Apple Fruit Quality Identification using Clustering

*Rahul J. Mhaske , Siddharth B. Dabhade , Prapti Deshmukh Siddharth Arts, Commerce and Science College, Jafrabad, Jalna, MS, India 12,3 MGM's Dr. G.Y. Pathrikar College of CS & IT, Aurangabad, Maharashtra, India rahulmhaske007@gmail.com, 2dabhade.siddharth@gmail.com, 3prapti.research@gmail.com

Abstract- "An apples a day keeps doctor away" this proverb gives us important of apple in our healthy life. Apples fruit is consist of plenty of nutrition's therefore, doctors are always prefer to advice to eat the apple in most of the diseases. Hence, there is a huge demand of apples in market. To fulfill this demands suppliers need to provide the good quality fruit. There is a need of quality fruits in market. In this work studied various types of apples quality by using clustering approach. Comparative analysis is performed and given results are much better as compare to earlier work.

Keywords- fruit quality, size detecting, fruit grading, image processing, clustering.

I. Introduction

India is an agricultural nation. The production from the farm is at second position in the world. It is observed that near about 60% Indians are directly or indirectly depends on the business which are on agricultural products. It plays very important role in Indian economy. Farmers are mostly depended on the yield of agricultural products, therefore, there is need to focus on agriculture products which can be in good quality. To scale the economic situation of current poor or low land holder farmers, there is a need to export their products in demanding market in a good quality[1]. One of the major farm product is fruit. Fruits are very useful to our health. Physicians are always suggest the good quality fruit to the patients to recover from the diseases. Fruits contains different types of vitamins, proteins, etc. which can be helpful to our body. Hence there is a need to provide good quality fruit in the market. Main reason for eating a fruits is to maintain a good health and will make happiness. Quality of fruit also consider the storage, transportation, disease, etc. factors [2] [3][4][5]. Loss of fruit quality in developing country is approximately 20-40% before the harvesting and also during the transportation or storage some sort of damages are occurred.

II. IMPORTANCE OF QUALITY FRUIT

The term quality implies the degree of excellence of a product or its suitability for a particular use. Quality is a term which can show the suitability of product in effective manner. Human being can compare N number of parameters to decide the quality of any product or item. The N number of parameters may be appearance, aroma, texture, taste, nutritional, chemical importance, functionality, mechanical, etc. Judith A. Abbott (1999) has worked on "Quality measurement of fruits and vegetables" says that only humans can decide the quality, but instruments that identify the attributes which are decided for the quality measurement [6]. Shewfelt (1999) points out that quality is either depend on product or consumer requirement [7][8] [9].

III. FEATURE EXTRACTION TECHNIQUES

Features can be extracted from the image in various ways like statistical features, non statistical features, color, texture, etc.

A. Color Feature

Color image is consist of three important colors like Red, Green and Blue. These colors having different intensity levels. By using this color intensity as features it is processed and according to that results were taken by the various researcher. Zou Xiaobo [10] has classified the apple fruit into four grades like Reject, Class-I, Class-II, and Extra. They have developed organization feature parameter (OFP) method and compared their results with BP-ANN with SVM classification. The mechanism behind the OFP is Genetic Algorithm. Overall accuracy is 95% on the fruit surface color. Guo Dong [11] has segmented the image on the basis color with the help of hierarchical prototype learning (HPL). For prototype generation self-organizing map (SOM) and simulated annealing (SA) were used. Color space is even very useful in the identification of internal structure or defects in pickles using hyper spectral camera [12]. Most of the peoples were attracted towards the fruits because of colors, hence, color is very important factor while deciding the quality of fruits. Every time only the surface color of the fruit is evaluated for grading the fruit quality[13]. There are plenty of fruits are available in market which are red in color like apple, pomegranate, tomato, etc. Audrey Darrigues, et. al. [14] has developed a tool which can work on phenotype and genotype based classification of various fruits like Potato, Cucumber, Red plum, Muskmelon, Carrot, Strawberry. Even they have created the tomato analyzer to find the phenotype. In computer vision colors are represented in RGB and HSV color space.

