *International Journal for Modern Trends in Science and Technology*, 6(12): 61-65, 2020.

Article Info

Received on 07-November-2020, Revised on 22-November-2020, Accepted on 28-November-2020, Published on 01-December-2020.

ABSTRACT

Many attempts were made to classify the bees that is bumble bee or honey bee , there have been such a large amount of researches which were made to seek out the difference between them on the premise of various features like wing size , size of bee , color, life cycle and many more. But altogether the analysis there have been either that specialize in qualitative or quantitative , but to beat this issue , thus researchers came up with an answer which might be both qualitative and quantitative analysis made to classify them. And making use of machine learning algorithm to classify them gives a lift . Now the classification would take less time as these algorithms are pretty fast and accurate . By using machine learning work is made easy . Lots of photographs had to be collected and stored for data set. And by using these machine learning algorithms we would be getting information about the bees which might be employed by researchers in further classification of bees. Manipulation of images had to be done so as on prepare them in such a way that they will be applied to the algorithms and have feature extraction done. As there have been a lot of photographs(data set) which take a lot of space and also the area in which bees were present in these photographs were too small so to accommodate it dimension reduction was done , it might not consider other images like trees , leaves , flowers which were there present in the photograph which we elect as a data set.

KEYWORDS: K- nearest neighbor , Histogram of oriented gradients

I. INTRODUCTION

The main objective of the project is to classify the 2 different types of bees namely bumble bee and honey bee ,honey bee we all have seen it but it represent only a little percent of bee species . They are from Apini tribe . They produce and store huge quantity of honey and these bees build huge nests by making use of their wax which is released from a fixed colony . Western honey bee is identified for producing honey and crop pollination and they're found on every continent on earth except antarctica , they have long tongue just like straw which helps them to drink nectar from blossom . Bumble bee is new , they're large , fuzzy insects with short , stubby wings. Their size is greater than honey bee but they are not producing as much honey as honey bee does but there are very important pollinators (without their help food won't grow), it's a flying bug which can be spotted at

higher places , but mainly in South USA , female bumble bee have a habit of stinging ,but it doesn't targets humans and animals. So it was absolutely difficult to spot these bees , so plenty of researches were in dire straits the identifying these bees into bumble bee and honey bee.

As many researches were already been made to classify the bees but they were either that specialize in quantitative or qualitative , not both of them . Like many researcher studied about the wing , they were doing chemical analysis and plenty quantitative analysis , they collected plenty of images and did the study and located out many exciting feature in doing so.. Some used cloud computing , some have used machine learning to classify the bees . So to classify them I made use of images and made classes two different classes for the identical .The classes has been described to spot bee as honey bee (class 1) or bumble bee (class

0). Firstly I collected different images of the bees from various resources like google , github and lots of more and so used RGB to grey so we'd make the image in 2-D array only and after converting it to 2-D it's also taking plenty of space still so I made use of HOG (Histogram of Oriented Images) in it to create it 1-D and performed several machine learning algorithm to classify them as Honey bee or Bumble bee so that so it might help farmer to search out more quickly which is bumble bee and which is honey bee and increase their production.

II. RELATED WORK

There are many researches which we were done supported this subject ,there are many studies on bees, strenuous botany of bees made conservation slowed down, botanist were finding it very difficult .So Schroder S [9] used an identification system which was automated , it had an electronic notebook which had a charge-coupled device camera on the top of a stereomicroscope. In his research he identified bees on the basis of characters of the fore wing of the bees , what he did was he made a video of it and then transferred the image of the wing to the diary . Whenever a user clicks the mouse it defined where is the vein junctions . This system then connects the junctions by automatic line-following and thus digitises the complete venation. He also pointed that training should be done with atleast thirty defined specimens of every sex of every species . As the model get a lot of data of bees , it learns and gets more efficient . Species were identified by automatic comparison of the new data with already present data

Many researchers like Rahul , Abhilakshya Agarwal [10] used soft computing and machine learning to classify the bees , by using machine learning and soft computing they made their work easier and would definitely speed-up the study of bee populations , they also made use of photograph of the bees but their method was different , they made use of algorithms like SVM , Logistic regression and lots of more.

