Scalable Computing: Practice and Experience, ISSN 1895-1767, http://www.scpe.org
© 2021 SCPE. Volume 22, Issues 3, pp. 302–312, DOI 10.12694:/scpe.v22i3.1856

BIG DATA ANALYTICS FOR ADVANCED VITICULTURE

JITALI PATEL,[*] RUHI PATEL,[†] SAUMYA SHAH,[‡] AND JIGNA PATEL[§]

Abstract. Big data analytics involve a systematic approach to find hidden patterns to help the organization grow from large volume and variety of data. In recent years big data analytics is widely used in the agricultural domain to improve yield. Viticulture (the cultivation of grapes) is one of the most lucrative farming in India. It is a subdivision of horticulture and is the study of wine growing. The demand for Indian Wine is increasing at about 27% each year since the 21st century and thus more and more ways are being developed to improve the quality and quantity of the wine products. In this paper, we focus on a specific agricultural practice as viticulture. Weather forecasting and disease detection are the two main research areas in precision viticulture. Leaf disease detection as a part of plant pathology is the key research area in this paper. It can be applied on vineyards of India where farmers are bereft of the latest technologies. Proposed system architecture comprises four modules: Data collection, data preprocessing, classification and visualization. Database module involves grape leaf dataset, consists of healthy images combined with disease leaves such as Black measles, Black rot, and Leaf blight. Models have been implemented on Apache Hadoop using map reduce programming framework. It applies feature extraction to extract various features of the live images and classification algorithm with reduced computational complexity. Gray Level Co-occurrence Matrix (GLCM) followed by K-Nearest Neighborhood (KNN) algorithm. The system also recommends the necessary steps and remedies that the viticulturists can take to assure that the grapes can be salvaged at the right time and in the right manner based on classification results. The overall system will help Indian viticulturists to improve the harvesting process. Accuracy of the model is 82%, and it can be increased as a future work by including deep learning with time-series grape leaf images.

Key words: Big Data, Viticlture, Disease Detection

AMS subject classifications. 68T09

1. Introduction.

India has a population of about 1.38 billion people. Out of those, around 60% of the population is employed in agriculture and it accounts for approximately 20% of the country's GDP. Most of these people belong to the rural areas and do not have access to any kinds of technologies or smart devices. As the demand for agriculture is growing with the increasing population, many farmers are slowly adopting the means of smart agriculture. In smart agriculture, farmers can monitor a wider range of crops, implement a dynamic irrigation system, find and stop diseases before most of the crops are lost. Big data analytics help in all these tasks, especially in the case of disease detection in the crops.

1.1. Motivation.

Big data is of immense importance these days. Because of the industrial revolution shift to the 4.0 stage, there has been the incorporation of technologies like the Internet of Things(IoT) and several other technologies. All these technologies work harmoniously to create a big data pool, and it is up to us to process this data and make it useful for the people. Big data can be understood as a variety of data collection that can be of structured, semi-structured, and unstructured nature [1].

In the big data context, the 3 types of data can be classified as:-

1. Structured data: It is the type of data that is available as a well-formatted repository database like spreadsheets, relational databases etcetera.

2. Semi-structured: This type of data is not in a properly formatted fashion, but it has some level of organized structure like XML(eXtensible Markup Language) data.

3. Unstructured: This type of data generally includes not organized data like multimedia data such as audio, video images fall under this category.

[*] Assistant Professor, CSE Department Nirma University(jitali.patel@nirmauni.ac.in).

[†] Student, CSE Department ,Nirma University(17bit097@nirmauni.ac.in).

[‡] Student, CSE Department, Nirma University (17bit101@nirmauni.ac.in).

[§] Assistant Professor, CSE Department Nirma University(jignas.patel@nirmauni.ac.in).

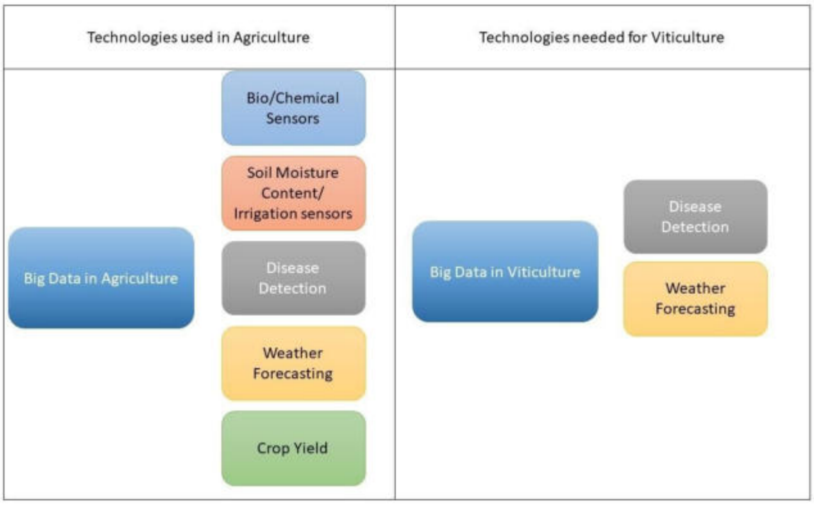

Fig. 1.1: Key research areas of agriculture and viticulture

Agriculture is the backbone of the Indian economy and contributes to up to 23% to Indian Gross Domestic Product (GDP) [2]. The inception of big data to the field of agriculture is relatively a novel concept and has not been explored extensively. Hence, this paper delves into the sphere of using big data for agriculture and in particular, viticulture. Some of the reasons for adopting big data in agriculture are that it helps farmers with planned harvesting and provides them with tools for better prediction, providing them with proper guidance and suggestions of possible remedies and fertilizers [3] [4] [5]. For these reasons, the application of big data in agriculture is of increasing importance.