RGB color space is used combination of red, green and blue in different ways which can reproduce the colors. Even on the display it is used various combination of intensity level of each color, the value of intensity color from 0 to 255. By formation of various additive color intensity it produces near about 16,777,216 various shades of colors. Most of the old displays were used 16 or 256 colors only with very few memory. HSV color space is advanced version it is known by hex-cone color. The working of HSV color space is similar to CMYK or RGB but there is a slightly difference. (H) Indicates the hue i.e. pure color. We have to consider the color wheel which can works upon the pure color value between the 0 to 1 and value is calculated the corresponding position of wheel in which the fractional values are considered for the representation of color for example red color refers 0 value whereas yellow color uses 1/6 value and green color uses 1/3 value and so on. Then (S) is a saturation for white color. Representation of pure red color saturation of white color value is 1. (V) is the representation of darkness of

38

the color. Black color is shows 0 value. For the formation of light color the wheel should be away from the black. Sometimes value of V is replaced by the brightness then this representation is called as HSB [15]. H.D. Cheng and other has shown the advantages of color based segmentation by using various color space like CIE spaces, HIS, YIQ, Munsell, Hybrid color space, Nrgb (Normalized rgb), etc. with different type of algorithms[16]. Guy Rosman has shown the color based image segmentation and processing with various approaches [17]. With the help of color of blueberry fruit identified the current growth stage of that fruit in a natural environment [18]. Some times watershed algorithms are used in combination with color for the segmentation of image for grading the quality of fruits [19]. While deciding the qualities of the fruit by various grading sometime ripeness of fruit is also considered for the Neural Network models are used for the recognition of the color [20][11]. Whenever we are processing the color by the manually by our necked eyes or through the system on sensors, it is divided into three different approach based color spacer i.e. hardware-orientated spaces, human-orientated spaces and instrumental space. Hardware-orientated space is defined on the properties of the hardware devices which used to reproduce the colors. Most of the popular color space are comes under this category such as RGB, YIQ, and CMYK. The human-orientated color spaces are based on hue and saturation. HSI, HSL, HSV and HSB are comes under this category. Here shade, tone tint type of perceptions are as per the ideas of artistic person. This type of representation of color is not sensitive to small differences of colors which works like a human vision hence it is not suitable during the processing of production of colors [21]. To get the more accurate fine detail information of the colors instead of normal RGB based camera multispectral and hyperspectral sensors were used [12][22].

B. Texture Features

Texture features are gives the outlier of the original scene. There are plenty of object recognition system were developed on the basis of texture based features. Jun Zhao has developed a system which can work on texture properties of fruits which are on the tree. They have recognized the fruit on the tree with the help of texture features [23]. S. Arivazhagan, et.al.[24] has shown the comparative study of various fruits fruit color and texture features gives 86%. Ravikant Sinha has identified the infection with the help of texture features on the basis of color [25]. Texture features are used to calculate the prediction of yield of apple fruit [26].

As per the objective or requirement texture feature evaluation categories into three classes (Szczesniak et al., 1963)—fundamental, empirical, and imitative[7]. For the selection of specific texture of a specific surface various kinds of statistical methods has been used (Gonzalez & Woods, 2008) [2][27] [28]. To find out the shapes like circle or other types of shapes in texture Hough transformation or Gabor, Tamura texture features extraction techniques are used [29]. For automatic quality detection texture plays a vital role [30] [31][32]. To detect the quality of fruit through the computer vision very first criteria is identification of texture [33]. Recent development in the domain of fruit quality determination survey indicates the authenticity for the use of texture features [34] [35].

IV. CLASSIFICATION TECHNIQUES

There are various approaches can be used to classify the object. Statistical approach like distance measurement, Template matching, Neural Network, etc. According to the use of classification techniques can be categorized into three main broad categories supervises, unsupervised and the combination of the both approaches. Before actual classification we have to extract the features from feature extraction techniques like partial least squares regression (PLS or PLSR), multivariate linear regression (MLR), principal component regression (PCR), principal component analysis (PCA), artificial neural networks (ANN) [36] . There is a need to develop a strong feature extraction technique and classification techniques which can classify the data more accurately. Kmeans, Support Vector Machine, Back Propagation, RGB to other color conversions, template matching, watershed, Euclidian distance or other distance measurement techniques. etc. are highlighted by the Sandesh Raut and Amit Fulsunge [37]. Multi-Layer Perceptron Neural Networks[2], Fuzzy and neural network can be a good approach for segmentation and analysis of quality identification [38]. Fruit grading can be decide by the classification approach [13], on the basis of multiple features [39].

