Machine Learning Models for Predicting Shelf Life of Processed Cheese

Sumit Goyal, Gyanendra Kumar Goyal

Abstract— **Feedforward multilayer machine learning artificial neural network (ANN) models were established for predicting shelf life of processed cheese stored at 7-8° C. Soluble nitrogen, pH, standard plate count, yeast & mould count, and spore count were input variables, and sensory score was the output variable. Mean Square Error, Root Mean Square Error, Coefficient of Determination and Nash–Sutcliffe Coefficient were used for comparing the prediction ability of the developed models. Feedforward ANN model with combination of 5→16→16→1 simulated best with high R^2: 0.998717294, suggesting that multilayer machine learning models can predict shelf life of processed cheese.**

Keywords— **Machine Learning, Soft Computing, Artificial Intelligence, ANN, Feedforward, Shelf Life, Prediction, Processed Cheese**

I. INTRODUCTION

An artificial neural network (ANN) is a system based on the operation of biological neural networks. Although, at present computing is quite advanced, but there are certain tasks that a program made for a common microprocessor is unable to perform; even so a software implementation of a neural network can be made with their advantages and disadvantages. Another aspect of the ANNs is that there are different architectures, which consequently require different types of algorithms, but despite to be an apparently complex system, a neural network is relatively simple [1]. ANNs are inspired by the early models of sensory processing by the brain. An ANN can be created by simulating a network of model neurons in a computer. By applying algorithms that mimic the processes of real neurons, one can make the network 'learn' to solve many types of problems. A model neuron is referred to as a threshold unit. It receives input from a number of other units or external sources, weighs each input and adds them up. If the total input is above a threshold, the output of the unit is one; otherwise it is zero. Therefore, the output changes from 0 to 1 when the total weighted sum of inputs is equal to the threshold. The points in input space satisfying this condition define a so called hyperplane. In two dimensions, a hyperplane is a line, whereas in three dimensions, it is a normal plane. Points on one side of the hyperplane are classified as 0 and those on the other side as 1. Thus, a classification problem can be solved by a threshold unit if the two classes can be separated by a hyperplane [2].
Feedforward backpropagation model consists of input, hidden and output layers. Backpropagation learning

Sumit Goyal – is at National Dairy Research Institute, Karnal, India. Email: thesumitgoyal@gmail.com
Gyanendra Kumar Goyal is at National Dairy Research Institute, Karnal, India. Email: gkg5878@yahoo.com

algorithm was used for learning these networks. During training this network, calculations were carried out from input layer of network toward output layer, and error values were then propagated to prior layers. Feedforward networks often have one or more hidden layers of sigmoid neurons followed by an output layer of linear neurons. Multiple layers of neurons with nonlinear transfer functions allow the network to learn nonlinear and linear relationships between input and output vectors. The linear output layer lets the network produce values outside the range –1 to +1. On the other hand, outputs of a network such as between 0 and 1 are produced, then the output layer should use a sigmoid transfer function. Multilayer networks consist of multiple layers of computational units, usually interconnected in a feed-forward way. Each neuron in one layer has directed connections to the neurons of the subsequent layer. In many applications the units of these networks apply a sigmoid function as an activation function. Multilayer networks use a variety of learning techniques, the most popular being back-propagation. Here, the output values are compared with the correct answer to compute the value of some predefined error-function. By various techniques, the error is then fed back through the network. Using this information, the algorithm adjusts the weights of each connection in order to reduce the value of the error function by some small amount. After repeating this process for a sufficiently large number of training cycles, the network usually converge to some state where the error of the calculations is small [3].

Processed cheese is very popular dairy product made from medium ripened Cheddar cheese, and sometimes a part of ripened cheese is replaced by fresh cheese. During its manufacture some amount of water, emulsifiers, extra salt, preservatives, food colorings and spices (optional) are added, and the mixture is heated to 70° C for 10-15 minutes with steam in a cleaned double jacketed stainless steel kettle, which is open, shallow and round-bottomed, with continuous gentle stirring (about 50-60 circular motions per minute) with a flattened ladle in order to get unique body & texture and consistency in the product. The determination of shelf life of processed cheese in the laboratory is very costly affair and takes a very long time to give results. It is alarming need of the day that ANN technique, which is fully equipped to predict the shelf life of food products, be employed for processed cheese as well. Hence, the present study was planned with the aim to develop feedforward multilayer machine learning models for predicting the shelf life of processed cheese stored at 7-8°C.