In this paper, we proposed a method for viticulture. Key research areas of agriculture and viticulture have been shown through Figure 1.1. The Indian subcontinent's climate ranges from tropical to temperate and it does not lie in the cold weather belt, which is favorable for wine production and consequently for viticulture. Despite not having the optimum weather conditions, Indian states like Maharashtra, Karnataka, Telangana, and Punjab have been highly successful with the production of grapes. Maharashtra alone accounts for more than 80% production of grapes in India [6]. With the arrival of Industry 4.0, the farming sector has increasingly incorporated precision farming techniques that help them maximize their throughput with automation, the Internet of Things, and other information technology. Precision viticulture (PV) helps in developing optimum viticulture techniques. It allows for many features like selective harvesting which could provide benefits that justify the input cost by the cultivators. PV creates a surge in the amount of data available to the farmer and big data technologies are required to convert this raw data to valuable information that the farmer can easily interpret [7]. Apart from the farmers, a good viticulture practice can benefit various stakeholders of its supply chain like having a high quality of the wine is beneficial for the winemakers; healthy grapes are also good for consumption by the end-user as well the phenolic compounds in grapes are used by the cosmetic industry [8].

To ensure that the information from big data is retrieved in the most efficient way possible there is a need to use effective machine learning algorithms to detect patterns within the data. Also, there is a need to make the output of such algorithms transparent and easily understood by the operator. Moreover, it is also beneficial that such algorithms are computationally feasible so that the applications can be successfully constructed [1] [9]. In this paper, we are using KNN to detect grape leaf diseases, the motivation behind choosing the K-Nearest Neighborhood (KNN) algorithm is that it allows for a reduced recognition time as well as reduced computational complexity (Figure 1.2). Reduced computational complexity is of key importance as the primary users of this system would be farmers who may not have access to advanced devices that can carry out high complexity algorithms [10].

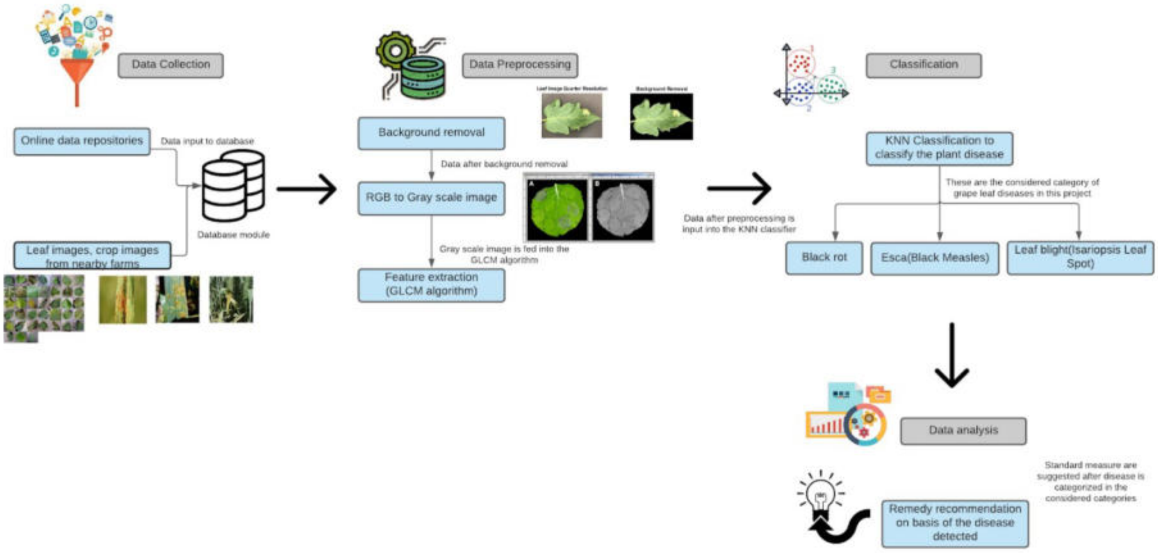
Data Collection
Online data repositories
Data input to database
Database module
Leaf images, crop images from nearby farms
Data Preprocessing
Leaf Image Quarter Resolution
Background Removal
Background removal
Data after background removal
RGB to Gray scale image
A
B
Gray scale image is fed into the GLCM algorithm
Feature extraction (GLCM algorithm)
Data after preprocessing is input into the KNN classifier
Classification
1
2
3
KNN Classification to classify the plant disease
These are the considered category of grape leaf diseases in this project
Black rot
Esca(Black Measles)
Leaf blight(Isariopsis Leaf Spot)
Data analysis
Standard measure are suggested after disease is categorized in the considered categories
Remedy recommendation on basis of the disease detected

Fig. 1.2: Methodology

1.2. Contribution. For our approach, we have taken the vineyards in Maharashtra, India, as they are the country's biggest vineyards. Vineyards have become an important part of our current economy. Along with the production of grapes for wine, vineyards have become the first step of the supply chain of many cosmetic products, food and wine producers and packagers, bottlers and distributors. Supply chains are now overseeded and improved with the help of big data analytics. This is because each and every step will generate lots of data. However, big data is not being applied on a large scale in vineyards, which is the very first step. Big data is important to implement smart farming so that the different agricultural challenges of production can be tackled

Big data is being applied for the creation of smart vineyards in some parts of the world. Most of these companies that provide smart vineyard solutions are located in Europe. All of them focus on these three solutions that gather big data and after analyzing, work towards decreasing the yield loss and decrease the working hours of the viticulturists as per [11].

