Principal Component & Linear Discriminant Analysis for Apple Sorting

Mrs. Manisha V. Bhanuse[1], Dr. Mrs. S. B. Patil[2]

[1]*Assistant Professor, D. Y. Patil College of Engg. & Tech. Kolhapur, India*
[2]*Incharge Principal, Dr. J.J. Magdum College of Engg. Jaysingpur, India*

ABSTRACT

Spectroscopic imaging is non-destructive, non-invasive & chemical free technique which can be used to analyze wide range of biological materials. There are many varieties of apple fruits found in Indian market. To identify the variety of apples such as Fuji, Red Star and Gala based on their parameters various machine learning models can be used. In order to distinguish the variety more precisely a comparison of machine Learning based Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is proposed in this paper. The spectral database provided by CAPA Apple Quality Grading Multi-Spectral Image Database, ULG (Gembloux Agro-Bio Tech) – Belgium is used here. LDA & PCA are the dimensionality reduction techniques used in Machine Learning. It uses 'feature similarity' to predict the values of new datasets which further means that the new data point will be assigned a value based on how closely it matches the points in the training set. The developed model is evaluated in terms of the performance parameters.

Index Terms — PCA, LDA,

INTRODUCTION

Machine learning with big data technologies and other fast computing techniques is developing in various fields or applications. It can be defined as the scientific method that allows machines to learn without programming the devices. In this fast developing world, healthcare is one of the important concern. It is said that 'Eat apple a day keeps a doctor away'. It shows the importance of apples in fruit market. A variety of apples are produced in agriculture field. [2] Each variety may yield better income based on the quality, texture, shape etc. As per market survey, it is observed that higher grade apples generate larger revenues. Conventional methods of sorting or grading is human-dependent. Very often mechanical devices are also used to differentiate the variety based on dimensions and weight. More recently, machine learning algorithms along with some image processing techniques are emerged to improve the process of sorting. [1] It provides the substitute to the human eye, enabling to detect features which humans cannot detect. The machine vision based algorithms for grading and sorting are much more powerful than conventional methods. They have automatic learning capabilities and ensure a detection performance far beyond the speed and accuracy of any trained operator. This paper presents use PCA& LDA algorithm to detect the variety of apples.

HYPOTHESIS

All the organic materials including fruits continuously emit and absorb energy at the molecular level by lowering or boosting their molecular energy levels. Molecules' structure determines the wavelengths at which they absorb, reflect, and transmit electromagnetic radiation

Ultraviolet (UV), visible light (VIS), near-infrared (NIR), mid-infrared (MIDIR), and far-infrared (FIR) are examples of electromagnetic waves (FIR). [5] Each zone corresponds to a distinct type of atomic or molecule transition with various frequencies. Food tissues, like any biological material, are held together by a variety of molecular connections and pressures. O-H or C-H bonds are abundant in water, carbohydrates, and lipids. C-H or N-H bonds are abundant in organic molecules and petroleum derivatives.

Electromagnetic waves are transferred through a fruit sample when it is exposed to light. The stretching and bending vibrations of chemical bonds such as O-H, N-H, and C-H affect the energy of incident electromagnetic waves. As a result, spectroscopy can offer unique and detailed fingerprints of fruit samples using the changes in molecular energy levels that have been observed. The electromagnetic wave is viewed as light at the macro level, and the transitioning of the incidence.

The reflection, dispersion, and transmission of light are all examples of electromagnetic waves. Because of the absorption a portion of light penetrates sample tissue, the intensity and wavelengths of emission, and physical and chemical states of the objective material influence absorption or reflection. This property of organic materials to behave differently at different wavelengths of Electromagnetic radiations according to physical and chemical properties is of great importance in determination of fruit quality non-destructively.

<div align="center">

METHODOLOGY

</div>

Various machine learning models or algorithms can be used to identify the variety of apple based on their features. In order to achieve this Principal Component and Linear Discriminant analysis are used in this paper.

