AUTOMATIC FRUIT CLASSIFICATION SYSTEM USING MACHINE LEARNING TECHNIQUES - A DETAILED ANALYSIS

L. Abirami [1], Pragya Bajpai[2] , S. Pavithraa[3], M. Adithya[4] and T.Kavitha[5]
Final year students[1234] and Assistant Professor(SS)[5]
Department Of Computer Science and Engineering,
Periyar Maniammai Institute of Science and Technology, Thanjavur 613403, India

ABSTRACT

Fruits vendors are now-a-days facing the problems in classifying fruits in the fruit packs. Fruits may be damaged due to climatic conditions, transportation and external factors, etc. This survey is focused on classifying the fruits with the aid of machine learning. Fruit vendors could also be able to find out the healthy and unhealthy fruits bought from the farms. SVM algorithm is used in the system for an efficient fruit classification. Vendors purchasing fruits from farmers or wholesale retailer can reduce the loss occurring due to purchase of defective fruits. The fruit vendors could be benefited out of this classification system by avoiding wastages due to damaged or unhealthy fruits.

KEYWORDS: Machine Learning, SVM Algorithm Classification, Artificial Neural Network

INTRODUCTION

Humans need to survive amongst the world of machines which are even trained to replace humans. The system is given the capability to learn from experiences without being programmed explicitly. Machine learning has the structural and hierarchical methods which are used for performing the tasks. It uses neurons for input, output and hidden layers. The number of layers are increased to reduce the error rates and attaining higher accuracy. As the amount of data provided increases, the performance of the system is increased in machine learning.

There are two types of machine learning: Reinforcement, Supervised and Unsupervised. Supervised machine learning includes labeled dataset and Unsupervised Learning includes unlabelled dataset. It is helpful to carry accurate result for huge dataset. Reinforcement includes learning in which output is based on the positively predicted dataset. It might take more time to train but testing is done in short period of time only. It provides advantage of faster detection of defects. It also analyse data in large quantity.

From pile of Fruits, healthy fruits are been classified automatically using SVM algorithm. Basically in this process according to healthy and defective , they are been classified and sold. Healthy Fruits are been sold first. The fruits are being transported to various regions to reach the vendors for sales, there is a chance that the fruit may get affected due to temperature, climatic changes and other external factors. The delivery of healthy fruits to the customers is a critical task for all the shopkeepers. The farmers harvest the fruits in different ripening stages for each fruits. Some may be pre-harvested or pre-matured, unripened fruits, over-ripened fruits. So the shopkeepers must decide and classify which fruits to be sold immediately according to the ripening stage and the fruits to be stored and sold to the buyers. This process is done by the Machine Learning which is a step-by-step process. First of all, Images need to be required through which system will be trained. Second step is to Pre-processing of images in which noise is removed ,contrast ,brightness is adjusted for better classification. Next step is Segmentation in which images will be broken into different blocks. Now , from each blocks features will be extracted and system will be trained. For classification, SVM Algorithm is used in this paper. This algorithm is chosen because it gives highest level of accuracy with minimum processing capacity. Input is taken from the training dataset(labeled) and output is calculated based on maximum hyperplane. SVM gives an advantage when used in high dimensional spaces and also uses less Memory. It also has a limitation i.e. it is not applicable for large dataset.

Fruit classification can be used by vendors and it will be useful for them and it will also reduce the loss in business. Not only this , it will also reduce time and will have benefit to the business.

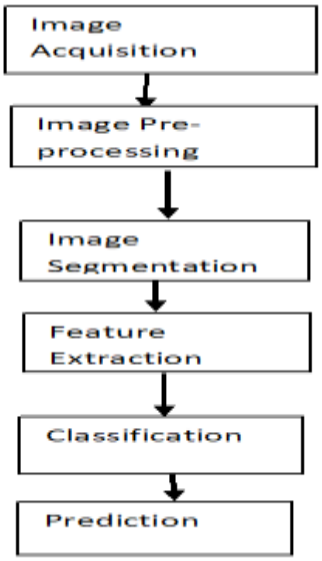

Fig.1 Flow chart for Fruit classification using Machine Learning.

