Implementation of Data Fusion in a Probabilistic Neural Network (PNN) for Classification of Fruit Ripening Stages

Jorge Isidro Aranda[1] and Arturo Baltazar [2]

[1]Facultad de Ciencias Físico-Matemáticas
Universidad Michoacana de San Nicolás de Hidalgo (UMSNH)
Morelia, Michoacán, México
jorge_isidro@yahoo.com

[2]Centro de Investigación y Estudios Avanzados
CINVESTAV-Saltillo
Robotics and Advanced Manufacturing Program
Km. 13.5 Ramos Arizpe, Coahuila, 25900, México
arturo.baltazar@cinvestav.edu.mx

Abstract. Quality control of postharvest fruits is moving to substitute traditional sensory testing methods for more reliable quantitative methods. Ripening in fruits, such as tomatoes, is a complex phenomenon which affects chemical and physiological properties as function of time. Attempts to solve the problem of ripening classification have been focus on single sensors: however, there is not yet a complete solution. In this work, data fusion from different non-destructive sensors implemented through a probabilistic neural network is used to improve the quality control of ripening in tomato fruit. Two independent sensors, a novel non-destructive acoustic impact and a colorimeter technique were used. The effect of the probabilistic neural network parameters was explained by an analysis of Bayes error. The acceptance rate as function of the number of feature is analyzed. The results showed that the error classification rate was reduced from 60% to 10% using the proposed data fusion scheme.

1 Introduction

The goal in quality control of postharvest fruits has been to substitute traditional sensory testing methods for more reliable and quantitative methods by correlating quality properties with properties that can be quantitatively measured. Firmness has been used by several researchers to describe internal and superficial properties and it is considered an important quality characteristic of fresh postharvested tomato fruits, whole or sliced [1]. Color change is another important parameter affected by ripening in tomatoes. It is influenced by many factors, including: temperature, maturity stage, atmosphere and storage time. Tomato color is sometimes determined using sensorial

© A. Gelbukh, A. Kuri (Eds.)

Received 04/06/07

methods; however, instrumental methods using colorimeter provide quantitative and more effective ways to determine a color index. Ripening in fruits is a complex phenomenon which affects chemical and physiological properties as function of time with global and local variations than can be better described in probabilistic terms.
Measurements of resonance frequencies and their relation with elastic properties in fresh commodities with like-spherical or spheroidal shape were first described by Abbott et al. [1], and the formulation was later modified by Cooke [2]. Several techniques have been studied in the past to excite the resonance frequencies in fruits including force vibration [1] and acoustic impact tests (introduced by Yamamoto Iwamoto & Haginuma [3]). The relation between the resonance frequency and the elastic properties of fruit with spherical shape is given by

$$Sc = f^2 * m^{3/2}, \qquad (1)$$

where Sc is typically referred to as stiffness coefficient, f is the dominant frequency and m is the bulk mass of the fruit. The determination of the dominant frequency f can be obtained by the acoustic impact test which consists of impacting the fruit with a small spherical object and recording the resonance vibration signals to determine the frequency spectrum [4]. In like-spherical fruits such as tomatoes two main types of vibration modes, assuming free boundary conditions, can be found: spherical modes and torsional modes. Some modes such as torsional can be filtered out using microphones, which are only sensitive to spherical vibration modes. However, a large uncertainty in the stiffness coefficient measurements can be found [4]. Although under controlled laboratory conditions they have been consistent and controllable results there is still presence of uncontrollable sources of noise [4].
The integration of sensor data using data fusion can be either complementary or use to enhance the response of a single sensor. In the area of robotics sensor fusion has been thoroughly investigated and used to overcome limitations of individual sensors [5]. Several schemes have been proposed to effectively develop and assess a methodology to integrate data [6]. The use of data fusion in robotics spans from improving position location and distance assessment to pattern recognition combining information from different sensors [5]. Artificial neural networks can be used for this fusion, and statistical models with probabilities help to integrate data at the evidential level, this level makes use of Bayesian approach theory.
Attempts to solve the problem of ripening classification have been mostly focused on single sensors; however, there is not yet a complete solution. Recently, application of artificial intelligence has been studied for sorting of fruits (i.e. [7] and [8]). Sensor fusion methodology has been applied using different non-destructive techniques including image analysis, near-infrared spectrophotometer, and colorimeter, among others. In this work, data fusion from different non-destructive sensors is used to improve the quality control of ripening in tomato fruit. The importance of data fusion of different sensors in foods such as melons and peaches to asses quality [8]. Sensor characteristics are a limiting factor in the performance of a data fusion system and good sensors are needed to obtain reliable results.
In this work, data integration and a fusion scheme is used to gather information of a proposed sensory system to produce a more complete solution to the ripening classification problem. Two independent sensors, a novel non-destructive acoustic impact [4] and a colorimeter technique [9] provided three different features: firmness,