Principal Component Analysis:
PCA is widely used method developed for dimensionality reduction. Dimensionality reduction is the method of taking high dimensional space data and mapping it into a different space with much smaller dimensionality. This procedure is closely linked to the concept of lossy compression used in information theory. There are quite a few reasons to decrease the dimensionality of data. First, high dimensional data impose computational challenges. Mathematical analysis for PCA is as follows [9]. It begins with p-dimensional vectors, which are then projected down into a q-dimensional subspace to summarise. The projection of the original vectors on to q directions, the primary components that span the sub-space, will constitute our summary.

There are numerous mathematical methods for deriving the principal components that are equivalent in weight. The most straightforward method is to determine the projection that maximises variance. The first and most important factor is the direction in space where projections differ the most. The direction that maximises variation among all directions to the first is the second main component. The kth Component is orthogonal to the previous k-1 components and maximises variance. In all, there are p main components.

Instead of maximising variance, it may appear that finding the projection with the shortest average (mean-squared) distance between the original vectors and their projections on to the principal components is a better option; this turns out to be identical to maximising variance. Assume that the data has been centered, meaning that every variable has a mean of 0. If we encode the centered data in a matrix x, where rows represent objects and columns represent variables, then $x^T x = nv$, where v represents the data's covariance matrix.

Minimizing projection residuals:
Let's start with a single-dimensional projection. To put it another way, we have p-dimensional vectors that we want to project onto a line passing through the origin. The line can be specified by a unit vector along it, w, and the projection of a a data vector $\vec{x_i}$ onto the line is $\vec{x_i}.\vec{w}$, which is a scalar. When we project on to one of the coordinate axes, we get the correct answer. This represents the projection's distance from the origin; in p-dimensional space, the real co-ordinate is $(\vec{x_i}.\vec{w})\vec{w}$. Because the mean of the vectors $\vec{x_i}$ is zero, the projection's main will be zero.

$$\frac{1}{n}\sum_{i=1}^{n} (\vec{x_i}.\vec{w})\vec{w} = \left(\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right).\vec{w}\right)\vec{w} \quad (1)$$

There will be certain errors if we try to use our projected or picture vectors instead of our original vectors. In most cases, the images do not match the source vectors. We need to know when they will match? What is the size of the difference in the projection's error or residual? Let's imagine we have a single vector x_i it can be

$$\| \vec{x_i} - (\vec{w}.\vec{x_i})\vec{w}\|^2 = (\vec{x_i}(\vec{w}.\vec{x_i})\vec{w}).(\vec{x_i} - (\vec{w}.\vec{x_i})\vec{w})$$

$$= \vec{x_i}.\vec{x_i} - \vec{x_i}.(\vec{w}.\vec{x_i})\vec{w} - (\vec{w}.\vec{x_i})\vec{w}.\vec{x_i} + (\vec{w}.\vec{x_i})\vec{w}.(\vec{w}.\vec{x_i})\vec{w} \qquad = \|\vec{x_i}\|^2 - 2(\vec{w}.\vec{x_i})^2 + (\vec{w}.\vec{x_i})^2 \vec{w}.\vec{w}$$

$$= \vec{x_i}.\vec{x_i} - (\vec{w}.\vec{x_i})^2 \qquad\qquad (2)$$

Since $\vec{w}.\vec{w} = \|\vec{w}\|^2 = 1$.