V. EXPERIMENTAL WORK

A. Proposed Appraoch

There are multiple ways to cluster the data but K-Means algorithm is the most used algorithm. Which tries to improve the inter group similarity while keeping the groups as far as possible from each other.

Basically K-Means runs on distance calculations, which uses "Euclidean Distance" for this purpose. Euclidean distance calculates the distance between two given points using the following formula:

Euclidean Distance =
$$\sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

K-Means is an iterative process of clustering; which keeps iterating until it reaches the best solution or clusters in problem space.

Here some of the basic steps in K-Means clustering which is generally used to cluster the data. Start with number of clusters we want e.g., 3 in this case.

Step 1: K-Means algorithm start the process with random centers in data, and then tries to attach the nearest points to these centers

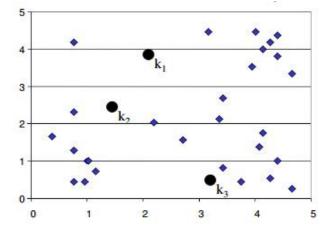


Fig. 1. Random Centers in Data

Step 2: Algorithm then moves the randomly allocated centers to the means of created groups

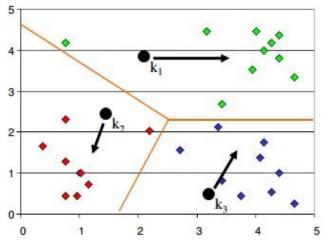


Fig. 2. Create Groups

Step 3: In the next step, data points are again reassigned to these newly created centers

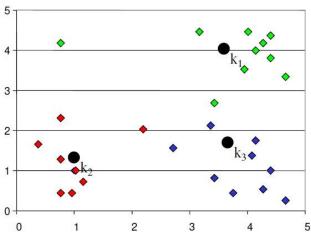


Fig. 3. Newly Created Centers

Step 4: Steps 2 & 3 are repeated until no member changes their association/ groups.

B. About our database

Local database is collected from the local market fruit suppliers and images were captured through the smart mobile phone i.e. Samsung Note 4. Samples own database are shown in Fig.4.



Fig. 4. Own Fruit Database Samples

Own Fruit database consist of 220 images i.e. apple and pomegrante collected from local market in a camera. Apple images are 59 with 2322 * 4128. There is variation in image size some of the images having the size of 647*483.

C. Computers and Optics in Food Inspection (COFILAB) database[40][41]:

For the experiments data set are taken from the Computers and Optics in Food Inspection (COFILAB). COFILAB is located in different laboratories located at Universidade de La Rioja (ULR), Universitat de València (UV), Universitat Politècnica de València (UPV) and the Valencian Institute for Agrarian Research (IVIA) and located the Valencian Institute of Agrarian Research (IVIA), Spain. The image acquisition system was composed of a digital camera (Canon EOS 550D) used to acquire high quality images with a size of 2592 x 1728 pixels and a resolution of 0.06 mm/pixel. However, the images for download in the database have been reduced to a size of 1200 x 800 pixels to reduce storage requirements.

Apple 'Royal Gala':

Database of 96 calibrated images of apples 'Royal' (24 apples photographed by four sides).

Apple 'Golden':

Database of 100 calibrated images of apples 'Golden' (25 apples photographed by four sides).

Apple 'Royal':

Database of 98 multispectral images of apples 'Royal Gala' (24 apples photographed by four sides in visible and near infrared).



Fig. 5. Apple 'Royal Gala'



Fig. 6. Apple 'Golden'

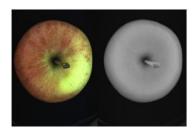


Fig. 7. Apple 'Royal'

Apple 'Golden':

Database of 100 multispectral images of apples Golden (25 apples photographed by four sides in visible and near infrared).

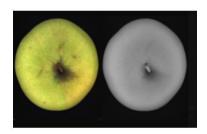


Fig. 8. Apple 'Golden'

VI. RESULTS

For calculation of results 14 different features were selected like as shown in Table 2, Table 3 and database with their details are given in Table 1.