color and luminosity. A PNN is used for data fusion and results are analyzed by a Bayesian classifier.

2 Probabilistic Neural Networks

Probabilistic Neural Networks (PNN) substitute the sigmoidal activation function commonly found in neural networks by an exponential function. Therefore, PNN results in a neural network that could compute nonlinear decision boundaries and approach the Bayes optimal [10]. PNN learns from a training data set, which is a collection of examples previously identified, belonging to specified classes. This training data are vectors with a determined number of components or elements each of them corresponding to the same number of features, each feature is obtained from a determined sensor.

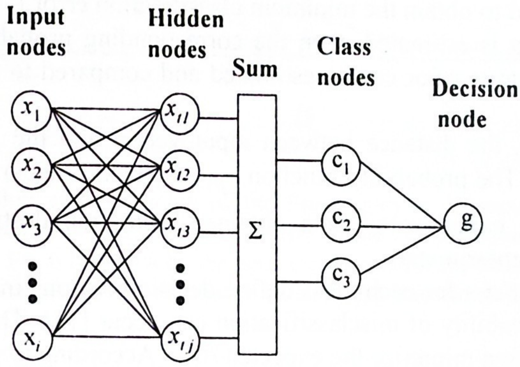


Figure 1. Probabilistic Neural Network

As described in Figure 1, a typical arrangement for a PPN is composed of two layers. The first layer contains input nodes where each receives an input value from the corresponding element in the input vector x and produces vectors indicating how close the input vector x is to the known pattern. This input vector data, labeled x, is a vector with the same number of elements or features as the training vector data. The second layer produces a vector of probabilities which is the sum for each class contribution in the first layer. Finally a hidden node with an activation function is given by

$$g = \exp(-\frac{D_i^2}{\sigma_i}), \tag{2}$$

where D_i is the distance between input vector and the i^{th} is the training data. This function g is a Gaussian function which returns a value of 1 if the input vector x and training data are equal, and drops to an insignificant value as the distance D_i increases.

Each hidden node is connected to a single class node C_i; if the output class of the input vector x is j, then such node is connected to the j^{th} class. Each *j-th* class node computes the sum of the activations (using Eq. 2) of the hidden nodes that are connected to it (i.e. all the hidden nodes for a particular class) and passes this sum to a decisión node, which outputs the class with the highest summed activation. The output is the class that x seems most likely to belong to.

It is important to note that we are assuming that σ_i has the same value for each data, $\sigma_i = \sigma$. The sum of the activations for data belonging to a certain class gives rise to a probabilistic function (linear normalizad combination of the Gaussians centered in each of the data in that class) and computes the probability of the input vector belonging to that class. The probabilistic function is written as [10]

$$p(x|w_i) = \frac{1}{\left(2\pi\sigma^2\right)^{3/2} M_i} \sum_{j=1}^{M_i} \exp\left[\frac{-\left(x-x_{ij}\right)^T\left(x-x_{ij}\right)}{2\sigma^2}\right], \tag{3}$$

where i (=1, 2, 3) is the corresponding class number, j (=1,..., M_i) is the pattern number, x_{ij} is the j[th] training vector from class i, x is the test vector, M_i is the number of training vector in class i, σ is a smoothing factor, which is a variable that needs to be optimized to obtain the minimum classification error.

Once the optimum σ is estimated, then the corresponding probability functions for each class and the Bayes error can be estimated and compared to error classification obtained in the PNN.

In the present work, the distance between input vector and the i^{th} training data is $D_i = \left(x - x_{ij}\right)^T\left(x - x_{ij}\right)$. The probability function $p(x|w_i)$ (using Eq. 3) for each class was estimated using the experimental data obtained from non destructive techniques acoustic impact and colorimetry .