1. Precision Sensors: These sensors are deployed between the grapevines, and they gather the acute weather details of the grape plant such as leaf moisture and humidity.
2. Microclimate Monitoring: Precision monitoring of climatic conditions per plant measuring its humidity so that monitoring the intensity of fungal diseases becomes possible.
3. Grape Disease Prediction: Monitoring the leaf images of the grape plant can help to identify the early onset of diseases which can be treated or the bad plant may be removed to protect the rest.

Big data is slowly being applied to the agricultural sector in India. While farming may have other factors like irrigation and soil moisture, for vineyards, weather forecasting, and disease detection are the main areas that require the help of analytics as per [4]. In this paper, we have worked towards implementing a machine learning algorithm on a big data platform that can distinguish between healthy plants and disease ridden grape plants.

A feature extraction algorithm along with a machine learning classification algorithm is used alongside big data technologies to classify the leaf on whether it is disease ridden or not which will provide an accuracy between 60-70%. The reason for using a machine learning algorithm instead of a deep learning algorithm which would provide 100% accuracy is to create a system that would be able to give instant results with maximum possible accuracy and can be implemented without any extensive hardware requirements. This is done so that the system can easily be made use by the farmers who do not have that much technology to run the high-end

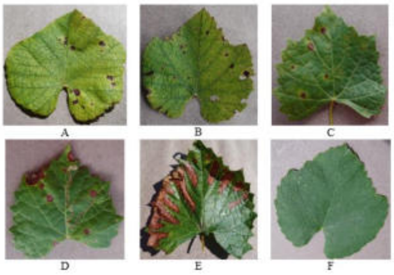

Fig. 3.1: Leafs

software.

The importance of big data technologies along with machine learning in the application of smart viticulture is given in this section of the introduction. The dataset used and literature reviewed is discussed in section 2. Section 3 talks about our proposed methodology and architecture in detail. Section 4 explains the implementation of our selected algorithm. Finally, section 5 shows the results of our implementation and the conclusion of the paper is given along with future scope and aspects.

2. Literature Review. Literature from reputed journals and conference proceedings have been referred to identify the suitable methodology and research gaps. Key objectives of the literature are as follows

1. Identification of data analytics algorithms to predict crop yield, crop production, crop price and crop protection from agriculture fields.
2. Review all grape disease detection algorithms

In Table 2.1, authors have figured out all the data analytics techniques implemented in the agriculture domain in the past 6 years. Table 2.2 depicts a comparison of the different algorithms implemented to detect disease in the grape plant and provide an analytical comparison with their advantages and disadvantages.

Table 2.1 given above provides a glimpse of what big data techniques have achieved in the field of agriculture in the last 6 years. Big data can help us predict crop yield [15] [17], crop production [14], crop price [18] and finally crop protection [16]. In this paper, we would be focusing on the concept of crop protection under big data in agriculture. We would specifically focus on big data in viticulture and grapefruit protection from diseases. To do this, we would be employing a machine learning model to predict the leaf disease in real-time collected from sensors in the vineyards.

Table 2.2 shows good accuracy offered by Convolutional Neural Networks [31] [32], but the major problem faced is the integration of these techniques with real-time data.

3. Proposed Architecture. In this paper, we have used the grape leaf dataset, which consists of images of healthy as well as grape leaves with diseases such as Black measles, Black rot, and Leaf blight. Moreover, we will also collect the data related to the diseases so that it can be used in the final recommendation of remedies for the detected disease(s).

Proposed model comprises four modules as per figure 4.1. Data collection module, data preprocessing module, Classification module and data analysis module. Data collection module deals with the gathered data(online and physically acquired) and will be stored in the database. The gathered data is then preprocessed and maintained by a preprocessing module. In the preprocessing the first step is to remove the background from the leaf images which will give us only the selected leaf portion, after this step the RGB image is converted to Grayscale image so that it can be processed with GLCM algorithm. The GLCM algorithm stands for Gray Level Co-occurrence matrix. This will help to convert the image data to numerical data, various features extracted from the image can be (contrast, correlation, energy, homogeneity, and entropy)[33]. As per Classification module achieved numerical data can be fed into the KNN classifier to classify images on the basis of the extracted features the dataset can be split into training, testing and validation as 80% train, 10% validation, 10% test. Visualization module proposes to recommend the necessary steps and remedies that the viticulturists