TABLE I. FRUIT DATABASE DETAILS

Fruit Name	Total Images	Image Size
Apple	41	964*1296
Golden Apple	67	800*1200
Royal Apple	79	800*1200
Royal Multispectral	40	964*1296
Local Apple	59	2322*4128/647*483

TABLE II. SAMPLES OF FEATURES WITH RESULTS



TABLE III. SAMPLES OF FEATURES WITH RESULTS



Fruit quality identified by using K-means clustering. Before apply the K-menas cluster 14 features like Contrast, Correlation, Energy, Entropy, Homogeneity, IDM, Kurtosis, Mean, RMS, Skewness, Smoothness, Standard_Deviation, Variance, Affected Area, etc. After calculation of accuracy it is found that minimm 95.16% and maximum 98.38% on the both databases. Most of the research articles does not provide the exact accuracy of thier work but as compare to earlier reoprted work of other fruits are shown in Fig. 9.

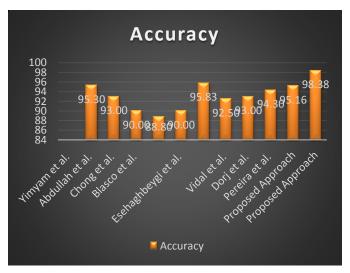


Fig. 9. Comparative results

TABLE IV. COMPARATIVE RESULTS WITH DETAILS [34] [42]

Year	Author (s)	Types Of Fruits &	Parameters	Color Space	Accurac y
		Vegetables			
2005	Yimyam et al.	Mango	Sorting by external quality	HSI	_
2006	Abdullah et al.	Carambola	Maturity discriminatio n	HSI	95.30%
2008	Chong et al.	-	Color evaluation	HSI	R2 = .93
2009	Blasco et al.	Pomegranat e	Grading by color	Gradin g by color	90.00%
2010	Liming and Yanchao	Strawberry	Grading by external quality	CIE Lab	88.80%
2010	Esehaghb eygi et al.	Peach	Color and size	HSI	90.00%
2012	Garrido- Novell et al.	-	Maturity discriminatio n	RGB	95.83%
2013	Vidal et al.	-	Color evaluation	RGB	R2 = .925
2017	Dorj et al.	Citrus	Grading by color	RGB	93.00%
2018	Pereira et al.	Papaya	Grading by color	RGB	94.30%
2019	Proposed Approach	Apple	Color based Segmentatio n	RGB	95.16%
2019	Proposed Approach	Apple	Color based Segmentatio n	RGB	98.38%

VII. CONCLUSION

In this research article fruit quality identified by using clustering technique. In the introduction need of fruit quality in huge market and current scenario of India has discussed. Importance of quality and apple fruit is highlited. Feature extraction approaches used by earlier reported work mentioned. Our own apple fruit database has developed and taken the results as compare to standard database. 98.38% is the highest accuracy given by the proposed approach. Comprative analysis has performed and it is observed that given approache gives better results.

