

Bayesian Regularized Neural Network for Prediction of the Dose in Gamma Irradiated Milk Products

M. Terziyska¹, Y. Todorov², D. Miteva³, M. Doneva³, S. Dyankova³, P. Metodieva³, I. Nacheva³

¹University of Food Technologies, 4000 Plovdiv, Bulgaria

²VTT Technical Research Center of Finland, Finland

³Institute of Cryobiology and Food Technologies, 1407 Sofia, Bulgaria

E-mails: m.terziyska@uft-plovdiv.bg yancho.todorov@ieee.org daniela.miteva@ikht.bg
maria.doneva@ikht.bg svetla.diankova@ikht.bg petya.metodieva@ikht.bg iliana.nacheva@ikht.bg

Abstract: Gamma irradiation is a well-known method for sterilizing different foodstuffs, including fresh cow milk. Many studies witness that the low dose irradiation of milk and milk products affects the fractions of the milk protein, thus reducing its allergenic effect and make it potentially appropriate for people with milk allergy. The purpose of this study is to evaluate the relationship between the gamma radiation dose and size of the protein fractions, as potential approach to decrease the allergenic effect of the milk. In this paper, an approach for prediction of the dose in gamma irradiated products by using a Bayesian regularized neural network as a mean to save recourses for expensive electrophoretic experiments, is developed. The efficiency of the proposed neural network model is proved on data for two dairy products – lyophilized cow milk and curd.

Keywords: Bayesian neural network, milk products, milk allergy, protein fraction.

1. Introduction

Nowadays, Artificial Intelligence (AI) techniques such as Neural Networks (NN), Fuzzy Logic (FL) and Neuro-Fuzzy Systems (NFS) are widely used in different domains like technology, medicine and biotechnology. In the scientific literature, there are also many studies, which are focused on solving specific problems in the food industry [4]. For instance, Artificial Neural Network (ANN) method was used to predict the drying characteristics of agricultural products such as hazelnut, bean and chickpea in [3], while in [5] a neural network model capable of predicting problems in fermentations of Cabernet Sauvignon has been presented. A dynamic model of the kneading process based on artificial neural networks which allows predicting the bread dough temperature and the delivered power necessary to carry out mechanical work, was proposed in [6]. To demonstrate the use of ANFIS in food process modelling, in [7] two case studies are discussed. The first case discusses the use of ANFIS in modelling of complex and nonlinear time-dependent intermittent

drying of spouted grains, while the second case demonstrates its use in modelling of fuzzy sensory attributes of espresso coffee by pod. The potential of the application of modern soft computing techniques into development stage of contemporary food products is also reviewed in [8, 9].

In many situations, the availability of data from the food industry may be limited, due to various reasons. For instance, some of experiments are time consuming and require tedious work, while other are too expensive to be repeated, in order to generate a comprehensive data set. Therefore, developing a good classical neural network or neuro-fuzzy one is quite difficult task, as the low availability of data can result to network overfitting and generation of inconsistent results.

An available solution to cope with the lack of extensive data is to use the Bayesian probabilistic approach, which has proven to be useful in different cases related to biological applications, including molecular biology [10-11], food industry [12-13] and medical diagnosis, areas where data often come from costly and difficult experimental work. The Bayesian neural networks are universally useful and can address a great variety of applications, but they are extremely difficult to scale to large problems.

This paper aims at studying the application of the classical Bayesian approach in a specific case related to milk and milk products. It has been proven already, that low dose gamma irradiation breaks the milk protein into small fractions which leads to substantial decrease in the allergenic effect of milk and dairy products, but extensive studies does not exist in the literature for a great variety of cases. However, developing Bayesian Regularized Neural Network (BRANN) model is expected to save resources on extremely expensive electrophoretic experiments, while providing valuable insights on the protein degradation for variety of dairy products. To test the efficiency of the proposed modeling approach data generated from experiments with two dairy products – lyophilized cow milk and curd were used. The results show the potential of the generated BRANN model to capture the trend on the dose of irradiation and its impact on the protein fractions.

2. Bayesian regularized neural networks

2.1. Feedforward Neural Networks and Regularization

Artificial Neural Networks (ANN) mimic biological neural systems. The most popular type ANN is the MultiLayer Perceptron (MLP) which is also known as feedforward neural network. The MLP architecture consists of an input layer, one or several hidden layers and an output layer. The neurons (nodes) in the input layer receive independent variables and transmit them to the hidden layer for processing. The nodes in the hidden layer take inputs from the input layer, perform some calculations and transform the result to the output nodes. In the output layer, the prediction or classification process is conducted and the model's output is deducted. Most neural networks are fully connected, which means that each hidden neuron and each output neuron are connected to every unit in the layers either side. The connections between one node and another are represented by weight coefficients that represent their relative importance. A given neuron can have multiple input and

output connections. In general, a backpropagation algorithm is used to train a feedforward network. During the training process, the weight coefficients are updated to minimize a sum of squared errors:

$$(1) \quad E_D = \frac{1}{2} e^2(k) = \sum_{i=1}^n \frac{1}{2} (y_i(k) - \hat{y}_i(k))^2,$$

where \hat{y}_i represents the neural network response, n is the number of training patterns, and E_D is performance function used for computing the distance between real and predicted data. The minimization of (1) can fall into one of many local minima, not in the global minimum.