Fig 1 is the general model which includes description of working of machine learning model. It shows step by step how image is processed and classified for fruit classification.

LITERATURE SURVEY

[1] Rabby, M.K.M.,et.al. (2018) processed the digital image using Modified Canny Edge Detection (MCED) algorithm. The image is acquired from the pile of fruits which is then pre-processed to obtain gray image and fed as input. Bi-cubic interpolation is used for resizing the images for conversion to gray scale image. The image is segmented into various parts for extracting more number of features. The color intensity and size of the fruit is calculated from the features. Traditional approach may be replaced with MCED for higher computational speed and accuracy.

[2] Gonzalez, J.P.B., et.al.(2017) proposed computer vision system to automatically estimate the characteristics in passion fruit. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) algorithms were used for statistical analysis. ANN is used as a tool for estimation which extracts the parameter functions. The features are extracted and correlation of size, shape and texture is analysed with the defined model for producing the output. Characteristics acquired was more but the representation of features are quite complex. The accuracy acquired was above 80% and error rate below 20%. The accuracy depends on the correlation coefficient(R).

[3] Ponce, J.M., et.al.(2019) proposed a system for olive industry, fruit grading plays a major role after harvesting the fruits where the fruits are classified based on the size and mass for high quality olives using regression analysis. The input is fed after image data acquisition and reference data acquisition, the acquired image is then pre-processed to remove the noises. Olives are then segmented individually using Watershed Transform as the olives may collide on each other which reduces the accuracy. The images after being segmented are post-processed for extraction of features. The Regression analysis is performed on the features and external validation is done for efficient grading which produces the result with higher accuracy. Trained models are used for size and mass estimation of olives. The error rates were below 0.80% and 1.05%.

[4] Kumari, R.S.S., et.al. worked with descriptors for classification phase. The system contains three phases: pre-processing, feature extraction and classification phase. Background subtraction is one of the steps in pre-processing. In feature extraction phase, statistical features of color and texture from co-occurence matrix are derived. In classification phase, Support Vector Machine (SVM) classifier is used for classification of both linear and non-linear datasets with the help of support vectors. The accuracy was achieved upto 95.3%.

[5] Akin, C., et.al. (2012) proposed a color based analysis technique**.** The fruits in the tree are detected using color-based technique. Pomegranate in trees are used as a dataset. The source image (Pomegranate tree) is loaded and the color index is applied for segmentation. The system uses local and shape analysis for analysing the color intensity for identifying the desired source. The estimated objects are counted. Out of 100 pomegranate spotted by human eye the system was able to identify 96 fruits from the tree. The system could be used in agricultural fields. The accuracy rate is 99%.

This paper proposed A system of automated classification system for classifying varioustypes of dates from the digital images was proposed by Muhammad, G.,(2014). SVM algorithm is used for classification. The pre-processing is the first step and the image is decomposed into color components to identify the size and shape features from the image. Descriptor fusion is the next step for selecting the features such as shape, size and texture. Classification is done using SVM along with the classified date. The system achieved accuracy rate of 97.12% .[6]

[7] Riyadi, S., et.al.(2008) The features are extracted using wavelet based method which efficiently extract the unique features. The classification task is performed from the extracted

statistical properties of the features for discriminating unhealthy papaya fruits amongst healthy fruits with the help of image processing technique. The algorithm used here is LDA (Linear Discriminant Analysis). LDA is used to classify well-formed and deformed shapes of fruits. The accuracy achieved is more than 98%.

[8]SwatiDewliyaet.al. (2015) did research with Multi class support vector machines for training and classification. Feature extraction is done using histogram of chain code and pixel's density of pixels. Training and Classification is performed on the multi class SVM kernel. The system has achieved more than 98% classification accuracy.

[9]Ronald M. et.al. (2016) investigated the Naive Bayes algorithm for classification of apple fruit. Image segmentation and Pre-processing is carried by transforming RGB to grayscale images and filtering the images to remove noise. The acquired average values of parameter such as accuracy, sensitivity, precision and specificity were 91%, 77%, 100% and 80%.