41

REFERENCES

- [1] S. M. Wasule, "Quality Determination and Grading of Tomatoes using Raspberry Pi," *Int. J. Recent Innov. Trends Comput. Commun.*, vol. 6, no. 7, pp. 86–89, 2018.
- [2] F. Vesali, M. Gharibkhani, and M. H. Komarizadeh, "An approach to estimate moisture content of apple with image processing method," *Aust. J. Crop Sci.*, vol. 5, no. 2, pp. 111–115, 2011.
- [3] Y. Al Ohali, "Computer vision based date fruit grading system: Design and implementation," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 23, no. 1, pp. 29–36, 2011.
- [4] S. Cubero, N. Aleixos, E. Moltó, J. Gómez-Sanchis, and J. Blasco, "Advances in Machine Vision Applications for Automatic Inspection and Quality Evaluation of Fruits and Vegetables," *Food Bioprocess Technol.*, vol. 4, no. 4, pp. 487–504, 2011.
- [5] J. N. Quinton, Encyclopedia of Agrophysics, vol. 32, no. 3. Springer, 2011.
- [6] J. A. Abbott, "Quality measurement of fruits and vegetables," Postharvest Biol. Technol., vol. 15, pp. 207–225, 1999.
- [7] D. M. Barrett, J. C. Beaulieu, and R. Shewfelt, "Color, flavor, texture, and nutritional quality of fresh-cut fruits and vegetables: Desirable levels, instrumental and sensory measurement, and the effects of processing," Crit. Rev. Food Sci. Nutr., vol. 50, no. 5, pp. 369–389, 2010.
- [8] W. J. Florkowski, S. E. Prussia, and R. L. Shewfelt, *Integrated view of fruit & vegetable quality*. 2018.
- [9] R. L. Shewfelt and Bernhard Briickner, Fruit and Vegetable Quality An Integrated View. CRC Press, Taylor & Francis Group, 2000.
- [10] Z. Xiaobo, Z. Jiewen, and L. Yanxiao, "Apple color grading based on organization feature parameters," *Pattern Recognit. Lett.*, vol. 28, no. 15, pp. 2046–2053, 2007.
- [11] G. Dong and M. Xie, "Color clustering and learning for image segmentation based on neural networks," *IEEE Trans. Neural Networks*, vol. 16, no. 4, pp. 925–936, 2005.
 [12] D. P. Ariana and R. Lu, "Evaluation of internal defect and surface
- [12] D. P. Ariana and R. Lu, "Evaluation of internal defect and surface color of whole pickles using hyperspectral imaging," *J. Food Eng.*, vol. 96, no. 4, pp. 583–590, 2010.
- [13] P. Gay, R. Berruto, and P. Piccarolo, "Fruit Color Assessment for Quality Grading Purposes," 2002 ASAE Annu. Int. Meet. CIGR XVth World Congr., vol. 0300, no. 02, pp. 1–9, 2002.
- [14] N. D. and S. G. Audrey Darrigues, Jack Hall, Esther van der Knaap, and David M. Francis, "Tomato Analyzer-color Test: A New Tool for Efficient Digital Phenotyping," *J. Am. Soc. Hortic. Sci.*, vol. 133, no. 4, pp. 579–586, 2008.
- [15] S. V Phakade, D. Flora, H. Malashree, and J. Rashmi, "Automatic Fruit Defect Detection Using HSV and RGB Color Space Model," *Int.* J. Innov. Res. Comput. Sci. Technol., vol. 2, no. 3, pp. 67–73, 2014.
- [16] H. D. Cheng, X. H. Jiang, Y. Sun, and J. Wang, "Color image segmentation: Advances and prospects," *Pattern Recognit.*, vol. 34, no. 12, pp. 2259–2281, 2001.
- [17] G. Rosman, X. C. Tai, L. Dascal, and R. Kimmel, "Polyakov action minimization for efficient color image processing," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 6554 LNCS, no. PART 2, pp. 50–61, 2012.
- [18] H. Li, W. S. Lee, and K. Wang, "Identifying blueberry fruit of different growth stages using natural outdoor color images," *Comput. Electron. Agric.*, vol. 106, pp. 91–101, 2014.
- [19] P. Deepa and S. N. Geethalakshmi, "A Comparative Analysis of Watershed and Color based segmentation for Fruit Grading," *Int. J. Soft Comput. Eng.*, vol. 2, no. 3, pp. 11–15, 2012.
- [20] R. P. Krishnan, S. Sofiah, and M. Radzi, "Color Recognition Algorithm using a Neural Network Model in Determining the Ripeness of a Banana," System, no. October, pp. 11–13, 2009.
- [21] W. Castro, J. Oblitas, M. De-La-Torre, C. Cotrina, K. Bazan, and H. Avila-George, "Classification of Cape Gooseberry Fruit According to its Level of Ripeness Using Machine Learning Techniques and