Kanmani S , Shunmugapriya Palanisamy [12] have used artificial bee colony algorithm in SVM(Support Vector Machine) , and it's better result and accuracies than other algorithm and selection of parameters played a important role on the accuracy of the classification algorithm and also impacted generalizing ability of SVM and thus a better bee colony algorithm which induces the

best and efficient parameter combination of SVM. Artificial bee Colony based featured selection was utilized in one amongst the research , and therein Feature Selection (FS) was used which helps to make the method of classification faster by extracting the needed and important information from the dataset .

Fanbing Zhus, Haiquan Wang , Wudai Liao , Hongnian Yu , Qian Zhang , Shuang Cang [4] made use of selecting such features subset and the parameter which is needed for classification so that it could improve the accuracy of the classification better . They made an algorithm in which optimal classification gave a boon to them and helped them to optimize the features subset and finding the desired parameters for the support vector machines which helped in boosting the performance. They collected dataset for UCI database and effectiveness was confirmed after seeing the results.

Jerzy Dembski Julian Szymański [1] came up with different thing they were ready to detect bee during a video stream , they made use of neural network for this. They took assorted color models which they used as an input format and did the comparison which was needed for the bee detection in a feedforward convolutional architecture . So the detection process was made up of a neural binary classifier that in which video streams were recorded and was viewed and then it classified ROI windows in the frames and was able to process whether it contained bee or not.

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.

Franziska Boenisch , Tim Landgraf , Fernando Wario, David Dormagen, Benjamin Wild and Benjamin Rosemann [3] came up with observing the honey bee members over their whole life span , their approach was analyzing the collective behavior in insects , they relied on paths of the motion as an middle or constructing layers of data can be used by anyone to infer behaviors of bee (each bee would have it's own behavior) and social

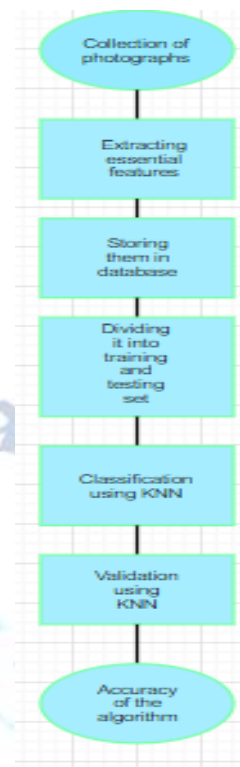
interaction among them , as per them after studying so many research paper they discovered lots of time was needed to mark bees , and by using their method , marking decoding accuracy would improve as a result speed also improved.

III. METHODOLOGY

The methods we utilized in this paper are different while compared to others. We collected photographs of bees from different means like clicking some picture , taking many pictures from google , github and plenty of more resources which are needed as dataset for this research . We ran different algorithms for extracting most essential part from the pictures that were collected , because the number of pictures was high , so lots of space would be needed to store such pictures and also a lot of time would be required train and test these all images and casting off the essential part from the pictures , would definitely take plenty of time so for saving time we did this . Then after all these information we gathered , we stored them within the database and while observing and analyzing come to an result , that may be needed to create my work easier , i.e dividing them into two classes , class 0(honey bee) and sophistication class 1(bumble bee) , but before all of this class division , we had to divide the info into training and testing set . So firstly we trained all the dataset and thereby extracting these above classes and then applying validation algorithm on the testing set and providing the results and accuracies while doing so.

IV. IMPLEMENTATION

Following diagram makes easier to grasp how we proceeded :-



1. Collection of photographs :-

We have collected the datasets from various sources like google photos , github and lots of more places , and used them in my research.

2. Extracting Essential Features :-

As the number of photographs were high and therefore the proportion of bee present within the photo was low , so it was absolutely difficult to store such a large data and further engaged on , so we made use of machine learning algorithm for extracting essential features of bee , like making the image in one D format from three D format (III channels to I channel) i.e from Red Green Blue format to Gray Scale format , but still it has been not upto that mark so we further used Histogram of Oriented Gradients for extracting features , but during the extraction various problem were there , as within the photo which contained bee at the corner was'nt recognized.

3. Storing in database:-

After extracting essential features and reducing the dimensions of images , they have to be stored in database in order that they can be used for classification and validation , so we converted the pictures into comma separated files and used this file as a database , used

their image number and sort for storing it within the database.