Add these residuals up through all the vectors

$$MSE(\vec{w}) = \frac{1}{n}\sum_{i=1}^{n} \|\vec{x_i}\|^2 - (\vec{w}.\vec{x_i})^2 \quad (3)$$

$$\frac{1}{n}\left(\sum_{i=1}^{n} \|\vec{x_i}\|^2 - \sum_{i=1}^{n} (\vec{w}.\vec{x_i})^2\right) \qquad (4)$$

Because the initial summation is independent on w, it makes no difference when attempting to reduce the mean squared residual. To reduce the MSE, we must increase the size of the second, i.e., we wish to maximize

$$\frac{1}{n}\sum_{i=1}^{n} \quad (\vec{w}.\vec{x}_i\ )^2 \tag{5}$$

Which we can perceive is a sample mean of $(\vec{w}.\vec{x}_i\ )^2$. Here the sample mean of square is each time equal to the square of sample mean added with the sample variance:

$$\frac{1}{n}\sum_{i=1}^{n} \quad (\vec{w}.\vec{x}_i\ )^2 \left(\frac{1}{n}\sum_{i=1}^{n} \quad \vec{x}_i\ .\vec{w}\right)^2 + \widehat{V}\ [\vec{w}.\vec{x}_i\ ]$$

Minimizing the residual sum of squares is equivalent to maximising the variance of the projections, as the mean of the projections is zero.[9] We don't wish to project on to just one vector, but on to several principle components in general. If the components are orthogonal and have the same unit vectors $\vec{w}_1, \vec{w}_2, \ldots .. \vec{w}_k$, they are said to be orthogonal, then the image of x_i is nothing but its projection onto the space spanned by these vectors,

$$\sum_{j=1}^{k} \quad (\vec{x}_i\ .\vec{w}_j\ )\vec{w}_j$$

The projection's mean for each component is still zero. When we apply the same algebra to the mean squared error, the cross-terms between distinct components cancel out, leaving us with the task of maximising the sum of the variances of the projections on to the components. [9]

Maximizing variance:

Let's see how to maximize the variance. Producing all the summations finds somewhat tedious, so let's perform the algebra in matrix form. If we load our n data vectors into the $n \times p$ matrix x, then the projection is given by xw i.e. $n \times 1$ matrix. [9] Thus the variance is

$$\widehat{V}\ [\vec{w}.\vec{x}_i\ ] = \frac{1}{n}\sum_i \quad (\vec{x}_i\ .\vec{w})^2$$

$$= \frac{1}{n}(xw)^T(xw)$$
$$= \frac{1}{n}w^T x^T xw$$
$$= w^T \frac{x^T x}{n} w$$
$$= w^T vw$$

Let's select a unit vector $\vec{w}$ which can maximize $\widehat{V}\ [\vec{w}.\vec{x}_i\ ]$. To perform this, we need to ensure that we just look at unit vectors. It means we need to limit the maximization. The limit or constraint is that $\vec{w}.\vec{w} = 1$, and/or $w^T w = 1$. To apply this limit, let's introduce multiplier λ and perform a superior and constrained optimization:

$$L(w,\lambda) \equiv w^T vw - \lambda\ (w^T w - 1) \tag{6}$$

$$\frac{\partial L}{\partial \lambda} = w^T w - 1$$
$$\frac{\partial L}{\partial w} = 2vw - 2\lambda w$$

Setting the derivative values to zero at the optimum, it is found that

$$w^T w = 1$$
$$vw = \lambda w$$

As a result, the desired vector w is an eigenvector of the covariance matrix v, and the maximising vector is the one with the greatest eigenvalue. This is good news because finding eigenvectors is a rather quick process, and eigenvectors have a number of useful mathematical qualities that we may exploit as follows. Because v is a pp matrix, it will have a maximum of p distinct eigenvectors. We know that v is a covariance matrix, thus it is symmetric, and that the eigenvectors must be orthogonal to one another, according to linear algebra. Because v is a matrix, it is a non-negative definite matrix in the sense that $\vec{x}.v\vec{x} \geq 0$ for any value of $\vec{x}$. This means that all of v's eigenvalues must be ≥ 0.