- Different Color Spaces," IEEE Access, vol. 7, pp. 27389-27400, 2019.
- [22] J. Li et al., "Multispectral detection of skin defects of bi-colored peaches based on vis-NIR hyperspectral imaging," Postharvest Biol. Technol., vol. 112, no. October, pp. 121–133, 2016.
- [23] J. Tow and J. Katupitiya, "On-tree fruit recognition using texture properties and color data," 2005 IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS, pp. 3993–3998, 2005.
- [24] L. G. S.Arivazhagan, R.Newlin Shebiah, S.Selva Nidhyanandhan, "Fruit Recognition using Color and Texture Features," J. Emerg. Trends Comput. Inf. Sci., vol. 1, no. 2, pp. 90–94, 2010.
- [25] P. P. Ravikant Sinha, "Infection Analysis Using Colour Feature Texture Using Image Processing," *Int. J. Eng. Res. Appl.*, vol. 2, no. 4, pp. 1861–1866, 2012.
- [26] D. Stajnko, J. Rakun, and M. Blanke, "Modelling apple fruit yield using image analysis for fruit colour, shape and texture," Eur. J. Hortic. Sci., vol. 74, no. 6, pp. 260–267, 2009.
- [27] J. Ortega-Garcia et al., "MCYT baseline corpus: a bimodal biometric database," *IEE Proc. - Vision, Image, Signal Process.*, vol. 150, no. 6, p. 395, 2003.
- [28] R. Gonzalez and R. Woods, Digital Image Processing. 2008.
- [29] S. Sengupta and W. Lee, "Identification and Determination of the Number of Green Citrus Fruit under Different Ambient Light Conditions," *Int. Conf. Agric.*, 2012.
- [30] M. R. Satpute and S. M. Jagdale, "Automatic fruit quality inspection system," *Proc. Int. Conf. Inven. Comput. Technol. ICICT* 2016, vol. 1, 2017.
- [31] G. Capizzi, G. Lo Sciuto, C. Napoli, E. Tramontana, and M. Woźniak, "Automatic Classification of Fruit Defects based on Co-Occurrence Matrix and Neural Networks," vol. 5, pp. 861–867, 2015.
- [32] F. López-García, G. Andreu-García, J. Blasco, N. Aleixos, and J. M. Valiente, "Automatic detection of skin defects in citrus fruits using a multivariate image analysis approach," *Comput. Electron. Agric.*, vol. 71, no. 2, pp. 189–197, 2010.
- [33] A. Askar and H. Treptow, *Quality Assurance in Tropical Fruit Processing*. Springer Laboratory, Springer-Verlag, 1993.
- [34] C. Zheng, D. W. Sun, and L. Zheng, "Recent developments and applications of image features for food quality evaluation and inspection - a review," *Trends Food Sci. Technol.*, vol. 17, no. 12, pp. 642–655, 2006.
- [35] S. V Phakade et al., "Quality measurement of fruits and vegetables," J. Food Eng., vol. 5, no. 2, pp. 1–5, 2017.
- [36] L. S. Magwaza, U. L. Opara, H. Nieuwoudt, P. J. R. Cronje, W. Saeys, and B. Nicolaï, "NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit-A Review," Food Bioprocess Technol., vol. 5, no. 2, pp. 425–444, 2012.
- [37] S. Raut and K. Ingole, "Review On Fruit Disease Detection Using Image Processing Techniques," *Int. Res. J. Eng. Technol.*, vol. 4, no. 4, pp. 22–24, 2017.
- [38] R. Swarnalakshmi and B. Kanchanadevi, "International Journal of Computer Science and Mobile Computing A Review on Fruit Grading Systems for Quality Inspection," *Int. J. Comput. Sci. Mob. Comput.*, vol. 3, no. 7, pp. 615–621, 2014.
- [39] H. N. Patel, "Fruit Detection using Improved Multiple Features based Algorithm," *Int. J. Comput. Appl.*, vol. 13, no. 2, pp. 1–5, 2011.
- [40] S. Cubero, M. P. Diago, J. Blasco, J. Tardáguila, B. Millán, and N. Aleixos, "A new method for pedicel/peduncle detection and size assessment of grapevine berries and other fruits by image analysis," *Biosyst. Eng.*, vol. 117, no. C, pp. 62–72, 2014.
- [41] A. Vidal, P. Talens, J. M. Prats-Montalbán, S. Cubero, F. Albert, and J. Blasco, "In-Line Estimation of the Standard Colour Index of Citrus Fruits Using a Computer Vision System Developed For a Mobile Platform," *Food Bioprocess Technol.*, vol. 6, no. 12, pp. 3412–3419, 2013
- [42] A. Bhargava and A. Bansal, "Fruits and vegetables quality evaluation using computer vision: A review," J. King Saud Univ. - Comput. Inf. Sci., pp. 1–15, 2018.

www.asianssr.org 42