International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2019 IJSRCSEIT | Volume 5 | Issue 2 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT195221

Fruit Disease Detection Using GLCM And SVM Classifier

Anu S[1], Nisha T[2], Ramya R[2], Rizuvana Farvin M[2]

[1]Assistant Professor, Department of Computer Science and Engineering, Sri Krishna College of Technology1, Coimbatore, Tamil Nadu, India

[2]Department of Computer Science and Engineering, Sri Krishna College of Technology1, Coimbatore, Tamil Nadu, India

ABSTRACT

Analytics plays a critical role in detecting and analyzing the diseases. The proposed system identifies the fruits that are affected with diseases. It is done by collecting the raw data which is subjected to pre-processing. It results in a HSV (hue, saturation, value) converted image. After pre-processing, the resized format of the data is used to extract the information. In feature extraction the image is segmented and converted into matrix using Gray level co-occurrence matrix algorithm. The further classification is done and result is represented in the form of a decision tree using Support Vector Machine (SVM). The disease that affected the fruit is displayed along with the right fertilizer to be used for the plant.

Keywords : Analytics, HSV, Gray level co-occurrence matrix, SVM.

I. INTRODUCTION

In the past twenty years, along with increased fruit tree diseases infections, disaster areas has also increased by 30% when compared with the stats recorded in the 90s. In recent years, pesticide is used mainly to control fruit tree diseases and insect pests and is used superfluously and abusively, leading to quality deterioration of fruit and endangering the human health. The naked eye observation of experts and farmers experience are the main approaches adopted in practice for detection and identification of diseases (Weizheng et al., 2008).This requires continuous monitoring by experts which might be prohibitively expensive in large farms. Furthermore, in some developing countries, farmers may have to go long distances to contact experts, and this makes consulting experts too much expensive and time consuming (Babu and Srinivasa, 2010). There are two main requirements regarding fruit disease detection methods that must be fulfilled: speed and accuracy (Hahn et al., 2009). Such pesticides are cost- effective and efficient only if the application is performed at the right time. Additionally, with precise prediction farmers will reduce number of chemical treatments thus reducing residues in fruits. In most cases, successful prediction depends largely on the selection of predictable parameters on which to base.

In paper (Georgiana, 2009) authors presented a small application of CART decision tree algorithm for weather prediction. Prediction is based on data collected in the period of four years. Selected prediction variables are: year, month, average pressure, relative humidity, clouds quantity, precipitation, and average temperature. From the aspect of appropriate conditions for disease infection, CART decision tree is also used as one of the models to forecast leaf wetness duration within the next 24 hours, working with site-specific weather data estimates as inputs (Kim et al., 2006). In paper(Gang Liu,2005) authors have collected a history of data and

using MATLAB the fruit disease prediction is done. Here they have used neural networks algorithm and collected the weather report data for prediction. Then the system was trained by the history record data. Finally, the ring spot, a fruit tree disease, was chosen as the research object to compare the predicted value with the actual value. The results indicate that the system can predict accurately, run fast, and function robustly. The severity and weather data were analysed using artificial neural network (ANN) models developed using data from some or all field sites in Australia and/or South America to predict severity at other sites(Chakraborty,2004).In the paper(Roman, T.,2004) Different splitting algorithms, historical data and pruning procedures have been used to construct the decision tree. The advantages and disadvantages of the CART algorithm are also discussed.

Early signals of disease have been found on the web and the classification is done by using the SVM classifier. It is a new measure for ranking of extracted terms in order to highlight the more relevant terms. To categorise documents retrieved from the web Naïve Bayes (NB) and Support Vector Machine (SVM) classifiers are used.This method is generic and its used for both humans and animals(Arsevska,E.,2016).Popular data mining algorithms CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3) and decision table (DT) have been used to extract a decision tree or rule-based classifier to develop the prediction models using a large dataset(Chaurasia,vikas,2013).The performance of multiple regression (MR) and three machine learning algorithms namely artificial neural networks, categorical and regression trees, and random forests (RF) were used in predicting the pre-planting risk of SNB in wheat. Here models were evaluated based on the regression of observed against predicted severity values of SNB, sensitivity-specificity ROC analysis, and the Kappa statistics (Lucky, M.,2016).