Table 2.1: Big data techniques in agriculture comparison

Data Source	Contribution	Pros	Cons
AkkerWeb Data for Satellite data and DairyCampus, Netherlands data [12]	Ontology is made of use for a combination of sensor data and applying big data techniques to the dairy farming	A system was created by the researchers that are based on the principle that ontology can be made use of to handle the big-data questions and create SPARQL federated queries on the data sources used by making use of ontology matching. As a result, farmers can pose questions in terms of the common ontology concepts instead of the detailed and specific concepts of the DairyCampus and Akkerweb data sources.	More advanced algorithm with hybrid approach could be applied
Web of Science and Scopus Database with search queries related to big data and farming [13]	Applied Ontology models as well as Machine learning models such as ANN, SVM and provided a deep comparison.	Made a list of all the insights got from the detailed study that can give rise to future scope and creation of applications	Not provided any suggestive list of all the challenges faced while implementing.
A dataset in agricultural sector Crop vise agricultural data Agricultural data of different districts Agriculture based on weather, temperature, and relative humidity [14]	wrote a paper that has its main focus on analyzing the agricultural data to find out the optimal parameters to maximize the production of crop by using data mining techniques such as Multiple Linear Regression, CLARA, PAM and DBSCAN.	A range of 5 different crops were selected with their yield data of over 6 years to provide as key measure. Comparison models were done between the different crops as well as the methods.	The research was limited to external quality metrics and not the internal ones too. The crops were selected only based on its economic importance, and not other factors like topography were included.
Cloud Database in Agriculture that gets data from sensors. [15]	Proposed a model to analyze the crop yield by combining the methodologies of IoT, cloud and big data concepts to deliver the prediction attributes to the farmers through the mobile computing technology. This is done using multiple nodes in MapReduce in Hadoop to predict the outcomes accurately at low costs.	The model will predict and notify the farmers regarding how much fertilizer to use for the crop and when to use it. This prediction will also try to decrease the total cost as much as possible.	They did not provide any insights on irrigation or how to deal with crop diseases. The focus of sensors was just on the amount of fertilizers to use.
A data source is from Charles Stuart University, Australia - herbicide resistance testing service from the years 2001-2005. Along with it, agricultural survey data is taken from the Australian Bureau of Statistics for each shire. [16]	TThe paper focuses on crop protection using big data with a focus on weed control and management using different machine learning techniques	A comparison between various machine learning approaches including discriminative/generative and supervised/unsupervised was done.	Samples were specific to just ryegrass. Out of the 173 shires available, only 121 were utilized to create the final dataset to avoid bias which could also have been removed using scalability.
Data collected from Onsite and remote farming [17]	The paper presents a hybrid model that first implements Grey Wolf Optimization on SVM classification to improve its results.	The proposed model has better precision, recall, accuracy and F1 values than a regular SVM classification.	No other optimization techniques were used as a comparison.
Created the dataset by finding the price of corn grain products in recent 10 years and the corn balance sheet for each year. [18]	Applied multiple linear regression to predict the price fluctuation in corn price under big data more accurately	Chinese special circumstances and the domestic and international market price of corn are also considered by the multiple linear regression model implemented in this paper to estimate the parameters by linear regression of corn prices.	No other regression techniques were used as a comparison

Table 2.2: Comparison of grape disease detection algorithms

Major Techniques	**Data Source**	**Pros**	**Cons**
First image processing is applied on the images and then classification of diseases is done using BPNN [19]	PlantVillage Dataset	Various techniques to segment the disease part of plant image were discussed	Other methods of classification were not compared.
Image segmentation is done using K-means clustering and classification is done using SVM [20]	Manually collected from different regions of Maharashtra, India	Provides a automatic, fast, accurate and less expensive method to detect and classify the grape leaf diseases	Other segmentation algorithms or classification methods were compared
Opposite Colour Local Binary Pattern Feature and Multiclass SVM [21]	Manually taken from farms	Provides a comparison with other methods for classification and their accuracy.	The algorthim was not implemented in a real time system by the researchers.
Detection using K means clustering [22]	-	A survey on different classification techniques	Segmentation of the disease area is difficult with these algorithms
Disease detection and classification using multiclass SVM [23]	PlantVillage Dataset	Clustering and Classification is done in Matlab with LAB and HSI color models	The number of images used in the dataset is just 160 which is less for real time data analysis.
Improving Classification accuracy by spatial-spectral analysis of hyperspectral images [24]	Manually capturing images of grape from a non-commercial vineyard.	Tested different dimenstionality reduction methods to study performance of spatial-spectral segmentation using Random Forest classifiers	The algorithm needs to be validated in other controlled conditions
Individual leafs are identified using leaf skeletons and then leaf disease classification is done using KNN classification. [25]	Created using agrnomical crop images	As the individual leafs are identified before classifiying, this approach is more efficient than other real time systems.	Apart from the luminance and linear characteristics of leaves, other features aren't taken.
A Transfer learning approach is applied in which Features from Rectified Linear unit layer of AlexNet (CNN) applied to MSVM [26]	PlantVillage Dataset	A high accuracy of about 99.23% is achieved.	The hardware requirements are stringent with GPU and high RAM requirements.
First, a local contrast haze reduction (LCHR) enhancement technique is applied. Next, LAB color transformation is done. Color, texture, and geometric features are extracted and fused by canonical correlation analysis (CCA) approach. Noise is removed by Neighborhood omponent Analysis (NCA). The classification is done by M-class SVM. [27]	PlantVillage Dataset	Provides a comparison between different algorithms for feature extraction and between different algorithms for classification	There is degraded accuracy in case of complex images.
Different machine learning algorithms are applied for classification [28]	PlantVillage Dataset	Provides a detailed comparison on the accuracy of the different models.	Apart from accuracy, other items like computational need, ease of apply etc are not compared.
Image analysis and back propogation Neural Network [29]	Manually taken from a farm in Zhengzhou City, Henan Province.	5 different grape diseases could be identified with high classification accuracy	The dataset just consisted of 60 images which are relatively less to give accurate results when applied elsewhere.
Artificial Bee Colony algorithm was applied to pre processed images. [30]	Plant village dataset	Comparison is done with two other algorithms namely Particle Swarm Optimization and Genetic Algorithm. Relatively better accuracy is observed	The accuracy is dependent on a SVM based classification to which the ABC is applied. Other classification algorithms are not applied for comparison.
Lightweight convolutional neural network along with channelwise attention with ShuffleNet V1 and V2 as backbones [31]	PlantVillage Dataset	A comparison is given between all the different models of CNN along with their accuracy and parameters.	Although the improved shufflenet provides maximum accuracy, its still not able to integrate with IoT for real time big data monitoring. The computing is cost effective but only compared to other CNNs.
A UnitedModel convolutional neural network based on multiple CNNs [32]	PlantVillage Dataset	Created a united model architecture based on InceptionV3 and ResNet50. It outperforms VGG16, InceptionV3, DenseNet121 and ResNet50.	The integration of the two CNNs is still not proper. Trained model cannot be applied in real time diagnosis needed in crop protection