Shelf life is defined as the length of time that a product is acceptable and meets the consumer's expectations regarding

food quality. It is the result of the conjunction of all services in production, distribution, and consumption. Shelf life dating is one of the most difficult tasks in food engineering. Market pressure has led to the implementation of shelf life by sensory analyses, which may not reflect the full quality spectra. Moreover, traditional methods for shelf-life dating and small-scale distribution chain tests cannot reproduce in a laboratory the real conditions of storage, distribution, and consumption on food quality. The consumer demands foods under the legal standards, at low cost, high standards of nutritional, sensory, and health benefits [4]. Shelf life studies provide important information to product developers enabling them to ensure that the consumer will see a high quality product for a significant period of time after production. Since long time taking shelf life studies do not fit with the speed requirement, hence new accelerated studies have been developed [5]. Machine learning models have been applied for predicting properties of potato chips [6], goat whole milk powder [7-8], for predicting total acceptance of ice cream [9], for prediction of meat spoilage [10], for predicting shelf life of processed cheese [11], for prediction of the type of milk, degree of ripening in cheeses [12], for predicting viscoelastic behavior of pomegranate [13], for estimating shelf life of burfi [14] and for estimating antioxidant activity and anthocyanin content of sweet cherry during ripening [15]. The results of this research would be very beneficial for consumers, dairy factories manufacturing processed cheese, wholesalers, retailers, food researchers, regulatory authorities and academicians.

II. METHOD MATERIAL

The input variables used in the ANN were the experimental data of processed cheese relating to soluble nitrogen, pH; standard plate count, Yeast & mould count, and spore count. The sensory score was taken as output variable for developing machine learning models (Fig.1).
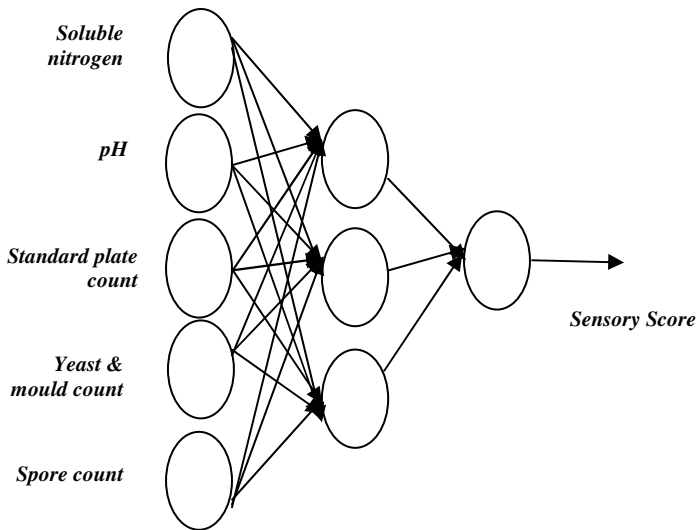

Fig. 1. Input and output variables for machine learning models

Thirty six observations for each input and output variables were used for developing the models. The dataset was randomly divided into two disjoint subsets, namely, training set having thirty observations (80% for training), and validation set consisting of six observations (20% for testing) [16-17].

Mean Square Error MSE (1), Root Mean Square Error RMSE (2), Coefficient of Determination R^2 (3) and Nash–Sutcliffe Coefficient E^2 (4) were applied in order to compare the prediction ability of the developed models. *Bayesian regularization* mechanism was used for training the artificial neural networks, as it exhibited the best results. The network was trained up to 100 epochs, and neurons in each hidden layers varied from 1 to 20. The ANN was trained with single as well as multiple hidden layers, and transfer function for hidden layer was *tangent sigmoid,* while for the output layer it was *pure linear* function. MALTAB software was used for performing the experiments.

$$MSE = \left[\sum_1^N\left(\frac{Q_{\exp}-Q_{cal}}{n}\right)^2\right] \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\left[\sum_1^N\left(\frac{Q_{\exp}-Q_{cal}}{Q_{\exp}}\right)^2\right]} \quad (2)$$

$$R^2 = 1-\left[\sum_1^N\left(\frac{Q_{\exp}-Q_{cal}}{Q_{\exp}^2}\right)^2\right] \quad (3)$$

$$E^2 = 1-\left[\sum_1^N\left(\frac{Q_{\exp}-Q_{cal}}{Q_{\exp}-\overline{Q_{\exp}}}\right)^2\right] \quad (4)$$

Where,

$Q_{\exp}$ = Observed value;

Q_{cal} = Predicted value;

$\overline{Q}_{\exp}$ =Mean predicted value;

n = Number of observations in dataset.