Most prominent advantage of backpropagation is its fast, simple and easy realization. At the same time, the actual performance of backpropagation on a specific problem is dependent on the input data. Therefore, it is not applicable for small dataset since it can result in overfitting. In addition, the backpropagation algorithm is quite sensitive to noisy data.

To avoid the shortcomings of the conventional backpropagation neural networks, some regularization techniques can be used. The process of regularization consists of putting an additional term to the performance function in which (1) is transformed to the form:

$$(2) \quad E(w) = \beta \sum_{i=1}^n \frac{1}{2} (y_i - \hat{y}_i)^2 + \alpha \sum_{i=1}^n \frac{1}{2} w^2 = \beta E_D + \alpha E_W.$$

The last term, αE_W , is called *weight decay*, because of its effect to bias the procedure in favor of small weights, while α is known as *decay rate*. The α and β parameters are also known as *hyperparameters* and their value is essential for the training process. If $\alpha \ll \beta$ then the training algorithm will project smaller errors and if $\alpha \gg \beta$, training will emphasize weight size reduction at the expense of network errors, thus producing a smoother network response [14].

2.2. Bayesian framework

Bayesian learning is one of the main approaches in Machine Learning (ML). The main idea of Bayesian learning lies on the Bayes' theorem given with the following equation:

$$(3) \quad P(A|B) = \frac{P(A)P(B|A)}{P(B)},$$

where A and B are events that have nonzero probabilities $P(A) > 0$ and $P(B) > 0$, $P(A)$ is the *prior* probability, $P(A|B)$ is the *posterior* probability, and $P(B)$ is the marginal probability of the event B , which acts as a normalization factor in Bayes' rule. It can be concluded that the Bayesian theorem (3) is used to infer the calculation of the conditional probability $P(A|B)$ which can be reduced to the calculation of the conditional probability $P(B|A)$, that could be a simple task.

2.3. Bayesian regularized artificial neural network

In contrast to the classical methods of ML, in which point estimates for the parameters of the neural network w are found, in Bayesian neural networks, the weight coefficients are considered random variables with probability distributions called a

priori distributions. The density function for the weights can be updated according to Bayes' rule (6):

$$(4) \quad P(w|D, \alpha, \beta, M) = \frac{P(D|w, \beta, M)P(w|\alpha, M)}{P(D|\alpha, \beta, M)},$$

where w is the vector of network weights, M is particular neural network model used and D represents the data set.

Assuming that the noise in the training set data is Gaussian and that the prior distribution for the weight coefficients is Gaussian, the probability densities can be written as:

$$(5) \quad P(D|w, \beta, M) = \frac{1}{Z_D(\beta)} \exp(-\beta E_D),$$

where the use of the sum-squared error E_D (2) corresponds to an assumption of Gaussian noise on the target variables, the parameter β defines a noise level $\sigma^2=1/\beta$ and $Z_D(\beta) = \int \exp(-\beta E_D) dD$ is a normalizer.

$$(6) \quad P(w|\alpha, M) = \frac{1}{Z_W(\alpha)} \exp(-\alpha E_W),$$

where, if E_W (2) is quadratic one (as it is assumed), then the corresponding prior distribution is a Gaussian with variance $\sigma^2=1/\alpha$ and the normalizer is $Z_W(\alpha) = \int \exp(-\alpha E_W) dw$.

Finally, with $P(D|w, \beta, M)$ and $P(w|\alpha, M)$ defined and according to Bayes rule, the posterior distribution can be obtained in the following form:

$$(7) \quad P(w|D, \alpha, \beta, M) = \frac{1}{Z_W(\alpha)Z_D(\beta)} \exp(-(\beta E_D + \alpha E_W)) = \frac{1}{Z_F(\alpha, \beta)} \exp(-E(w)),$$

where $Z_F(\alpha, \beta) = \int \exp(-E) dw$.

Equation (5) is called the *likelihood function*. It is a function of the network weights w , and it describes how likely a given data set is to occur, given a specific set of weights [1]. The maximum likelihood method selects the weights to maximize the likelihood function, which in this Gaussian case is the same as minimizing the squared error E_D . Equation (6) is called the *prior density*. It embodies our knowledge about the network weights before we collect any data. The last term in (4) – $P(D|\alpha, \beta, M)$ – is called *evidence*. It is a normalizing term that is not a function of w . If the objective is to find the weights w that maximize the posterior density (4), then there is no need to be concerned with the evidence. This term is important for optimizing the hyperparameters α and β . In order to do this, using the Bayes' theorem, it can be written:

$$(8) \quad P(\alpha, \beta|D, M) = \frac{P(D|\alpha, \beta, M)P(\alpha, \beta|M)}{P(D|M)}.$$

A detailed description of the algorithm for Bayesian optimization of the regularization parameters, with the Gauss-Newton approximation to Hessian matrix can be found in [14].