The primary components of the data are the eigenvectors of v. They span the entire p-dimensional space because we know they are orthogonal. The direction along which the data have the maximum variation is represented by the first principle component, i.e. the eigenvector with the biggest value of λ. The path orthogonal to the first component with the most variance is the second principle component, or second eigenvector.]9]Their projections will be uncorrelated since it is orthogonal to the first eigenvector. In reality, projections on to all of the major components are unrelated to one another. Our weight matrix w will be a p×q matrix if we choose q principle components, with each column representing a different eigenvector of the covariance matrix v. The eigenvalues gives us the variance of the projection on to each component. The variance of projections on to the leading q principle components is then

$$given\ as \sum_{i=1}^{q} \lambda_{i.}$$

Linear Discriminant Analysis (LDA)
The LDA approach works by projecting the original data matrix into a lower-dimensional environment. Three stages were required to accomplish this goal. The between-class variance or between-class matrix is used to compute the separability between distinct classes i.e. the distance between the means of different classes. The within-class variance, also known as the within-class matrix, is calculated as the distance between the mean and the samples of each class. The final stage is to create a lower-dimensional space that optimizes between-class variance while minimizing within-class variance. All these stages are discussed below.

Calculating the between-class variance (S_B)
The between-class variance for the ith class (S_{Bi}) signifies the distance among the mean for the ith class (μ_i) and the total mean(μ). LDA method searches for the lower dimensional space, which is most widely used to improve the between-class variance i. e. it simply maximizes the separation distance of classes. Taking consideration of the following assumption, the between-class variance and the between-class matrix (S_B) are determined, The assumption is that, given the original dataset matrix $X = \{x_1, x_2, \dots x_N\}$, in which x_i denotes the ith sample, pattern or observation whereas N represents the entire number of samples. Every sample is denoted by M number of features$(x_i \in R^M)$. It means, every sample is denoted as the point in M – dimensional space. Let's consider, the data matrix is subdivided into $c = 3$ classes as $X = [\omega_1, \omega_2, \omega_3]$. Let every class has five number of samples (i.e. $n_1 = n_2 = n_3 = 5$), where n_i denotes the number of samples in ith class. The overall number of samples (N) can be then calculated as follows,

$$N = \sum_{i=1}^{3} n_{i.}$$

To compute the between-class variance(S_B), separation distance of various classes represented by $m_i - m$ will be determined as follows:

$$(m_i - m)^2 = (W^T \mu_i - W^T \mu)^2$$
$$= W^T(\mu_i - \mu)(\mu_i - \mu)^T W$$

Where m_i denotes the mean projection of ith class and is determined as follows,
$$m_i = W^T \mu_i,$$

Where m signifies the projection of total mean of each class and is determined as follows,
$$m = W^T \mu,$$

Where, W denotes transformation matrix of LDA and $\mu i(1 \times M)$ signifies mean of the ith class and is calculated as defined in Equation (7), $\mu(1 \times M)$ the total mean of altogether classes and is calculated as defined in Equation (8) .

$$\mu_j = \frac{1}{n_j} \sum_{x_i \in \omega_j} x_i \qquad (7) \qquad\qquad \mu =$$

$$\frac{1}{N} \sum_{i=1}^{N} x_i = \sum_{i=1}^{c} \frac{n_i}{N} \mu_i \qquad (8)$$

Where c denotes total number of classes as an example $c = 3$.

The term $(\mu_i - \mu)(\mu_i - \mu)^T$ of Equation (6) denotes the separation distance among the mean of ith class (μ_i) and total mean (μ), or simply it signifies the between-class variance for the ith class (S_{Bi}). Substituting (S_{Bi}) into Equation (6) gives us equation (9)

$$(m_i - m)^2 = W^T S_{Bi} W \qquad (9)$$

The overall between-class variance is determined as follows,

$$(S_B = \sum_{i=1}^{c} \ n_i \ S_{Bi}).$$

This is how the between-class matrix for the first class (S_{Bi}) can be calculated and then total between-class matrix (S_B) is computed by adding all these between-class matrices.

RESULTS

Evaluation of the PCA & LDA models is carried out on the basis of accuracy, R^2, Root Mean Square Error (RMSE).