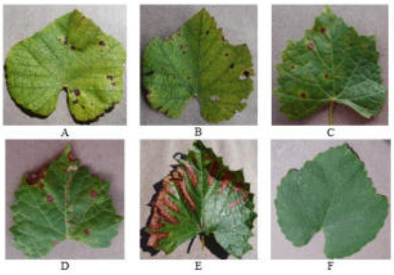

Fig. 4.1: Leafs

can take to make sure that the grapes can be salvaged at the right time and in the right manner and overall help them in the harvesting process.

4. Dataset Description. An important aspect of viticulture is to take care of the health of the grapevine. This paper focuses on three main diseases that afflict the grape plant, namely black rot, esca (black measles) and leaf blight (Isariopsis leaf spot). Black rot is a fungal disease that affects all the parts of the plants, i.e. leaves, berries, shoots and stems. The fungal disease spreads via spores and has to be contained during the first two weeks or all the plants grown for that season are lost. Fungicides need to be sprayed, and the rot needs to be detected as early as possible. Such rot especially spreads in warm and humid conditions as per [34] like the weather condition in Maharashtra, and thus it is important for our proposed algorithm to detect this disease early on to curb its growth. Esca or black measles affect the grapefruit and leaves and has plagued the viticulturists for a long time. With no apparent management strategy for measles, field crews are needed to check for them every day and remove the infected plant before the nearby plants can be infected. This problem can be solved with the help of big data analytics where each leaf is monitored, and our proposed algorithm can detect the measles along with the location, so that field crews just need to remove them. Isariopsis leaf spots is another fungal disease that if not controlled, will lead to a reduction in the population of the plant as per [35].

To deal with these three major diseases, the dataset was taken from Kaggle [36] in which there are 1180 images of black rot, 1383 images of esca, 1076 images of isariopsis leaf spots and 423 images of healthy grape leaves. Total of 6000 images were 4062 images taken from Kaggle and 1938 images taken from Ambapur village farms situated in Kalwan district, Maharashtra as shown in Figure 4.1.

5. Implementation. Execution of the proposed model is carried out on three nodes set up with commodity computers of 64 bit operating system with 8GB RAM, windows 10,core i5 CPU, 120GB hard disk(built in),Extra 1TB hard disk on specific node. One of the nodes in the cluster worked as Hadoop Master and the rest of the two nodes worked as Hadoop Slave. Selection of competitors is on MapReduce and parallel processing ground. The implementation details are as follows:

Step 1: At first the images of all the healthy and the diseased leaf images are combines and are split into 80:20 ratio of training to testing.

Step 2: After that the images are preprocessed, the background is removed and conversion of the RGB to Grayscale image takes place, the function cvtColor() is used to change the color space of the image.

Step 3: Then, the HSV(Hue, Saturation, Value) and their mean values are calculated, these values are then concatenated with the GLCM features in the final CSV sheet.

Step 4: The GLCM features used for this project are Contrast, Dissimilarity, Homogeneity, ASM(angular second moment), and Energy , these terms are calculated as given in equations 5.1, 5.2, 5.3, 5.4, and 5.5.

Contrast:

$$\sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \tag{5.1}$$

Dissimilarity:

$$\sum_{i,j=0}^{N-1} P_{i,j}|i-j| \tag{5.2}$$

Homogeneity:

$$\sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \tag{5.3}$$

ASM:

$$\sum_{i,j=0}^{N-1} P_{i,j}^2 \tag{5.4}$$

Energy:

$$\sqrt{ASM} \tag{5.5}$$

where $P_{i,j}$ is the element i, j of the normalized symmetrical GLCM and N is the number of gray levels in the image as specified by Number of levels in under Quantization.