The probability functions for each class define decision regions in which boundaries with the highest probability of misclassification can occur [11]. The Bayes strategies for pattern classification minimize the expected risk. According to the Bayes decision rule to assign a pattern x to a class w_i, with minium error, we have that

$$p(w_i)p(x|w_i) > p(w_k)p(x|w_k) \quad k,i = 1,\ldots,n; k \neq i, \tag{4}$$

where $p(w_1)$,..., $p(w_n)$,are known prior probabilities [11]. The average probability of error for the case of three classes is given in terms of decision regions in which vector x falls

$$p(error) = \int_{R_1} p(w_1)p(x|w_1)dx + \int_{R_2} p(w_2)p(x|w_2)dx + \int_{R_3} p(w_3)p(x|w_3)dx \tag{5}$$

which is known as Bayes risk [12] and x satisfies the following equation $p(w_i)p(x|w_i) > \max\limits_{\substack{1\le k\le 3\\ k\neq i}}\{p(w_k)p(x|w_k)\}$ with this the Bayes risk becomes the Bayes error, considering Eq. 4. The exact calculation of Eq. 5 for multi-class and multivariate problems is quite difficult, even assuming normal distribution [11]. This is due to the discontinuous nature of the decision regions. If the intersection area between classes 1 and 3 vanishes, then the three-class problem can be approached as a two-class problem and only the first two terms in Eq. 5 are needed [13].

However, Eq. 5 can be approximated analytically to give an upper and lower bound on the error [11]. These can be estimated in terms of the Bhattacharyya bound, which is a particular case of Chernoff bounds as follows

$$p(error) \equiv \varepsilon \leq \sqrt{p(w_1)p(w_2)} \int \sqrt{p(x|w_1)p(x|w_2)}\ dx. \tag{6}$$

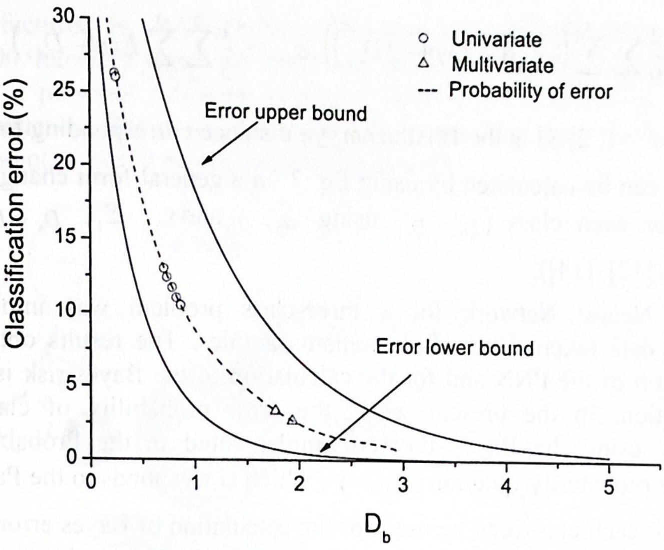


Figure 2. Classification error in terms of the Bhattacharyya distance. Solid lines are the theoretical classification error bounds. The estimate of the classification error for data in the Univariate case using Eq. 6 is shown by the open circles and for the Multivariate case by open triangles. The theoretical probability of error as given by Eq. 8 is shown by the dashed line

However, Eq. 5 can be approximated analytically to give an upper and lower bound on the error ([14]). Chulhee and Choi [14] presented an approach to estimate the classification error ε in the case of two classes (in this case $p(w_1) = p(w_2) = 1/2$), which reduces to the probability error for low values of Bhattacharyya distance D_b [12], [13], [14], given by

$$D_b = \frac{1}{8}(\mu_2 - \mu_1)^t \left[\frac{\Sigma_1 + \Sigma_2}{2}\right]^{-1} (\mu_2 - \mu_1) + \frac{1}{2}\ln \frac{\left|\frac{\Sigma_1 + \Sigma_2}{2}\right|}{\sqrt{|\Sigma_1||\Sigma_2|}}, \tag{7}$$

where μ_l and Σ_l (l = 1,2) are the mean and covariance for each multivariate class. The approach uses a fitted polynomial curve obtained using different class statistics. The polynomial curve is described as [14]

$$E = 40.219 - (70.019)\ D_b + (63.578) D_b^2 - (32.766)\ D_b^3 + (8.7172) D_b^4 - (0.91875) D_b^5, \tag{8}$$

this equation is valid only for $D_b \leq 3$ (see Figure 2) and shows t[illegible]n the number of features increases, the classification error diminishes significa[illegible]