Based on confusion matrix shown in figure,

The closeness of the proposed fruit identification method to the actual value is referred as accuracy and is expressed as

$$M_1 = \frac{X^{tn} + X^{tp}}{X^{fn} + X^{fp} + X^{tn} + X^{tp}}$$

Where, the true positive and negative is denoted as X^{tp} and X^{tn} False negative and positive is expressed as X^{fn} and X^{fp} respectively. Accuracy for PCA & LDA for various training percentage is observed as follows.
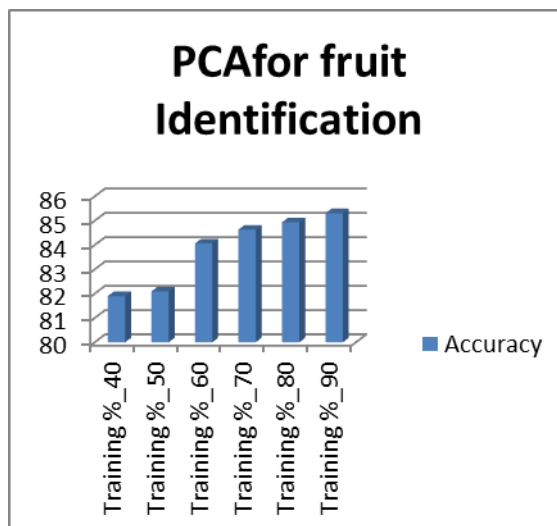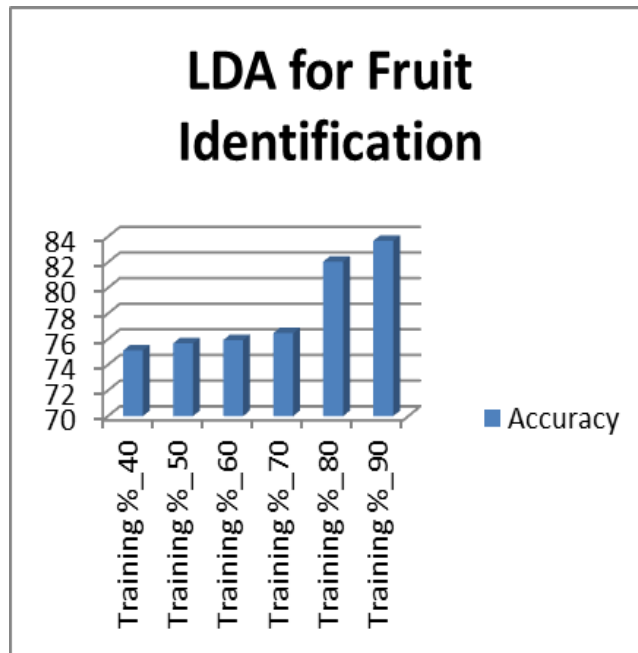
The future prediction is evaluated using the metric R2 concerning the variability proportion and is expressed as,

$$M_3 = 1 - \frac{\sum\left(x^c - x^a\right)^2}{\sum\left(x^c - x^{mean}\right)^2}$$