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

We had to divide data into two parts , one part was kept for training purpose and second one for testing purpose as if we used the identical dataset for both training and testing it should give hundred percent accuracies , because we are training our model and also the testing it on the same .

5. **Classification using KNN:-**

KNN (K-Nearest Neighbor) It's a non-parametric method which is used for regression and classification . In both of the above cases , the input consists of the k closest training examples within the feature space. The output depends on whether k-NN is employed for classification or regression:

- In k-NN classification, the output may be a class membership. An object is assessed by a plurality vote of its neighbors, with the thing being assigned to the category commonest among its k nearest neighbors (k must be a positive integer, typically small). If $k = 1$, then the item is solely assigned to the class of that single nearest neighbor.
- In k-NN regression, the output is that the property value for the object. This value is that the average of the values of k nearest neighbors.

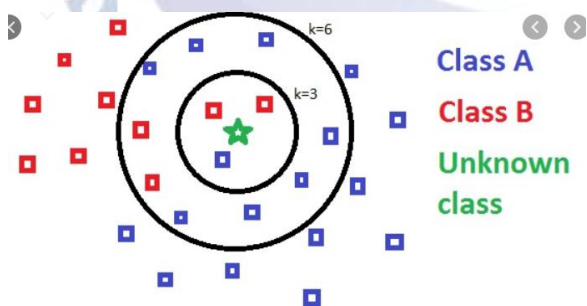


Fig.1 KNN Classification

6. **Validation using KNN:-**

In cross-validation, rather than splitting the info into two parts, we split it into 3. Training data, cross-validation data, and test data. Here, we use training data for locating nearest neighbors, we use cross-validation data to search out the simplest value of “K” and

eventually we test our model on totally unseen test data. This test data is similar to the longer term unseen data points.

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

After doing the classification and validation , we got the accuracy of our algorithm and displayed the same in results.

V. RESULTS

As the classification of bee was difficult and was time consuming process for farmers and for people , so by using machine learning algorithm we've made this part easy and faster. Thousands of images of bees were collected , so that we could every type possible , and so classify them. Because the data collected was too big so it had been consuming lots of space , so to beat this issue we made RGB format to Gray scale , so that we can find out that part of the entire image on which we've to stay target , and discarding behind the left over part . We have made use of Histogram of oriented gradients for extracting features from the pictures that we had converted into gray scale format .

But still the pictures were in 3-D format and storing it , still requires a plenty of space so we need to convert into 1-D format which would surely require less space , it was done absolutely by inbuilt function “img.flatten()”

So all the pictures were converted into 1-D format , but during loading the pictures we found out one image was not recognized from the converted image , so there was problem which has to be removed . We also have to reduce the amount of features we've got using principal component analysis (PCA), PCA is method of transforming the data linearly such that most of the information within the data is contained within a smaller number of features called components.

The Receiver Operator Characteristic (ROC) curve is used in binary classification problems for evaluation metric. It's a probability curve which is plotted between the TPR and FPR on the basis of different threshold values which most importantly separated the signal from noise . The Area Under the Curve (AUC) measure the ability of a classifier to tell difference between classes and is used as a summary of the ROC curve.

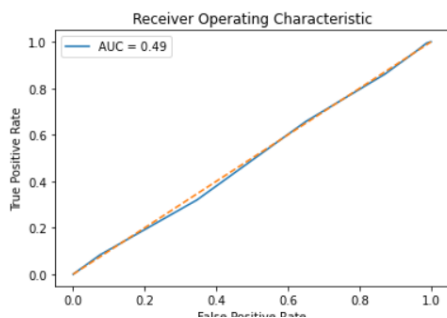


Fig.2 AUC-ROC curve

Then the data was trained and was tested using KNN algorithm , for it confusion matrix was used and a confusion matrix is an M x M matrix that is used in a classification model for evaluating the performance , where M is the number of target classes. The matrix compares the particular target values with those predicted by the model. This provides us a holistic view of how well our classification model is performing and what sorts of errors it is making. and it showed the subsequent accuracy:-

```
Confusion Matrix [[ 11 229]
 [ 43 908]]
Accuracy Score 77.16204869857263
Classification Report
precision    recall  f1-score   support

   0.0         0.20     0.05     0.07         240
   1.0         0.80     0.95     0.87         951

 accuracy         0.77         1191
 macro avg         0.50         0.50         0.47         1191
 weighted avg         0.68         0.77         0.71         1191
```

Fig.3 Accuracy of the algorithm

VI CONCLUSION

In this report , we have classified different types , size of bee into two bumble bee and honey bee successfully . We have explained the technologies and methodology behind it. We can collect more images and perform KNN and try to improve accuracy of our algorithm. We conclude that we need further improvement in our methods to improve the accuracy which is just above average .