For a three-class problem, Garber and Djouadi [12] introduced a technique to estimate bounds for the classification error (ε); such bounds are expressed as a function of pairwise classification errors which for a three-class problem is evaluated by adding the probability error resulting in three combinations of pairs of classes. The classification error for deciding among the original three classes ε is bounded from above and below as ([12])

$$\varepsilon_{min} \geq \frac{2}{9}\sum_{j=1}^{2}\sum_{i=j+1}^{3}\left(1-\sqrt{1-Exp\left(-2D_{i,j}\right)}\right), \varepsilon_{max} \leq \frac{1}{3}\sum_{j=1}^{2}\sum_{i=j+1}^{3}Exp\left(-D_{i,j}\right), \quad (9)$$

where $D_{i,j}$ $(i, j = 1, 2, 3)$ is the Bhattacharyya distance corresponding to the events w_i and w_j, each can be calculated by using Eq. 7 in a general form changing mean and covariance for each class ($D_h = D_{3,1}$ using μ_3, μ_1 and Σ_3, Σ_1; $D_h = D_{3,2}$ using μ_3, μ_2 and Σ_1, Σ_2) ([12], [14]).

Probabilistic Neural Network for a three-class problem was implemented for experimental data taken from whole tomato samples. The results obtained for the implementacion of the PNN and for the calculation of the Bayes risk is discussed in the next section. In the present work, the error probability of classification is calculated by using the Bayes theorem implemented in the Probabilistic Neural Network. The probability function $p(x|w_i)$, which corresponds to the Parzen window in the PNN for each class, can be used for the calculation of Bayes error by using Eq. 5; the Bayes error becomes Bayes risk [12] when the PNN is at the stage of decision node.

3 Experimental Description

In this section, materials and methods for the classification of tomatoes are described in detail. Whole tomato fruit samples, Charleston ("Lycopersicum esculentum"), were obtained from a greenhouse located in Sonora (México). The tomatoes were harvested and preliminarily sorted with colorimeter leaving only those with roughly Breaker color (USDA classification). A total of 236 whole tomato samples were used in the experiment. The experimental considered a storage temperature of 20 °C. The relative humidity of the cold chamber was 75%. Colorimeter, weight loss and nondestructive firmness measurements were taken for these samples at selected testing days.

Color measurements were performed using a Minolta CR 300 colorimeter; three replicas were taken for each intact tomato sample. The equipment provides an estimate of values L*, a*, and b* recommended by CIE [9] from where a*/b* ratio, chroma and hue angle = tan-1(a*/b*) can be calculated. Measurements of weight loss were done using a laboratory weighing instrument, Mettler Toledo model PR2003 Deltarange ® with readability of 1 mg. The impact acoustic tests were carried out using the experimental setup described in Figure 3. The specimen was insulated from external vibrations or noise using a simple plastic foam layer located underneath it, allowing the sample to vibrate freely. The impacting force was generated by a mechanical pendulum consisting of a small solid plastic sphere with a 24.9 mm diameter, 3.09 grms in weight and a string of 180 mm in length. The impacting sphere was released from an angle of θ =30 degrees as shown in Figure 3. Acoustic signals were recorded at two points: 0 degrees and 90 degrees in relation to the impact direction of the pendulum. The signals were captured by a sound sensor model CI-

6506B manufactured by PASCO Scientific and then digitalized by the Science Workshop 500 Interface from the same manufacturer. Finally, signals were post-processed by a personal computer equipped with FFT to determine the dominant resonance frequency. Three repetitions of impact acoustic tests were performed at each impact point to total 6 measurements per sample.