Where, x^c refers to the calculated value, and x^{mean} refers to the average value respectively. R-squared is a statistical quantity that signifies the goodness of fit for the model. R-square for PCA & LDA for various training percentage is observed as follows.
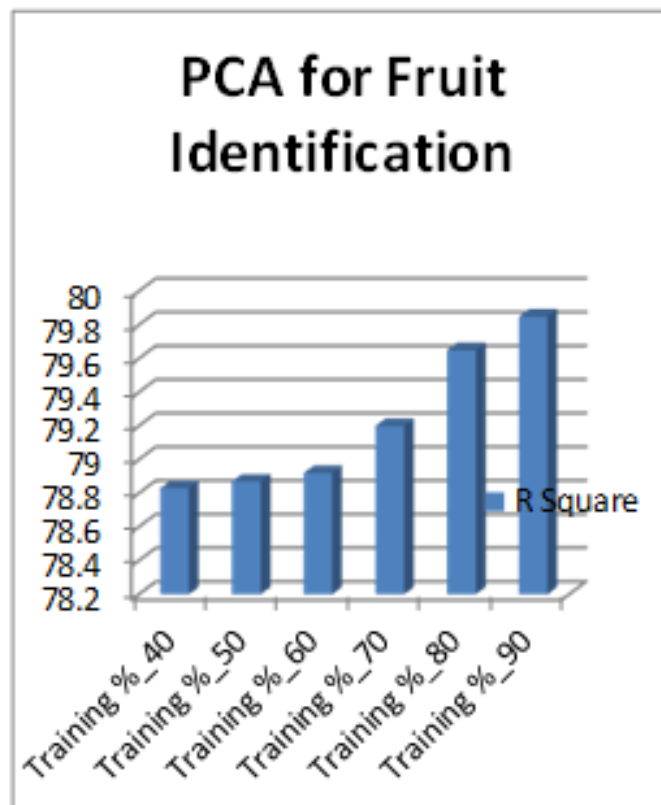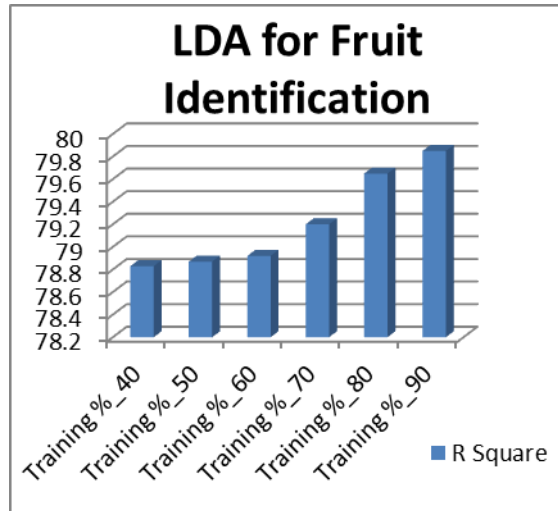
The most frequently used metric for ML evaluation is RMSE (root-mean-square error). "This is defined as the square root of the average squared distance between the actual score and the predicted score" And mathematically represented as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

Where yi denotes the true score if ith data point, whereas ŷi represents the predicted value. RMSE is a measure to custom when it is required to assess the standard deviation σ for observed value from model's prediction. RMSE for PCA & LDA for various training percentage is observed as follows.
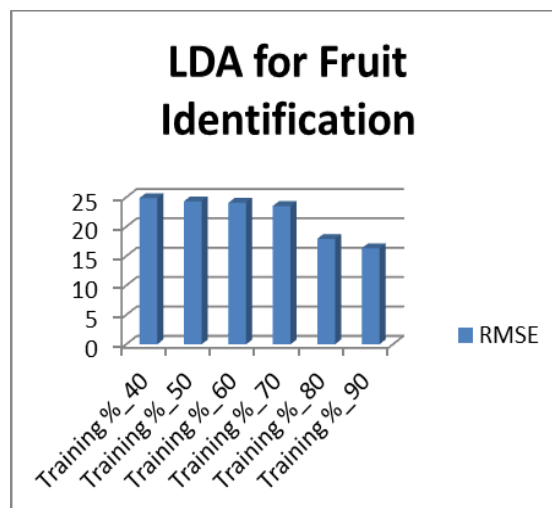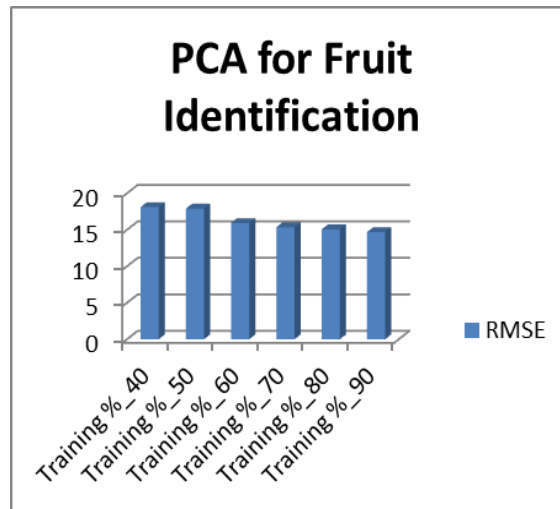
CONCLUSION