Step 5: The GLCM calculation and feature extraction process:

GLCM: GLCM is used to calculate how frequently a pixel with gray with a certain gray-level grayscale intensity or level (grayscale intensity or Tone), namely value i occurs either horizontally, vertically, or diagonally to adjacent pixels with the value j. In the process of calculating the co-occurrence matrix we firstly quantize the image data. Each sample on the echogram is treated as a single image pixel and the value of the sample is the intensity of that pixel. Quantization is used to quantize the levels of intensity of the pictures by specifying the discrete gray levels.

5.1 Create the GLCM matrix, which is an $N \times N$ square matrix where N signifies the Number of levels specified under Quantization. The matrix is built as follows:

5.1.1 S are the samples taken for the calculation.

5.1.2 W are the samples around sample S that fall in a window centered around sample S with the size specified by the window size.

5.1.3 Considering only the samples in the set W, define each element i, j of the GLCM as the number of times two samples of intensities i and j occur in specified Spatial relationship (where i and j are intensities between 0 and Number of levels-1). The sum of all the elements i, j of the GLCM will be the total number of times the specified spatial relationship occurs in W.

5.1.4 To make a symmetric GLCM matrix, we create a transposed copy of the GLCM matrix.

5.2 Adding this copy to GLCM itself produces a symmetric matrix in which the relationship i to j is indistinguishable for the relationship j to i (for any two intensities i and j). As a consequence the sum of all the elements i, j of the GLCM will now be twice the total number of times the specified spatial relationship occurs in W (once where the sample with intensity i is the reference sample and once where the sample with intensity j is the reference sample), and for any given i, the sum of all the elements i, j with the given i will be the total number of times a sample of intensity i appears in the specified spatial relationship with another sample.

5.2.1 We normalize the GLCM by dividing each element by the sum of all elements. The final matrix contains elements that represent the probability of finding the relation between i and j in W.

5.3 Calculate the selected Feature. This calculation uses only the values in the GLCM. The features are given as follows:

Table 6.1: Calculation metrics

Test	**Positive**	**Negative**	**Total**
Disease Present	P(TP)	R (FN)	P+R(TP+FN)
Disease Absent	Q(FP)	S (TN)	Q+S(FP+TN)
Total	P+Q	R+S	P+Q+R+S

Table 6.2: Example table

Sr. No	**Measures**	**Values**
1	Accuracy	0.824
2	F1_score	0.84
3	Precision	0.846
4	Recall	0.77

1. Energy
2. Entropy
3. Contrast
4. Homogeneity
5. Correlation
6. Shade
7. Prominence

5.4 The sample s in the resulting virtual variable is replaced by the value of this calculated feature.

Step 6: At the end of pre-processing and feature extraction we are left with a CSV data sheet containing the selected GLCM and HSV values along with the label of the type of leaf.

Step 7: In the KNN code, the label from the CSV data are represented on the Y axis, whereas the features (GLCM and HSV) are represented on theX axis.

Step 8: The KNN algorithm used to find the nearest neighbor is the Brute force algorithm which calculates the distances between all pairs of points in the dataset. Efficient brute-force neighbors searches can be very competitive for small data samples.

Step 9: We also randomized the value of K and found the most accuracy when the value of K is set to 15 we get an accuracy of 72%.

6. Result and Discussion. As per table 6.1, TP represent true positive, TN represent true negative, FP represents false positive, and FN represents the false negative. The fraction of the population expresses accuracy of the model are rightly classified as disease affected leaves, and it is calculated using the formula $(P+Q)/P+Q+R+S$. With the application of the KNN model on grapes leaves, 82% accuracy have been achieved. The precision of the model is expressed by the fraction of true positive samples among all positive sample by the formula $P/P+Q$. Table 6.2 shows achieved results for the metrics accuracy, F1_score, precision and recall.

Disease type would be identified by a classification module. Based on the class of the grape leaf visualization and recommendation module would suggest primary level solutions.

1. Black measles: There is no operational solution to control black measles till date. Agriculture experts' solution is to remove the infected grapes, leaves and. Shield them by pruning wounds which spread minimum fungal infection using spiral sealant (5% boric acid in acrylic paint). Use of neem oil or suitable fungicides.
2. Leaf blight: Only sprinkle of fungicides would help in reducing this disease.
3. Black rot: Removal of all wrinkled grapes from vines during latent pruning. Also, cultivation of new soil during outgrowth break to bury mummies. One can apply suitable fungicides to control the disease.

7. Conclusion and Future Scope. In this paper, we have proposed an approach to implement big data analytics in vineyards which extracts the features from the leaf images using GLCM and then classifies them using KNN and with a primary level of recommendations. In our experiment dataset, the leaves were classified into three diseases, namely black rot, esca, leaf blight and healthy leaves. The final accuracy of our experiment comes to about 82% but this algorithm can be directly applied to a big data environment to create a smart vineyard where technological limitations are prevalent. This system can be easily implemented to a mobile application so that farmers in India can easily find out if their crops have any diseases or not. As GLCM and KNN do not require any high end graphic cards or high technical configuration to run, a smart vineyard system to monitor the grapes can be implemented easily at low costs. In the area of agriculture, the absence of demonstrative devices in underdeveloped nations impacts improvement. In this way, it is essential to determine at the beginning phase to have open and economical solutions. As a future scope, other grouping strategies in machine learning like decision trees, Naïve Bayes classifier, reinforcement learning might be utilized for infection discovery in plants and help the farmers to make an automated harvesting decision-maker, with the help of the statistics collected form economical classification algorithms which can be accessed by mobile devices and collaborate it with other precision farming techniques which will ultimately contribute towards making a sound harvesting decision or suggested most beneficial remedies for the infected plants covering more diseases for various crops.