In paper (Mohammadhassani,M.,2014) Fuzzy modelling approach is being used for the prediction of shear strength of high strength concrete (HSC) beams without stirrups. Along with 122 experimental datasets were extracted from the literature review on the HSC beams with some comparable cross sectional dimensions and loading conditions. A comparative analysis has been carried out on the predicted shear strength of HSC beams without stirrups via the ANFIS method. The observed ratio between the

ANFIS method and the real shear strength shows that the shear strength isbetter in nonlinear iterations such as the ANFIS for shear strength prediction of HSC beams without stirrups.

Image processing is a rapidly growing technology where we obtain an enhanced image so that we can obtain useful information from that image. It is a signal processing technique therefore the given input will be an image and output can be either an image or it can also be the characteristics that are associated with that image. The extracted important details are used to provide a proper description, interpretation and clear understanding of the scenario. In our day to day life, we come across a lot of situations that can be related with image processing. Our human brain combined with eyes is a best example for image processing because we can automatically recognize our parents and relatives only because our brain stores their faces as images and the processing is done faster when we look at them.

Nowadays images are presented in the form of two dimensions and more while processing. So here it is modelled in the form of multidimensional systems. The important step in image processing is feature extraction. It is a dimensionality reduction process the data used here will be clean and non-redundant in nature. The data will be a reduced format where it still remains to be accurate and describes the original data without any change in the actual information.

Digital image processing includes various applications such as remote sensing through satellites, space craft image transmission and storage for business applications. This information will assist us to achieve a greater perception without adding any additional information to the provided data. Using this image processing we can also enlarge and shrink images to gather finest details of the image. Compression tools that work with BMP format are also present to compress the image to the required specification. Rotation tools are used to rotate the images to different angles if necessary. Additionally filters like lightening, embossing, softening are alsoavailable.

II. PROBLEM DEFINITION

The diseases that affect the fruits have increased overtime. The farmers who primarily depend upon these plant crops get highly affected if the production gets decreased. The only way to increase the yield is to find the disease in advance and apply the right fertilizer. It not only prevents the fruits from getting affected by pests it also produces high quality fruits to the consumers.

III. MATERIALS AND METHODS

Data used in this paper are collected and organized particularly for the purpose of this research. This research is focused on two types of data sets. First data set type is the one containing data used for prediction model training. Second data set type represents data that are used for prediction model testing and accuracy evaluation. The entire dataset is collected in the image format.

Methods for: data pre-processing, prediction model creation, evaluation and prediction of future class attribute values are implemented in MATLAB tool. The basic data flow from the raw data to the final training and test data sets, and the steps needed for obtaining the best prediction model are shown on Fig. 1. Using the raw data on the input, prediction model is developed and tested, and ready for the future predictions.

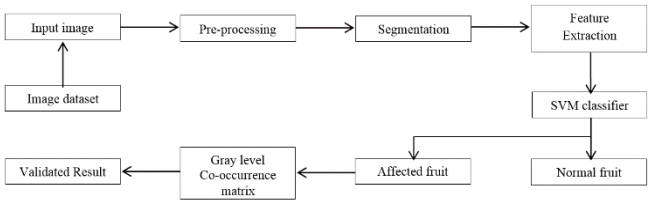

Fig. 1. Data flow diagram, and mutual relations of implemented methods.

The above diagram shows the flow of data and the relation between the methods. When we feed the input with the image from the dataset, it is subjected to pre-processing and a segmented image is obtained. Then feature extraction is performed in order to enhance the contrast of the image and the fruit is classified by using SVM classifier whether it is an affected fruit or a normal fruit. Finally Gray level Co-occurrence matrix is used to validate the result based on the classification.

Pre-Processing

Pre-processing is the first step that is performed in order to enhance the given input image. It does not increase the image information content rather it is useful on a variety of situations like suppress information that is not relevant to the specific image processing or analysis task. The main aim of pre-processing is to improve the image data so by suppressing undesired distortions so that it enhances image features that are relevant for further processing of the images. In pre-processing, no high-level mechanisms of image analysis are involved. The types of pre-processing are contrast enhancement for contour detection. Restoration is to suppress degradation using the knowledge about its nature like the relative motion of camera and wrong lens focus etc. Compression is about searching for ways to

eliminate redundant information from images.Fig.2. shows the pre- processed and hsv converted image for strawberry.