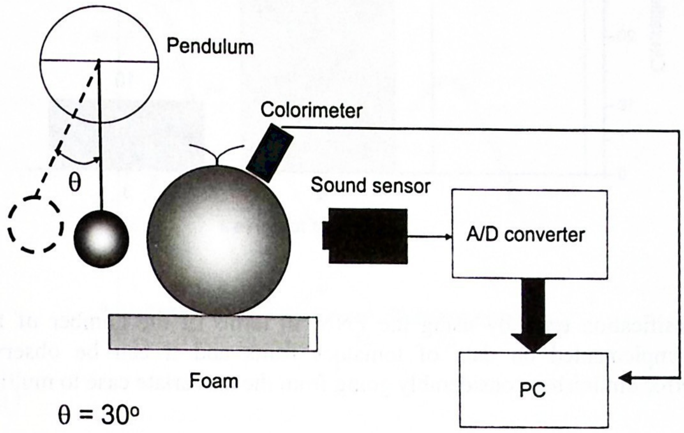


Figure 3. Experiment setup for impact acoustic and colorimeter tests on tomato samples

Three classes were chosen for training of the PNN, a total of 239 data vectors were considered for classification analysis. Data was sorted into three classes based on time in storage, Class 1: from days 0 to 6, class 2: days 6 to 15 and class 3: days 15 to 27. Then, 45 neurons (with known class) were randomly chosen for training data in the input layer. The features were put in vectors as follows: [a*/b*, Sc, L*]. Matlab 7.0 ® was used for the design and construction of the PNN.

4 Results

Tomato samples at a temperature of 20^0C degrees were studied. Data taken from the acoustic impact and colorimeter sensors for the tomato samples was presorted into two sets: training data and input data vectors [a*/b*; Sc; L*]. The data set consisted of 239 data vectors taken from samples at 20^0C; 45 vector data were used for training the network and 194 for classification. The optimum σ value for the three features (a*/b*, Sc, L*) with the best classification results was found at $\sigma = 0.1$. It is important to note that when only one feature is considered the corresponding error classification is about 60%. This is better observed in Figure 4, where a reduction in error classification from 60% to 10% as a function of number of features was estimated using the proposed PNN.

It can be also observed, that the results in Figure 5 agree with those in Figure 4.

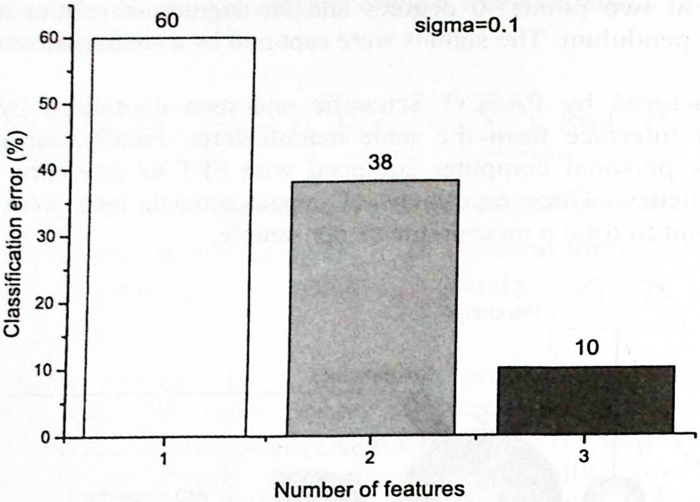


Figure 4. Classification error by using the PNN, in terms of the number of features. The network was implemented on data of tomatoes fruits and it can be observed that the classification error diminishes considerably going from the univariate case to multivariate case