[10]Macanhã P.A.et.al.(2018) did a research using Zoning and Character-Edge Distance methods with Discrete Fourier Transform. In this, all feature spaces were classified with Multilayer Perceptron and k-NN classifiers from Weka.The best feature space computed with this approach reaches 97.54% of correct classification rates.

[11]Zihai Lua,et.al.(2017)proposed a system withFractional Fourier entropy (FRFE)which tremendously diminish the feature space and multilayer perceptron (MLP), which was used as the classifier and genetic algorithm (GA) and Improved hybrid genetic algorithm.MLP is used for classification and function approximation. The optimization is done using Standard genetic algorithm (GA). The proposed method resulted into an overall accuracy of 89.59% in different categories.

[12]Rocha, A. et.al.(2010) worked with Statistical, structural and spectral approaches with Supervised learning and K-Nearest Neighbors (K-NN) .The objects are described using statistical approaches. The object appearance is described using primitives like the patches in the part of a displayed object in structural approaches. Fourier spectrum is done in spectral approaches.The implemented solution achieves a classification error less than 2%.

[13]Akif Birol et.al.(2016) proposed LabVIEW vision development in the study. Real-time images were used as dataset. A data acquisition (DAQ) card was used to control the system. A LabVIEW-based proportional–integral–derivative (PID) controller was used to enable real-time DC motor speed control. The optimum parameters and results were attained in response to the five different belt speeds (0.518, 0.691, 0.749, 0.806 and 0.864 m/s).

[14]Raja Sekar et.al.(2018) research was based on Artificial Neural Network, Adaptive Neural Fuzzy Interference &SVM .SVM (Support Vector Machine) gave highest accuracy, but ANFIS (Adaptive Neuro Fuzzy Interference System) showed the best result out of these techniques. The lowest accuracy rate was given by Fuzzy system. The implementation was easy when fuzzy model is used.

A system with recursive elimination of featurestechnique combined with SVM algorithm was used for prediction system. Image frames of 27 features of mangoes (16400 images) were provided as input. Images werepre-processed using wiener filter to eliminate themotion blur and noises. Thecolor based features were extracted fromthe pre-processed image.Their main scope was to predict the maturity level of mangoes . [15] Nandi, C.S. ,et.al.(2014) proposed system with the average accuracy in classification phase was obtained up to 96%.

[16] Leekul, P.,et.al.(2016) proposed a system that used non-invasive and a non-destructive system embedded with sensors to classify the durian fruits for classifying maturity stages with the help of features.The principles of wireless communication and Rician k-factors were used to analyse the maturity stage of the fruit.The durian fruits hadbeenanalysed by measuring the scattered waves at 433, 915 and 2450 mhz.The accuracy rate achieved upto 92.7% according to the maturity of durian fruit.

Dubey, S.R.,et.al.(2016) had appliedimageprocessing-basedmethods which was used for detecting infected part of fruit .K-means

clustering method is asloused in detection . They have used digital images of apple as an input.Shape , colorand texture based features were extractedand multi-class SVM was used to identify,whether the fruit was infected or healthy fruit . The accuracy achieved was 95.94%.[17]

[18]Dubey, S.R., et.al. (2012) had used Improved Sum and Difference Histogram(ISADH) system. Algorithm used was SVM. Texture based features were extracted using difference and sum of the depth values of nearby pixelsby using color images. Images of kiwi,orange,onion, potato and various type of fruits and vegetables weretaken as input .Image segmentation is done for feature extraction. The output of the feature was more discriminative. Accuracy achieved upto 99%.

[19]Marimuthu.S.,et.al.(2017)proposed Computer Vision based system and Regression tree algorithm. 3108 images of banana were taken as input to this system. Data pre-processing and feature is combined and extracted using destructive feature extraction method.Fuzzy Model Classification Banana Ripeness (FMCBR) method has attained 94.85% of classification rate.