Several problems were faced while training ANN's, too many neurons in the hidden layers resulted in overfitting. Overfitting occurs when the neural network has so much information processing capacity that the limited amount of information contained in the training set is not enough to train all of the neurons in the hidden layers. Using too few neurons in the hidden layers also resulted in underfitting. Underfitting occurs when there are too few neurons in the hidden layers to adequately detect the signals in a complicated dataset. A second problem can occur even when there is sufficient training data resulting in increase in training time of the network. Obviously some compromise must be reached between too many and too few neurons in

the hidden layers. Ultimately the selection of the architecture of neural network came down to trial and error. There are two trial and error approaches that are used in determining the numbers of hidden neurons: the "forward" and "backward" selection methods. The first method, the "forward selection method", begins by selecting a small number of hidden neurons. The second method, the "backward selection method", begins by using a large number of hidden neurons. Then the neural network is trained and tested. This process continues until the performance improvement of the neural network is no longer significant [18].

III. RESULTS AND DISCUSSION

The comparison of Actual Sensory Score (ASS) and Predicted Sensory Score (PSS) for machine learning feedforward multilayer models is illustrated in Fig.2.

Table 1. Results of Feedforward multilayer ANN model

Neurons	MSE	RMSE	R^2	E^2
3:3	9.1366E-05	0.009558554	0.990441446	0.999908634
4:4	9.14141E-05	0.009561074	0.990438926	0.999908586
5:5	9.14141E-05	0.009561074	0.990438926	0.999908586
6:6	9.14623E-05	0.009563594	0.990436406	0.999908538
7:7	0.000118738	0.010896701	0.989103299	0.999881262
8:8	9.15105E-05	0.009566114	0.990433886	0.999908489
9:9	0.000381535	0.019532917	0.980467083	0.999618465
10:10	0.000414509	0.020359493	0.979640507	0.999585491
11:11	0.000427435	0.0206745	0.9793255	0.999572565
12:12	0.000208365	0.014434853	0.985565147	0.999791635
13:13	3.7192E-05	0.006098524	0.993901476	0.999962808
14:14	0.000571338	0.023902685	0.976097315	0.999428662
15:15	0.000822872	0.028685743	0.971314257	0.999177128
16:16	**1.64533E-06**	**0.001282706**	**0.998717294**	**0.999998355**
17:17	0.000278146	0.016677699	0.983322301	0.999721854
19:19	0.000331967	0.01821997	0.98178003	0.999668033
20:20	0.000424107	0.020593858	0.979406142	0.999575893

A combinations of 5→16→16→1 (**MSE:** 1.64533E-06; **RMSE: 0.001282706**; **R^2: 0.998717294**; E^2: 0.999998355) gave the best result (Table 1). Goyal and Goyal [18] established Elman machine learning ANN models for predicting shelf life of processed cheese stored at 7-8ºC. Input parameters for their models were: Body & texture, aroma & flavour, moisture, and free fatty acid, while sensory score was output parameter. Bayesian regularization was training algorithm for the models. The network was trained up to 100 epochs, and neurons in each hidden layers varied from 1 to 20. Transfer function for hidden layer was tangent sigmoid, while for the output layer it was pure linear function. MSE, RMSE, R^2 and E^2 were used for comparing the prediction ability of the developed models. Elman model with combination of 4→17→17→1 (MSE: 3.68747E-06; RMSE: 0.001920279; R^2:0.998079721; E^2: 0.999996313) performed significantly well for predicting the shelf life of processed cheese stored at 7-8º C.

Bahramparvar et al. [9] used machine learning ANN models to predict the total acceptance of ice cream. The experimental sensory attributes (appearance, flavor, body and texture, coldness, firmness, viscosity, smoothness and liquefying rate) were used as inputs and independent total acceptance was output of ANN.
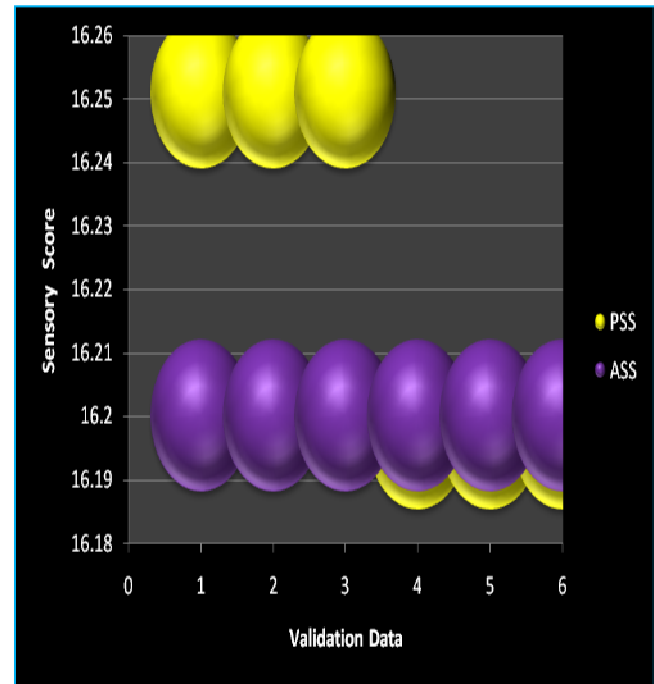