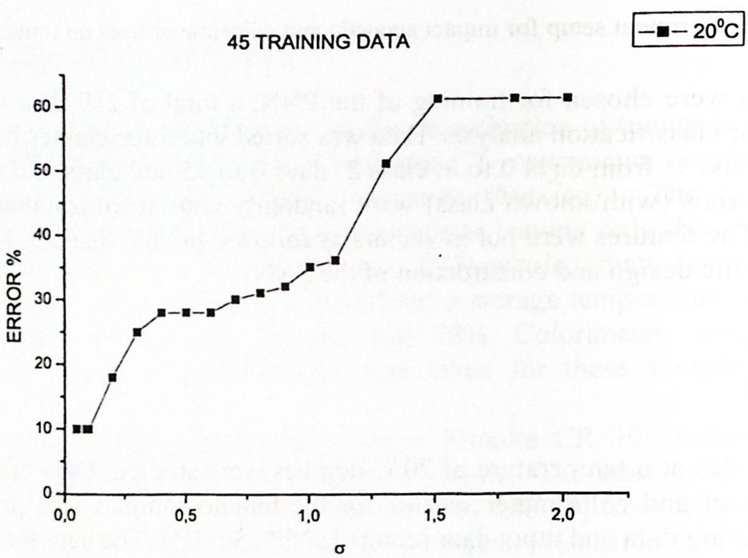
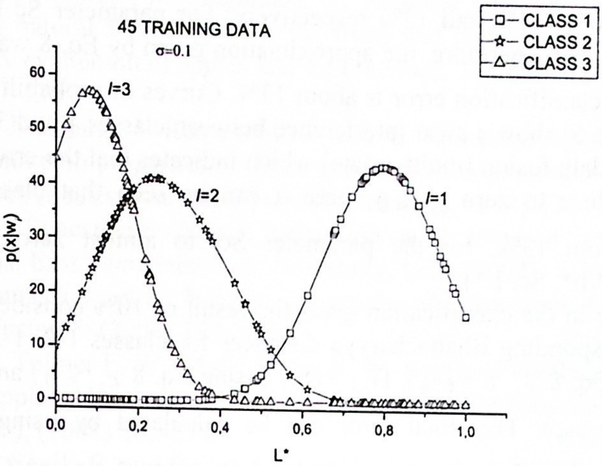
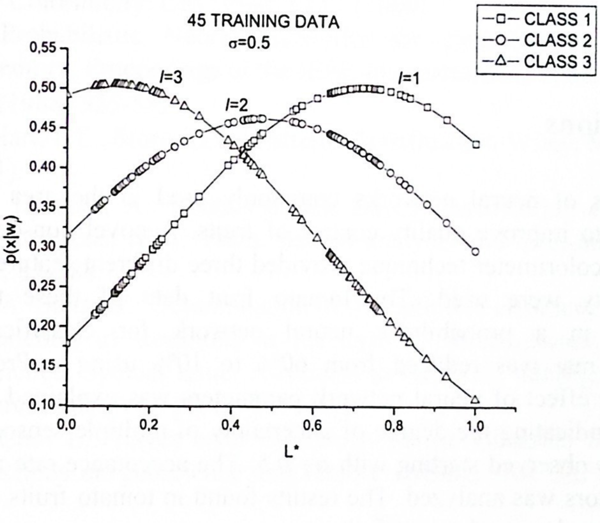


Figure 5. Results for the classification error in function of σ for data belonging to three classes and three features (a*/b*, Sc and L*). 45 vector data at 20^0 C were used for training and 239 input data vector for classification for tomato samples at 20^0 C were considered. The minimum error was obtained with a value of σ=0.1

The classification error in terms of the smoothing parameter σ in the PNN design is shown Figure 5 which was estimated as the number of times that the network incorrectly classified a known sample divided by the total number of input data. The results are the fusion of the three vector features corresponding to two sensors.
The probability functions for the multivariate multiclass case were obtained for each class and compared with the Bayes error classification calculated for $\sigma = 0.1$ (Eq. 9).



a)



b)

Figure 6. Probability functions corresponding to color parameter L* where data correspond are at 20^0 C. 45 training vector data and 239 input vector data, at 20^0 C, for classification were used for all figures with probability functions for three classes (*class 1 for square points, class 2 for circle points, class 3 for triangle points*). The figures correspond to different values of σ . a) $\sigma = 0.1$; b) $\sigma = 0.5$