[20]Tao, Y., et.al. (2017) worked with Color-FPFH (animproved3-dimension descriptors) by the combination of color features and geometry3-dimension features wereextracted from the pre-processed cloud points. Algorithm used was Genetic Algorithm (GA),Support Vector Machine (SVM) and Random Sample Consensus (RANSAC) algorithmwere used to construct an apple model for training.Images were segmented and computer 3D descriptors were attained from which the training dataset was trained. SVM and Genetic Algorithms were used for classification and the results were evaluated. Accuracy is 92.30%.

Zhang, Y. et.al.(2014) had applied hybrid classification method in this system. The algorithms such as Fitness-Scaled Chaotic Artificial Bee Colony, FNN, PCA were used in the system.[21] In this 1653 colorimages of fruit, from the 18 categories were attained as input . Images of fruits were acquired and background removal was done by split and merge algorithm. Fruit texture, shape,and color based Features were extracted using PCA. Accuracy rate of 89.1% was achieved using this system.

S., Litananda,.et.al. (2016) utilized Gas sensor for identification and classification of fruit. Gas sensor was used to detect or operated under a controlled temperature and data analysis software.Itwas controlled by using a neural network. 3 kindsof fruits(durain,jackfruit,pakel mango) were taken as input.The best accuracy could be obtained by combining principal component analysis (PCA) Pre-processing for extractingchromatogram of fruit.The unique patterns amongst the samples was extracted using the chromatograms from which the repeatable and dependable results were generated.It couldidentified the three distinct flavours with the level of accuracy of 82%.[22]

[23]Pholpho, T.,et.al. (2011) proposed classification system works with aid ofspectrometer which uses visible spectra to classify the longan fruits with bruise and non-bruise spots. The input images of the fruit was approximately 400 fruits. The algorithms implemented were Principal Component Analysis (PCA), Soft Independent Modeling of Class Analogy and PLSDA were used to generate classification of normal and bruised longan fruits .By usingthe visible spectra of the spectrometer,classification of bruised and non-bruised fruits were obtained with 10 nm resolution. The accuracy rate of PLS-DA was higher than compared to other algorithms.

The system worked with algorithms such as Partial Least Squares, Support Vector Machine (SVM) and Back Propagation Neural Network.210 fruit images were taken as input. The range of multispectral imaging must be within 405-970 nm with 19 wavelengths. [24]The moisture content, soluble solid contents in total and the pH level in the strawberry fruit was determined using hyperspectral imaging in between visible and near-infrared region. The higher accuracy rate of 100% was achieved by SVM model. This system was proposed by Liu, C.,et.al.(2014)

[25]Kheiralipour, K.,et.al. (2017) proposed a system that has the advantage of higher accuracy.102 cucumber fruits were taken as input.Image processing technique and artificial neural network methods are used for detection. A new algorithm was introduced for pre-processing

and MATLAB 2010a software was used for extraction of shape features from the images.Two features such as centroid with non homogeneity and width non homogeneity were extracted. Many models of ANN were used to evaluate thefeatures selected after extraction to classify the effective features.Cucumber sample was put on a white background paper under a same illumination system including two florescent and four LED lamps. The finest classifier model reached the accuracy rate of 97.1%.

[26]Jana, S.,et.al. (2017) proposed with Image analysis and pattern recognition methodologies were used in this system. 210 images of 7 fruit classes were taken as input. The image needs to be normalized by pre-processing step proportionate to fold stratified cross validation (SCV).1653 fruit images from 18 categories. The accuracy achieved was 89.5% which has

appearance descriptors to categorize fruits.The descriptors are used to classify the fruits. The the deviation in translation, scaling, dimensions, rotationandfeaturesutilized must not vary due to distances, growth stages of image and surface appearances of fruits. Test dataset used Naïve Bayes classifier and KNN classifier. Accuracy ranges from 88–95%.

[27]Wang, S.,et.al.(2015) worked with machine-learning based classification methods. The system developed for classification consists of Wavelet Entropy, Principal Component Analysis. The functions were optimized using Bio-geography Optimization. The feature extraction and training was done effectively with the help of these techniques. The statistical analysis was done with the help of K-system.