REFERENCES

1. "Bees Detection on Images: Study of Different Color Models for Neural Networks" By Jerzy Dembski Julian Szymański.
2. "Monitoring populations of solitary bees using image processing techniques" By Ngair H. Hart , Loulin Huang.
3. "Tracking All Members of a Honey Bee Colony Over Their Lifetime Using Learned Models of Correspondence" By

Franziska Boenisch, Benjamin Rosemann, Benjamin Wild, David Dormagen, Fernando Wario and Tim Landgraf.

4. "Parameters optimization of classifier and feature selection based on improved artificial bee colony algorithm" By Haiquan Wang , Hongnian Yu , Qian Zhang , Shuang Cang , Wudai Liao , Fanbing Zhus.
5. "The effect of training set on the classification of honey bee gut microbiota using the Naïve Bayesian Classifier" By Irene LG Newton & Guus Roeselers.
6. "Efficient algorithm for classification of electrocardiogram beats based on artificial bee colony-based least-squares support vector machines classifier" By S. Jain ,V. Bajaj , A. Kumar
7. "Fast artificial bee colony algorithm with complex network and naive bayes classifier for supply chain network management" By Jianhua Jiang, Di Wu, Yujun Chen, Dianjia Yu, Limin Wang & Keqin Li .
8. "A new ECG arrhythmia clustering method based on Modified Artificial Bee Colony algorithm, comparison with GA and PSO classifiers" By Selim Dilmac , Mehmet Korurek.
9. "The New Key to Bees: Automated Identification by Image Analysis of Wings" By Schroder S
10. "Bees Classifier Using Soft Computing Approaches" By Abhilakshya Agarwal , Rahul Pradha.
11. "An improved support vector machine classifier based on artificial bee colony algorithm" By Yongcun Cao , Saisai Ji and Yong Lu.
12. "Classifier Ensemble Design using Artificial Bee Colony based Feature Selection" By Shunmugapriya Palanisamy , Kanmani S
13. "A COST-SENSITIVE CLASSIFICATION ALGORITHM: BEE-MINER" BY PINAR TAPKAN, LALEOZBAKIR , SINEMKULLUK ,ADILBAYKASOGLU .
14. "DISCERNMENT OF BEE POLLEN LOADS USING COMPUTER VISION AND ONE-CLASS CLASSIFICATION TECHNIQUE" BY MANUEL CHICA , PASCUAL CAMPOY.
15. "FEATURE SELECTION BASED ON ARTIFICIAL BEE COLONY AND GRADIENT BOOSTING DECISION TREE" BY HALDI RAO, Lichuan Gu , Juanjuan Feng.
16. "DETECTION OF THE BEE QUEEN PRESENCE USING SOUND ANALYSIS" BY TYMOTEUZS CEJROWSKI , JULIAN SZYMANSKI
17. "A BUMBLE BEES MATING OPTIMIZATION ALGORITHM FOR THE FEATURE SELECTION PROBLEM" BY MAGDALENE MARINAKI AND YANNIS MARINAKIS
- 18 "PERFORMANCE-BASED CLASSIFIER COMBINATION IN ATLAS-BASED IMAGE SEGMENTATION USING EXPECTATION-MAXIMIZATION PARAMETER ESTIMATION" BY T.ROHLFING , DB RUSSAKOFF , CR MAURER.
19. "A Hybrid Estimation System for Medical Diagnosis using Modified Full Bayesian Classifier and Artificial Bee Colony" by Ahmed T. Sadiq, Noor T. Mahmood
20. "A HYBRID ALGORITHM USING ANT AND BEE COLONY OPTIMIZATION FOR FEATURE SELECTION AND CLASSIFICATION (AC-ABC HYBRID)" BY P. SHUNMUGAPRIYA , S.KANMANI