The intersection between the corresponding curves probability functions for L* (luminosity) is more significant as the value of the σ increases (Figure 6).
The calculation of the Bhattacharyya distance and the classification error for the univariate and multivariate cases indicates that in the univariate case, Bhattacharyya distance between classes 1 and 3 is larger than 3.0. Therefore, bounds in Eq. 6 are valid to estimate the mean classification error; same criteria was used for classes 1-2 and 2-3 as defined for parameter L*. The classification error for classes 1-2 and 2-3 were estimated as 11% and 12% respectively. For parameter Sc the Bhattacharyya distance is $D_b = 0.7$, therefore, the approximation given by Eq. 8 was used; the results show that the classification error is about 13%. Curves of probability functions of L* feature (Figure 6) show a clear interference between classes 1 and 3. However, this is reduced after data fusion (multivariate) which indicates that the error between classes 1 and 3 is close to zero, $\varepsilon_{3,1} \approx 0$, here it can be seen that classification error is diminished from 13%, for the parameter Sc, to almost zero considering three parameters (a*/b*, Sc, L*).
The total error in the classification gives the result of 10% considering Eq. 9, in this case the corresponding Bhattacharyya distances for classes 1-2, 1-3 and 2-3 are $D_{2,1}$ =1.93 (using Eq. 8 $\varepsilon_{2,1} = 2.6\%$), $D_{3,1}$ =18.9 (using Eq. 8 $\varepsilon_{3,1} \approx 0$) and D_b =1.77 (using Eq. 8 $\varepsilon_{3,2} = 3.3\%$). This total error can be calculated by using Eq. 10 and the corresponding classification errors calculated for pairs of classes. Classification error obtained by PNN for two sensors and three features is in agreement with results of Bayes risk. This consideration can stand for three features and σ = 0.1 as discussed above.

5 Conclusions

Methodologies of neural networks commonly used in the area of robotics were implemented to improve quality control of fruits. A novel non-destructive acoustic impact and a colorimeter technique provided three different features: firmness, color and luminosity were used. The tomato fruit data of these two sensors was implemented in a probabilistic neural network for classification. The error classification rate was reduced from 60% to 10% using a Probabilistic Neural Network. The effect of neural network parameters was explained by an analysis of Bayes error, indicating the degree of uncertainty of multiple sensors during ripening stages. This is observed starting with σ= 0.5. The acceptance rate results as function of feature factors was analyzed. The results found in tomato fruits can be potentially extrapolated to other postharvest fruits.

References

1. Abbott, J., R. Lu, B.L., & Upchurch R.L.: Technologies for Nondestructive Quality of fruits and vegetables. Horticultural Reviews, 20, (1997) 1-120.
2. Cooke J.: An interpretation of the resonant behavior of intact fruits and vegetables. Transactions of the ASAE, 15(6), (1972) 1075-1080.
3. Yamamoto, H., Iwamoto, M. Haginuma, S.: Acoustic impulse response method for measuring natural frequency of intact fruits and preliminary applications to internal quality evaluation of apples and watermelons. Journal of Texture Studies, 11, (1980) 117-136.
4. Baltazar, A., J. Espina, I. Ramos-Torres and G. Gonzalez-Aguilar.: Effect of Methyl Jasmonate on properties of intact tomato fruit monitored with destructive and nondestructive tests. Journal of Food Engineering, 80 (4), (2007) 1086-1095.
5. Stephen S.S.: Synesthetic sensor fusion via a cross-wired artificial neural network, thesis. East Tennessee State University, USA (2002).
6. Kokar M., Tomasik J., et al: Formalizing classes of information fusion systems. Information Fusion 5: (2004) 189-202.
7. Shahin M. A., Tollner E. W., McCLendon R. W. Artificial intelligence classifiers for sorting apples based on Watercore. Journal of Agricultural Engineering Research, 79(3), (2001) 265-274.
8. Steinmetz V., Sévila F. and Bellon-Maurel V.: A methodology for sensor fusion design: Application to fruit quality assessment. Journal of Agricultural Engineering Research, 74(1), (1999) 21-31.
9. Marcus, R.T.: Colorimetry. CRC Press LLC. (2000).
10. Specht D.: Probabilistic Neural Networks for classification, mapping or associative memory. Proceedings of the IEEE International Conference on Neural Networks, 1, (1988) 525-532.
11. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, Second Edition. (2001).
12. Garber, F.D., Djouadi, A.: Bounds on the Bayes classification error based on pairwise risk functions. IEEE Trans. pattern analysis and machine intelligence. 10(2), (1988) 281-288.
13. Aranda, J. I., A. Baltazar.: Determination of the optimum error and reject validity regions as function of threshold in the case of three classes. Research in computing Science. Special issue: Advances in Artificial intelligence, Gelbukh and Reyes-García Eds., Vol. 26, (2006) 61-72.
14. Chulhee L., Choi E.: Bayes error evaluation of the Gaussian ML Classifier. IEEE: Transactions on geoscience and remote sensing, Vol. 38, No. 3, (2000) 1471-1475.