REFERENCES

[1] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, *Big data analytics: Computational intelligence techniques and application areas*, Technol. Forecast. Soc. Change, vol. 153, p. 119253, Apr. 2020, doi: 10.1016/j.techfore.2018.03.024.

[2] Dolli Manoj, Divya K.S, *A study on present indian agriculture: Status, Importance, and Role in Indian Economy*, Accessed: Sep. 09, 2020. [Online]. Available: http://www.indianjournals.com/ijor.aspx? target=ijor:zijmr&volume=10&issue=3&article=003

[3] A. Kamilaris, A. Kartakoullis, and F. X. Prenafeta-Boldú, *A review on the practice of big data analysis in agriculture*, Comput. Electron. Agric., vol. 143, pp. 23–37, Dec. 2017, doi: 10.1016/j.compag.2017.09.037.

[4] R. Frelat et al., *Drivers of household food availability in sub-Saharan Africa based on big data from small farms*,Proc. Natl. Acad. Sci., vol. 113, no. 2, pp. 458–463, Jan. 2016, doi: 10.1073/pnas.1518384112.

[5] K. E. Giller et al., LU-*Communicating complexity: Integrated assessment of trade-offs concerning soil fertility management within African farming systems to support innovation and development*, Agric. Syst., vol. 104, no. 2, pp. 191–203, Feb. 2011, doi: 10.1016/j.agsy.2010.07.002.

[6] K. L. Chadha, *INDIAN VITICULTURE SCENARIO*, in Acta Horticulturae, May 2008, no. 785, pp. 59–68, doi: 10.17660/ActaHortic.2008.785.3.

[7] P. Spachos and S. Gregori, *Integration of Wireless Sensor Networks and Smart UAVs for Precision Viticulture*, IEEE Internet Comput., vol. 23, no. 3, pp. 8–16, May 2019, doi: 10.1109/MIC.2018.2890234.

[8] M. Soto, E. Falqué, and H. Domínguez, *Relevance of Natural Phenolics from Grape and Derivative Products in the Formulation of Cosmetics*, Cosmetics, vol. 2, no. 3, pp. 259–276, Aug. 2015, doi: 10.3390/cosmetics2030259.

[9] S. Suthaharan, *Big Data Classification: Problems and Challenges in Network Intrusion Prediction with Machine Learning*,SIGMETRICS Perform Eval Rev, vol. 41, no. 4, pp. 70–73, Apr. 2014, doi: 10.1145/2627534.2627557.

[10] N. Krithika and A. G. Selvarani, *An individual grape leaf disease identification using leaf skeletons and KNN classification*, in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, Mar. 2017, pp. 1–5, doi: 10.1109/ICIIECS.2017.8275951.

[11] SmartVineyardTM , *SmartVineyardTM Precision Viticulture System to monitor grape diseases.*, http://smartvineyard.com/home/ (accessed Sep. 08, 2020)

[12] P. Shvaiko, J. Euzenat, E. Jiménez-Ruiz, M. Cheatham, and O. Hassanzadeh, *Proceedings of The Tenth International Workshop on Ontology Matching (OM-2015)*, p. 249.

[13] C. Kempenaar, C. Lokhorst, E.J.B. Bleumer, R.F. Veerkamp, Th. Been, F.K. van Evert, M.J. Boogaardt, L. Ge, J. Wolfert, C.N. Verdouw, Michael van Bekkum, L. Feldbrugge, Jack P.C. Verhoosel, B.D. Waaij, M. van Persie, H. Noorbergen, *Big data analysis for smart farming: Results of TO2 project in theme food security,* Wageningen Plant Research report 65, 2016

[14] . Majumdar, S. Naraseeyappa, and S. Ankalaki, *Analysis of agriculture data using data mining techniques: application of big data*,. Big Data, vol. 4, no. 1, p. 20, Dec. 2017, doi: 10.1186/s40537-017-0077-4.

[15] S. Rajeswari, K. Suthendran, and K. Rajakumar, *A Smart Agricultural Model by Integrating IoT, Mobile and Cloud-based Big Data Analytics,*, 2017 International Conference on Intelligent Computing and Control (I2C2), 2017, doi: 10.1109/I2C2.2017.8321902

[16] R. H. L. Ip, L.-M. Ang, K. P. Seng, J. C. Broster, and J. E. Pratley, *Big data and machine learning for crop protection* ,Comput. Electron. Agric., vol. 151, pp. 376–383, Aug. 2018, doi: 10.1016/j.compag.2018.06.008.

[17] S. Sharma, G. Rathee, H. Saini, *Big Data Analytics for Crop Prediction Mode Using Optimization Technique*, 2018.

[18] Y. Ge, *Prediction of corn price fluctuation based on multiple linear regression analysis model under big data*, Neural Comput. Appl., p. 13, 2020

[19] *[No title found]*,presented at the 2015 International Conference on Computing, Communication, Control and Automation (ICCUBEA), Pune, Feb. 2015.