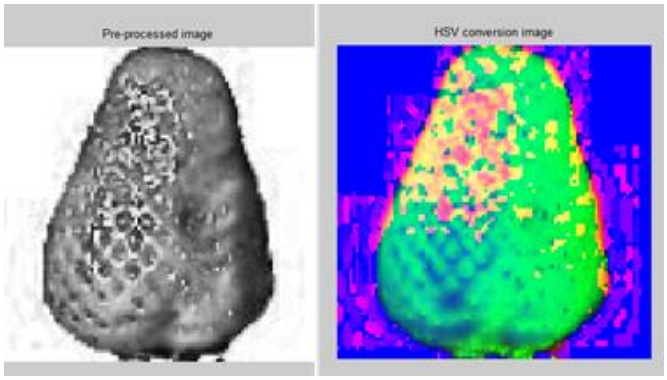

Fig. 2.Pre-processed image and hsv converted image.

FEATURE EXTRACTION

Feature extraction is a dimensionality reduction process which plays a very important role in the area of image processing. Initially, various image pre-processing techniques like binarization, thresholding, resizing, normalization etc. are applied on the given input image. After that, feature extraction techniques are applied to get features that will be useful in classifying and recognizing the images. Features define the behavior of an image they depict its place in terms of storage taken, efficiency in classification and time consumption also. It starts from collecting a measured data and builds derived values that are intended to be informative and non-redundant. It facilitates the subsequent learning and generalization steps, and in some cases it leads to better human interpretations. The initial set of raw variables is reduced to more manageable groups for processing while still being accurate and completely describing the original data set. When the input data is too large to be processed by the algorithm it is suspected to be redundant so it is transformed into a reduced set of features. The feature selection is performed where the subset of the initial features is determined. These selected features are expected to contain only the relevant information with respect to the input data. Therefore, the desired task is performed by using this reduced representation instead of the complete data set.Fig.3 shows the segmented images for strawberry.

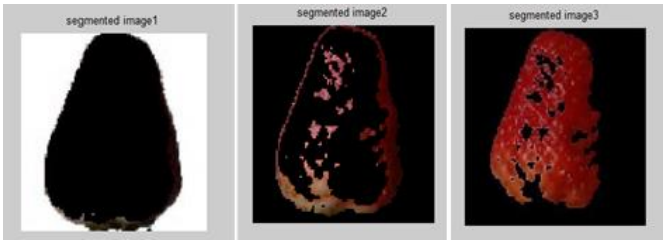

Fig. 3. Segmented images.

SUPPORT VECTOR MACHINE

Support Vector Machine is a supervised learning algorithm which is mostly used in classification problems. Here the number of features is taken as “n” and plotted in an n-dimensional space where each n value will correspond to a particular coordinate. For a dataset containing feature set and label set SVM assigns new example/data points to one of the classes. Basically there are two kinds of SVM classifier they are linear and non-linear. In linear model, the points are plotted in space and are separated by an apparent gap predicting a straight hyper plane which divides both the classes. Maximizing the distance between both hyper plane and nearest data point is the focus of this linear model. As a result the maximum margin hyper plane is produced. In a non-linear model, real world data set is used where the data points are plotted in a higher dimensional space in order to avoid the dispersion in the dataset. This non-linear method not only avoids the dispersion but also overcomes the drawbacks of the linear model. The diagrammatic representation of linear and non-linear model is depicted below.