Fig. 2. Comparison of ASS and PSS

Thirty, ten and sixty percent of the sensory attributes data were used to train, validate and test the ANN model, respectively. It was found that ANN with one hidden layer comprising 10 neurons gave the best fitting with the experimental data, which made it possible to predict total acceptance with acceptable mean absolute errors (0.27) and correlation coefficients (0.96). Their sensitivity analysis result showed that flavor and texture were the most sensitive sensory attributes for prediction of total acceptance of ice cream.

Time-delay and linear layer (design) intelligent computing expert system models for predicting shelf life of soft mouth melting milk cakes stored at 6ºC were implemented. The best results for time-delay model with single hidden layer having 20 neurons were MSE: 0.001332342, RMSE: 0.036501259, R^2: 0.984011897 for time-delay model with double hidden layers having 8 neurons in the first and second layers MSE: 0.001318004, RMSE: 0.036304329, R^2: 0.984183948. Best results for linear layer (design) model were MSE: 0.000293366, RMSE: 0.017127919, R2: 0.996479613, suggesting that the intelligent computing expert system models are efficient in predicting the shelf life of soft mouth melting milk cakes [19]. Radial Basis (Exact Fit) and Linear Layer (Design) models were developed for predicting shelf life of processed cheese stored at 30º C. Several experiments were carried out in order to get to good results. The best results were observed for Radial Basis (Exact fit) model with 30 neurons and spread constant as 20 (MSE: 1.81045E-06, RMSE: 0.001345528, R^2: 0.998654472, E^2: 0.99999819) for predicting shelf life of processed cheese stored at 30º C [20]. The efficiency of Cascade hidden layer models was tested for shelf life

prediction of Kalakand, a sweetened desiccated dairy product. For developing the models, the network was trained with 100 epochs. Cascade models with two hidden layers having twenty neurons in the first layer and twenty neurons in the second layer gave best result (MSE 0.000988770; RMSE: 0.03144471; R^2: 0.988125331) [21]. Recently, linear layer (train) and generalized regression models were developed and compared with each other for predicting the shelf life of milky white dessert jeweled with pistachios. Neurons in each hidden layers varied from 1 to 30. Datasets were divided into two sets, i.e., 80% of data samples were used for training and 20% for validating the network. MSE, RMSE, R^2 and E^2 were applied in order to compare the prediction performance of the developed models. The study revealed that artificial neural networks are quite effective for determining the shelf life of milky white dessert jeweled with pistachios [22].

In principal, these results are in harmony with the findings of this research. Therefore, Feedforward machine learning ANN models have the potential for predicting shelf life of processed cheese.

IV. CONCLUSION

Machine learning feedforward multilayer ANN models were established for predicting the shelf life of processed cheese stored at 7-8° C. The results of the study established very good correlation between the experimental data and the predicted values, with a high determination coefficient, establishing that the developed feedforward models were able to analyze non-linear multivariate data with excellent performance, fewer parameters, and shorter calculation time.

REFERENCES

[1]   Learn artificial neural networks Website: http://www.learnartificialneuralnetworks.com/ (accessed on 1.4.2011).

[2]   A. Krogh, "What are artificial neural networks?," *Nature Biotechnology,* vol.26, no.2, pp.195-197, 2008.

[3]   H. Demuth, M. Beale and M. Hagan, "*Neural Network Toolbox User's Guide,*" The MathWorks, Inc., Natrick, USA. 2009.

[4]   R.C. Martins, V.V. Lopes, A.A. Vicente, and J.A. Teixeira, "Computational shelf-life dating: complex systems approaches to food quality and safety," *Food and Bioprocess Technology*, vol.1, no.3, pp. 207-222, 2008.

[5]   Medlabs Website: http://www.medlabs.com/Downloads/food_product_shelf_life_web.pdf (accessed on 21.5.2011)

[6]   T. Marique, A. Kharoubi, P. Bauffe, and C. Ducattillon, " Modeling of fried potato chips color classification using image analysis and artificial neural network," *Journal of Food Science*, vol.68, no.7, pp. 2263-2266, 2003.