[20] P. B. Padol and A. A. Yadav, *SVM classifier based grape leaf disease detection*, in 2016 Conference on Advances in Signal Processing (CASP), Pune, India, Jun. 2016, pp. 175–179, doi: 10.1109/CASP.2016.7746160.

[21] H. Waghmare, R. Kokare, and Y. Dandawate, *Detection and classification of diseases of Grape plant using opposite colour Local Binary Pattern feature and machine learning for automated Decision Support System*, in 2016 3rd International Conference on Signal Processing and Integrated Networks (SPIN), Noida, Delhi NCR, India, Feb. 2016, pp. 513–518, doi:

10.1109/SPIN.2016.7566749.

[22] R. Patil, S. Udgave, S. More, D. Nemishte, and M. Kasture, *Grape Leaf Disease Detection Using K-means Clustering Algorithm*, vol. 03, no. 04, p. 4.

[23] N. Agrawal, J. Singhai, and D. K. Agarwal, *Grape leaf disease detection and classification using multi-class support vector machine*,in 2017 International Conference on Recent Innovations in Signal processing and Embedded Systems (RISE), Bhopal, Oct. 2017, pp. 238–244, doi: 10.1109/RISE.2017.8378160.

[24] U. Knauer, A. Matros, T. Petrovic, T. Zanker, E. S. Scott, and U. Seiffert, *Improved classification accuracy of powdery mildew infection levels of wine grapes by spatial-spectral analysis of hyperspectral images*, Plant Methods, vol. 13, no. 1, p. 47, Dec. 2017, doi: 10.1186/s13007-017-0198-y.

[25] N. Krithika and A. G. Selvarani, *An individual grape leaf disease identification using leaf skeletons and KNN classification*, in 2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS), Coimbatore, Mar. 2017, pp. 1–5, doi: 10.1109/ICIIECS.2017.8275951.

[26] K. R. Aravind, P. Raja, R. Aniirudh, K. V. Mukesh, R. Ashiwin, and G. Vikas, *Grape Crop Disease Classification Using Transfer Learning Approach*, in Proceedings of the International Conference on ISMAC in Computational Vision and Bio-Engineering 2018 (ISMAC-CVB), vol. 30, D. Pandian, X. Fernando, Z. Baig, and F. Shi, Eds. Cham: Springer International Publishing, 2019, pp. 1623–1633.

[27] A. Adeel et al., *Diagnosis and recognition of grape leaf diseases: An automated system based on a novel saliency approach and canonical correlation analysis based multiple features fusion*,Sustain. Comput. Inform. Syst., vol. 24, p. 100349, Dec. 2019, doi: 10.1016/j.suscom.2019.08.002.

[28] U. Shruthi, V. Nagaveni, and B. K. Raghavendra, *A Review on Machine Learning Classification Techniques for Plant Disease Detection*, in 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, Mar. 2019, pp. 281–284, doi: 10.1109/ICACCS.2019.8728415.

[29] J. Zhu, A. Wu, X. Wang, and H. Zhang, *Identification of grape diseases using image analysis and BP neural networks*, Multimed. Tools Appl., vol. 79, no. 21–22, pp. 14539–14551, Jun. 2020, doi: 10.1007/s11042-018-7092-0.

[30] A. D. Andrushia and A. T. Patricia, *Artificial bee colony optimization (ABC) for grape leaves disease detection*,Evol. Syst., vol. 11, no. 1, pp. 105–117, Mar. 2020, doi: 10.1007/s12530-019-09289-2.

[31] Z. Tang, J. Yang, Z. Li, and F. Qi, *Grape disease image classification based on lightweight convolution neural networks and channelwise attention*,Comput. Electron. Agric., vol. 178, p. 105735, Nov. 2020, doi: 10.1016/j.compag.2020.105735.

[32] M. Ji, L. Zhang, and Q. Wu, *Automatic grape leaf diseases identification via UnitedModel based on multiple convolutional neural networks*, Inf. Process. Agric., vol. 7, no. 3, pp. 418–426, Sep. 2020, doi: 10.1016/j.inpa.2019.10.003.

[33] W. Lumchanow and S. Udomsiri, *Combination of GLCM and KNN Classification for Chicken Embryo Development Recognition*,in 2019 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT-NCON), Nan, Thailand, Jan. 2019, pp. 322–325, doi: 10.1109/ECTI-NCON.2019.8692272.

[34] *Managing Black Rot | Viticulture and Enology*, https://grapesandwine.cals.cornell.edu/newsletters/appellation-cornell/2014-newsletters/issue-17/managing-black-rot/ (accessed Sep. 12, 2020).

[35] *Defoliation of Grape Leaves Associated with Downy Mildew, Anthracnose, and Isariopsis Leaf Spot*, American Society for Enology and Viticulture. https://www.asev.org/abstract/defoliation-grape-leaves-associated-downy-mildew-anthracnose-and-isariopsis-leaf-spot (accessed Sep. 12, 2020).

[36] *Grapevine Disease Images*, https://kaggle.com/piyushmishra1999/plantvillage-grape (accessed Sep. 12, 2020).

Edited by: Dana Petcu
Received: Feb 19, 2021
Accepted: Sep 30, 2021