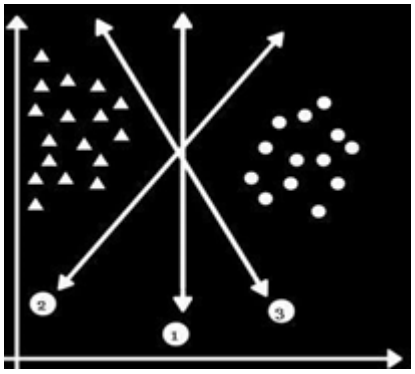

Fig. 4.Linear Model

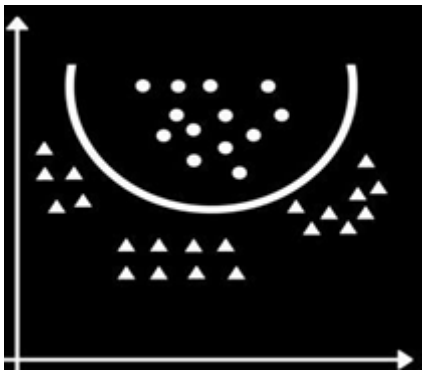

Fig. 5.Non-Linear Model

It constructs a hyper plane that can be used for classification and regression purposes at an infinite plane. Therefore we obtain a good separation that has a largest distance to the nearest data point of any functional margin. The reason SVM stands the best is that it is an effective algorithm is even if the features is larger than the samples and it works the best if the features are large. Nonlinear model, data is classified by using the kernel trick i.e. by increasing the maximum margin. Therefore it is considered as a robust model to solve the problems. The Fig.6 shows how the classification is done based on the disease affected and the corresponding fertilizer is also displayed along with it.

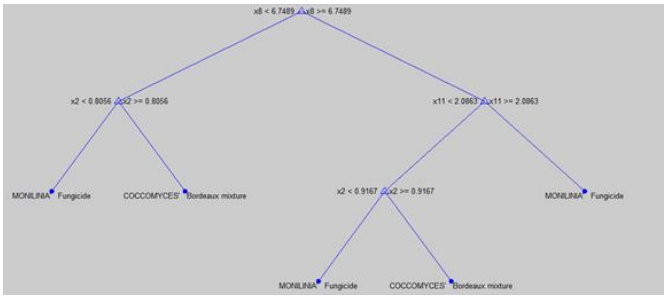

Fig. 6.Classification of diseases and fertilizers.

The relational distance between the given point and the minimum distance data points is called margin. The optimal line that separates two classes is called Maximal-Margin hyper plane. SVM training data include numerical inputs and binary classification. Maximize the margins to obtain optimized result. SVM is used in various domains such as text categorization, digital image analysis, and bio informatics. For example, consider strawberry image is given as input then the suspected diseases are monilinia and coccomyces. The above classification is done based on the level of how much the disease has affected the fruit.

GRAY LEVEL CO-OCCURRENCE MATRIX

Gray Level Co-occurrence Matrix is used for analysing texture of input images. It considers spatial relationship of pixels to determine texture. It calculates the texture by comparing how frequently pair of pixels with certain value and spatial relationship present in an image. There are four statistics which includes, contrast, correlation, energy, homogeneity. Contrast identifies slight variation in gray-level co-occurrence matrix. Correlation determines joint probability and calculates sum of squared values in GLCM. Homogeneity determines relation between distributions of elements and follows second order statics in relation to a pair of pixels that are collected. First order texture measures variance, second order calculates comparison among a group of two pixels from the actual image and third order textures are not practically possible because of time complexity. It compares the relation between two image pixels at the same time referred to as a neighbouring pixel and explains distance and angular spatial relationship in a particular region of fixed size. The frequency of

pixel along with a gray level is considered neither horizontally or vertically.

IV. RESULT

The result is displayed along with the fertilizer to be used to avoid the further spoiling of the crop and to prevent the nearby crops from getting affected by the disease.Fig.7 shows the sample result for a selected fruit from the data set.

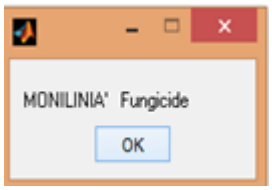

Fig. 7.Result.

V. CONCLUSION

Fruit farms contribute a lot to the food production industry. The farmers who plant fruit crops get highly affected when the fruits get infected by diseases. Thus proposed system acts as a solution to find the fruits that are affected by diseases. The pre- processed images are subjected to feature extraction which segments the images based on the clumps formed. Then the gray level co-occurrence matrix will display the texture of the image based on the values that are obtained. Further, Support Vector Machine is used to classify the diseases and corresponding fertilizer. Thus the system helps us in detecting the fruit diseases easily.