Rocha, A.,et..al.(2010) worked with Naïve method in this system.2633 images were taken as input. Analyses statistical color and texture descriptors as well as structural results obtained shows that the introduced system can reduce the classification error upto 15%.[28]

COMPARITIVE STUDY

Algorithm	Advantages	Disadvantages	Datasets
Support Vector Machine(SVM)	It is scalable to high dimensional data.	The error percentage will increase if incorrect support vectors are chosen.	16400 fruit images were taken as input.
Linear Discriminant Analysis(LDA)	It is often used in dimensionality reduction .	Prediction accuracy rate is comparatively low.	100 images were taken as input.
Principal Component Analysis(PCA)	Pre-requisite details about the data is not necessary.	Reduces the overfitting in models.	1653 images of different categories of fruit were taken as input.
Naïve Bayes	Less complex and implementation is easy.	Variables dependency exsist.	50 images of each fruits were taken as input.
Regression Analysis	Prediction is sharp because of the variables.	It is limited in the case of linear relationship.	3108 images were taken as input.
K-NN Classifier	There is no training period for the models generated.	Unscalable when the dataset is massive.	210 images were taken as input.

CONCLUSION

Delivery of healthy fruits and detection of fruit is the equitable aim of fruit classification. The fruits are captured and converted into digital images which are then classified based on color, texture and shapes with the help of various algorithms and training models. The accuracy range varies from each method according to the training data set provided to the model. The above survey discusses about the assorted methods prevailing in machine learning and deep learning techniques

REFERENCES

[1]Rabby, M.K.M., Chowdhury, B. and Kim, J.H., 2018, December. A Modified Canny Edge Detection Algorithm for Fruit Detection & Classification. In 2018 10th International Conference on Electrical and Computer Engineering (ICECE) (pp. 237-240). IEEE.

[2]Gonzalez, J.P.B., Ortiz, F.A.P. and Rodriguez, C.P.P., 2017. Mass and volume estimation of passion fruit using digital images. IEEE Latin America Transactions, 15(2), pp.275-282.

[3] Ponce, J.M., Aquino, A., Millan, B. and Andújar, J.M., 2019. Automatic Counting and Individual Size and Mass Estimation of Olive-Fruits Through Computer Vision Techniques. IEEE Access, 7, pp.59451-59465.

[4] Kumari, R.S.S. and Gomathy, V., 2018, February. Fruit Classification using Statistical Features in SVM Classifier. In 2018 4th International Conference on Electrical Energy Systems (ICEES) (pp. 526-529). IEEE.

[5] Akin, C., Kirci, M., Gunes, E.O. and Cakir, Y., 2012, August. Detection of the pomegranate fruits on tree using image processing. In 2012 First International Conference on Agro-Geoinformatics (Agro-Geoinformatics) (pp. 1-4). IEEE.

[6] Muhammad, G., 2014, October. Automatic date fruit classification by using local texture descriptors and shape-size features. In 2014 European Modelling Symposium (pp. 174-179). IEEE.

[7]Riyadi, Slamet & Ishak, Asnor & Mustafa, Marzuki & Hussain, Aini. (2008). Wavelet-based feature extraction technique for fruit shape classification. Proceeding of the 5th International Symposium on Mechatronics and its Applications, ISMA 2008. 1 - 5. 10.1109/ISMA.2008.4648858.

[8] Dewliya, S. and Singh, M.P., 2015. Detection and classification for apple fruit diseases using support vector machine and chain code. International Research Journal of Engineering and Technology (IRJET), 2(04).

[9] Misigo, Ronald & Evans, Miriti. (2016). Classification Of selected apple fruit varieties using Naive Bayes. Indian Journal of Computer Science and Engineering (IJCSE). 7. 13.

[10] Macanhã, P.A., Eler, D.M., Garcia, R.E. and Junior, W.E.M., 2018. Handwritten feature descriptor methods applied to fruit classification. In Information Technology-New Generations (pp. 699-705). Springer, Cham.