[7]   Sumit, Goyal, S. Kar, and G.K. Goyal, "Artificial neural networks for analyzing solubility index of roller dried goat whole milk powder," *International Journal of Mechanical Engineering and Computer Applications*, vol.1, no.1, pp. 1-4, 2013.

[8]   Sumit Goyal and G.K. Goyal, " Radial basis artificial neural network models for predicting solubility index of roller dried goat whole milk powder," In: V. Snášel et al. eds. Soft Computing in Industrial Applications. *Advances in Intelligent Systems and Computing 223*, DOI: 10.1007/978-3-319-00930-8_21. Chapter No.: 21, Book ID: 311964_1_En Book. ISBN: 978-3-319-00929-2. Publisher: Springer International Publishing, Switzerland, 2013.

[9]   M. Bahramparvar, S. Fakhreddin, and S. Razavi, "Predicting total acceptance of ice cream using artificial neural network," *Journal of Food Processing and Preservation*, doi: 10.1111/jfpp.12066, 2013.

[10] A.A. Argyri, R.M. Jarvis, D. Wedge, Y. Xu, E.Z. Panagou, R. Goodacre, and G.J.E. Nychas, " A comparison of Raman and FT-IR spectroscopy for the prediction of meat spoilage," *Food Control*, vol. 29, no.2, pp. 461-470, 2013.

[11] Sumit Goyal and G.K. Goyal, "Intelligent artificial neural network computing models for predicting shelf life of processed cheese," *Intelligent Decision Technologies*, vol.7, no.2, pp. 107-111, 2013.

[12] M.C. Soto-Barajas, M.I. González-Martín, J. Salvador-Esteban,J.M. Hernández-Hierro, V. Moreno-Rodilla, A.M. Vivar-Quintana, I. Revilla, I.L. Ortega, R. Morón-Sancho, and B. Curto-Diego, " Prediction of the type of milk and degree of ripening in cheeses by means of artificial neural networks with data concerning fatty acids and near infrared spectroscopy," *Talanta*, vol.116, pp. 50-55, 2013.

[13] M.H. Saeidirad, A. Rohani, and S. Zarifneshat, "Predictions of viscoelastic behavior of pomegranate using artificial neural network and Maxwell model," *Computers and Electronics in Agriculture*, vol. 98, pp. 1-7, 2013.

[14] Sumit Goyal, and G.K. Goyal, "Artificial vision for estimating shelf life of burfi," *Journal of Nutritional Ecology and Food Research*, vol.1, no.2, pp. 134-136, 2013.

[15] S. Taghadomi Saberi, M. Omid, Z. Emam Djomeh, and H. Ahmadi, " Evaluating the potential of artificial neural network and neuro fuzzy techniques for estimating antioxidant activity and anthocyanin content of sweet cherry during ripening by using image processing," Journal of the Science of Food and Agriculture. doi: 10.1002/jsfa.6202, 2013

[16] Sumit Goyal, "Artificial neural networks (ANNs) in food science–A review," *International Journal of Scientific World*, vol.1, no.2, pp. 19-28, 2013.

[17] Sumit Goyal, "Artificial neural networks in vegetables: A comprehensive review," *Scientific Journal of Crop Science*, vol.2, no.7, pp. 75-94, 2013.

[18] Sumit Goyal, and G. K. Goyal, "Artificial neural network simulated Elman models for predicting shelf life of processed cheese," . *International Journal of Applied Metaheuristic Computing*, vol.3, no.3, pp. 20-32, 2012.

[19] Sumit Goyal, and G. K. Goyal, "Time – delay simulated artificial neural network models for predicting shelf life of processed cheese," *International Journal of Intelligent Systems and Applications*, vol.4, no.5, pp.30-37, 2012.

[20] Sumit Goyal, and G. K. Goyal, "Radial basis (exact fit) and linear layer (design) ANN models for shelf life prediction of processed cheese," *International Journal of u- and e- Service, Science and Technology*, vol.5, no.1, pp. 63-69, 2012.

[21] Sumit Goyal, and G. K. Goyal, "Advanced computing research on cascade single and double hidden layers for detecting shelf life of kalakand: An artificial neural network approach," *International Journal of Computer Science & Emerging Technologies*, vol.2, no.5, pp.292-295, 2011.

[22] Sumit Goyal, and G. K. Goyal, "A new scientific approach of intelligent artificial neural network engineering for predicting shelf life of milky white dessert jeweled with pistachio,". *International Journal of Scientific and Engineering Research*, vol.2, no.9, pp.1-4, 2011.