VI. FUTURE WORK

Future work concerns, a deeper analysis of the data for accurate results. This study mainly focuses on finding the fruits affected by diseases. Further we would like to include a dataset of fruits in different sizes so that we can analyze the diseases of the fruits irrespective of their sizes.

VII. REFERENCES

[1]. Babu, M.S.P., Srinivasa, B.R., 2010. Leaves recognition using back propagation neural network - advice for pests and disease control on crops. Technical report, Department of Computer Science & Systems Engineering, Andhra University, India.

[2]. Roman, T., 2004. Classification and Regression Trees (CART) Theory and Applications, A Master Thesis, Center of Applied Statistics and Economics Humboldt University, Berlin, pp. 1–40.

[3]. Kim, K.S., Wang, T.C., Yang, X.B., 2005. Simulation of apparent infection rate to predict severity of soybean rust using a fuzzy logic system. Phytopathology 95 (10), 1122–1131.

[4]. Georgiana, E., 2009. A decision tree for weather prediction. Buletinul. LXI 1, 77–82.

[5]. Hahn, F., 2009. Actual pathogen detection: sensors and algorithms-a review. Algorithms 2(1), 301–338.

[6]. Liu, G., Shen, H., Yang, X., Ge, Y., 2005. Research on prediction about fruit tree diseases and insect pests based on neural network. In: IFIP International Conference onArtificial Intelligence Applications and Innovations. Springer, US, pp. 731–740.

[7]. Chakraborty, S., Ghosh, R., Ghosh, M., Fernandes, C.D., Charchar, M.J., Kelemu, S., 2004. Weather-based prediction of anthracnose severity using artificial neural networkmodels. Plant. Pathol. 53, 375–386.

[8]. Arsevska, E., Roche, M., Hendrikx, P., Chavernac, D., Falala, S., Lancelot, R., Dufour, B.,2016. Identification of terms for detecting early signals of emerging infectious disease outbreaks on the web. Comput. Electron. Agric. 123, 104–115.

[9]. Chaurasia, Vikas, Pal, Saurabh, 2013. Early prediction of heart diseases using data mining

techniques. Caribbean J. Sci. Technol. 1, 208–217 ISSN: 0799-3757.

[10]. Lucky, M., Christina, C., Kevin, G., Peter, O.S., 2016. Predicting pre-planting risk of Stagonospora nodorum blotch in winter wheat using machine learning models.

[11]. Mohammadhassani, M., Nezamabadi-Pour, H., Suhatril, M., Shariati, M., 2014. An evolutionary fuzzy modelling approach and comparison of different methods for shear strength prediction of high-strength concrete beams without stirrups. Smart Struct.Syst. 14 (5), 785–809.

[12]. Mirko, I. (Ed.), 1992. Mikoze biljaka (Mycosis of plants). Nauka, Beograd.

[13]. Pan, G., Li, K., Ouyang, A., Li, K., 2016. Hybrid immune algorithm based on greedy algorithmand delete-cross operator for solving TSP. Soft. Comput. 20 (2), 555–566.

[14]. Rakesh, K., Amar, K.S., Gajendra, R., 2006. Machine learning techniques in disease forecasting: a case study on rice blast prediction. BMC Bioinformatics 7 (1), 485.

[15]. Weizheng, S., Yachun, W., Zhanliang, C., Hongda, W., 2008. Grading method of leaf spot disease based on image processing. In: Proceedings of the international Conference on Computer Science and Software Engineering. 6, CSSE. IEEE Computer Society,

[16]. Washington, DC, pp. 491–494. http://dx.doi.org/10.1109/CSSE. 2008.1649.

[17]. M. Ilic et al. Computers and Electronics in Agriculture 150 (2018) 418–425

Cite this article as :

Anu S, Nisha T, Ramya R, Rizuvana Farvin M, "Fruit Disease Detection Using GLCM And SVM Classifier", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 2, pp. 365-371, March-April 2019. Available at doi : https://doi.org/10.32628/CSEIT195221 Journal URL : http://ijsrcseit.com/CSEIT195221