[11] Lu, Zhihai & Lu, Siyuan & Wang, Shuihua & Li, Yujie & Zhang, Yu-Dong & Lu, Huimin. (2017). A Fruit Sensing and Classification System by Fractional Fourier Entropy and Improved Hybrid Genetic Algorithm. 293-299. 10.12792/iciae2017.053.

[12] Rocha, A., Hauagge, D.C., Wainer, J. and Goldenstein, S., 2010. Automatic fruit and vegetable classification from images. Computers and Electronics in Agriculture, 70(1), pp.96-104.

[13] Dumanay, A.B., Sakin, R., & Istanbullu, A. (2016). A New Design of Olive Fruit Sorting Machine Using Color Image Processing.

[14] Raja Sekar L, Ambika N, Divya V and Kowsalya T "Fruit Classification System Using Computer Vision: A Review" Published in International Journal of Trend in Research and Development (IJTRD), ISSN: 2394-9333, Volume-5 | Issue-1 , February 2018.

[15] Nandi, Chandra & Tudu, Bipan & Koley, Chiranjib. (2014). A Machine Vision-Based Maturity Prediction System for Sorting of Harvested Mangoes. Instrumentation and Measurement, IEEE Transactions on. 63. 1722-1730. 10.1109/TIM.2014.2299527.

[16] Leekul, Prapan et al. "Rician k-Factors-Based Sensor for Fruit Classification by Maturity Stage." IEEE Sensors Journal 16 (2016): 6559-6565.

[17] Dubey, Shiv Ram & Jalal, Anand. (2016). Apple disease classification using color, texture and shape features from images. Signal Image and Video Processing. 10. 819-826. 10.1007/s11760-015-0821-1.

[18] Dubey, S.R. and Jalal, A.S., 2012. Robust approach for fruit and vegetable classification. Procedia Engineering, 38, pp.3449-3453.

[19] Marimuthu, S. and Roomi, S.M.M., 2017. Particle swarm optimized fuzzy model for the classification of banana ripeness. IEEE Sensors Journal, 17(15), pp.4903-4915.

[20] Tao, Y. and Zhou, J., 2017. Automatic apple recognition based on the fusion of color and 3D feature for robotic fruit picking. Computers and electronics in agriculture, 142, pp.388-396.

[21] Zhang, Y., Wang, S., Ji, G. and Phillips, P., 2014. Fruit classification using computer vision and feedforward neural network. Journal of Food Engineering, 143, pp.167-177.

[22] Ciptohadijoyo, S., Litananda, W.S., Rivai, M. and Purnomo, M.H., 2016. Electronic nose based on partition column integrated with gas sensor for fruit identification and classification. Computers and Electronics in Agriculture, 121, pp.429-435.

[23] Pholpho, T., Pathaveerat, S. and Sirisomboon, P., 2011. Classification of longan fruit bruising using visible spectroscopy. Journal of Food Engineering, 104(1), pp.169-172.

[24] Liu, C., Liu, W., Lu, X., Ma, F., Chen, W., Yang, J. and Zheng, L., 2014. Application of multispectral imaging to determine quality attributes and ripeness stage in strawberry fruit. PloS one, 9(2).

[25] Kheiralipour, K. and Pormah, A., 2017. Introducing new shape features for classification of cucumber fruit based on image processing technique and artificial neural networks. Journal of Food Process Engineering, 40(6), p.e12558.

[26] Jana, Susovan & Parekh, Ranjan. (2017). 2017 : Shape-based Fruit Recognition and Classification. 10.1007/978-981-10-6430-2_15.

[27] Wang S, Zhang Y, Ji G, Yang J, Wu J, Wei L. Fruit Classification by Wavelet-Entropy and Feedforward Neural Network Trained by Fitness-Scaled Chaotic ABC and Biogeography-Based Optimization. Entropy. 2015; 17(8):5711-5728.

[28] Rocha, A., Hauagge, D.C., Wainer, J. and Goldenstein, S., 2010. Automatic fruit and vegetable classification from images. Computers and Electronics in Agriculture, 70(1), pp.96-104.