The primary distinction between the two algorithms is in the fundamental learning strategy. Both techniques depend on deconstructing matrices of eigenvalues and eigenvectors. LDA is supervised whereas PCA is unsupervised. By examining the association between various characteristics, PCA decreases dimensionality. In order to achieve this, a new subspace with orthogonal axes—or principal components—with the direction of maximum variance is created.

In essence, PCA creates components based on the direction in which the data has the most variance, for example, when the data is spread out the maximum. Principals and eigenvectors are two names for the same component, which is a subset of the data that contains the majority of the information or variance of our data. On the other hand, LDA accomplishes nearly the same thing, but it also includes a "pre-processing" step that determines mean vectors from class labels before extracting the eigenvalues.

REFERENCES

[1]   Qionglei LIU, Baohua TAN1, ― Nondestructive Determination of Fruit Based on Sensory and Near-Infrared Technology‖ Advances in Social Science, Education and Humanities Research, volume 87, ICEMEET 2016

[2]   Ebrahiema Arendse, ―Non-destructive measurement of pomegranate fruit quality‖, Stellenbosch University https://scholar.sun.ac.za

[3]   Parichat Theanjumpol, Pichet Noimanee, et al., ―Tangerine and Mango Fruits Quality Detection by Near Infrared Spectroscopy‖ CMU.J.Nat.Sci.Special issue on Agricultural & Natural Resources (2012) Vol.11 (1) 231

[4]   Nicácia P. Machado, José C. Fachinello, ―Pear quality characteristics by Vis / NIR spectroscopy‖, Annals of the Brazilian Academy of Sciences, (2012) 84(3): 853-863, ISSN 1678-2690

[5]   Yong He, Yun Zhang, et al., ―Nondestructive Determination of Tomato Fruit Quality Characteristics Using Vis/NIR Spectroscopy Technique‖, International Journal of Information Technology Vol. 11 No. 11 2005

[6]   Alexander Wendela, , James Underwooda, Kerry Walsh ―Maturity Estimation of Mangoes Using Hyperspectral Imaging from a Ground Based Mobile Platform‖, Computers and Electronics in Agriculture, Volume 155, December 2018, Pages 298-313

[7]   Xiao-Yan Chen, Xiang Long , et al. ―Apple Internal Quality Inspection Using hyperspectral Image Technology‖ Advances in Computer Science Research, volume 44 WCSN 2016

[8]   Cinmayii A. Garillos-Manliguez, John Y. Chiang, ―Multimodal Deep Learning and Visible-Light and Hyperspectral Imaging for Fruit Maturity Estimation‖, Sensors 2021, 21, 1288. https://doi.org/10.3390/s21041288

[9]   Cosma Rohilla Shalizi, ―Advanced Data Analysis from an Elementary Point of View‖, http://www.stat.cmu.edu/~cshalizi/ADAfaEPoV/

[10]   Alaa Tharwat, Tarek Gaber, ―Linear discriminant analysis: A detailed tutorial‖, 0921-7126/17/$35.00 © 2017 – IOS Press, AI Communications 00 (20xx) 1–22 DOI 10.3233/AIC-170729

AUTHORS PROFILE

Mrs. M. V. Bhanuse, Assistant Professor, D. Y. Patil College of Engg. & Tech. Kolhapur, India

Prof. Dr. S. B. Patil, Principal, Dr. J.J. Magdum College of Engg. Jaysingpur, India.