Summarizing the above, Bayesian Regularized Artificial Neural Networks (BRANN) are conventional backpropagation networks with an additional term added to the objective function (2) and with a Bayesian-based method for terminating the training [2]. In order to develop BRANN, two optimizations must be performed. The first one is minimize (2) with respect to the weights, while the second one – maximize (11) with respect to hyperparameters α and β until self-consistency is achieved. This whole process is illustrated in Fig. 1.



Fig. 1. Flowchart for a BRANN learning process

3. Results and discussions

3.1. Food irradiation process

Food irradiation has the ability to destroy the microorganism's DNA thereby prolonging shelf-life and enhancing food safety without detrimental effect on the sensorial and nutritional quality when applying the appropriate dose. Nowadays, irradiation is being widely used for animal feed decontamination, sterilization of food for immune-compromised patients needful of sterile diets and in medicine for diagnosis, treatment, sterilization of equipment, etc. [15].

There is slow adoption of the irradiation of dairy products due to the effective elimination of pathogens by pasteurization. Reports about development of off-flavors in irradiated dairy products also hindered the use of the process, though this has been contradicted by research concluding that irradiating at low doses and/or in frozen conditions can be an effective treatment without compromising the organoleptic properties. However, success in the improvement of microbial quality of dairy

products by gamma irradiation has been reported in [16]. In addition, the allergy to cow milk and dairy products is one of the most widespread food allergies. It was proven by [17] that the allergenic effect of milk and products can be reduced by low dose gamma irradiation that causes various changes in the nutritional components, including breakdown of the protein epitopes, which is the main reason to reduce their allergenic effect. Conducting studies in this topic is quite complicated as relatively large number of experiments have to be conducted in determining the appropriate dose for different types of milk product that further require expensive analyses of the proteins and by products of irradiation, which hinder the further developments in this area. Therefore, this paper is focused on proposing an approach to deal with the limited data availability by using the Bayesian inference, as potential approach on saving resources for experiments.

3.2. Simulation experiments with BRANN

In this paper, a BRANN was developed to predict the dose in gamma irradiated milk products. Different structures of BRANN are studied and the results are summarized in Table 1. Since the smallest Mean Squared Error (MSE) is reached in a case with three-layer BRANN with 10 neurons in hidden layer, it is chosen to make further simulations with it in Matlab environment. In this study, as test mediums lyophilized cow milk and curd irradiated with low dose of γ -rays in the range 0-10 kGy are used. The densitograms for each dose after a number of repeats is obtained by electrophoresis, that shows the protein fractions as bands, whose area is used as BRANN inputs and correspond to the protein raw/molar volume, which is affected by irradiation. Two data sets of 125 samples for each product, are obtained during biological experiments and used for training of the proposed model, whose initially generated weights have a normal distribution.

Table 1. Different BRANN structures MSE results

No	BRANN structures	MSE milk data set	MSE curd data set
1	Three-layer, 5 neurons in hidden layer	0.0285114 at epoch 27	0.038612 at epoch 22
2	Three-layer, 10 neurons in hidden layer	0.023525 at epoch 24	0.034862 at epoch 17
3	Three-layer, 15 neurons in hidden layer	0.031759 at epoch 32	0.037149 at epoch 27
4	Four-layer, 5 neurons in hidden layers	0.033622 at epoch 43	0.040182 at epoch 33
5	Four-layer, 10 neurons in hidden layers	0.029123 at epoch 37	0.043357 at epoch 37
6	Four-layer, 15 neurons in hidden layers	0.026397 at epoch 40	0.039084 at epoch 31

On Fig. 2 and Fig. 7 the graphical representation of the normalized dose of γ -rays estimated based on the BRANN structure and the true normalized value, respectively for lyophilized cow milk and curd, are depicted. It can be seen that almost all predicted values are close to true values. There is only negligible difference between estimated and true values. This indicates that the proposed BRANN network is efficient.

Figs 3 and 8 show the results of BRANN estimation plotted as normalized target values versus predicted γ -rays dose. The dashed line indicates a line with a slope of one, which would indicate a perfect fit. It can be seen that most of the data lie close to the best fit line (or regression line). This means that the residual errors are very

small. This is confirmed by the value of the correlation coefficient, which for milk dataset is $R=0.96388$ and for curd dataset is $R=0.94242$.

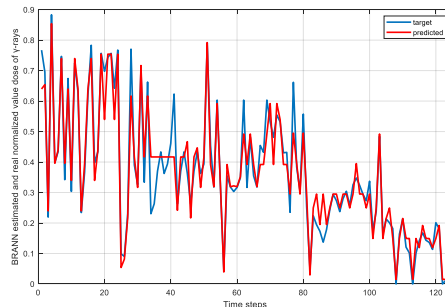


Fig. 2. Normalized values of the predicted and actual dose of γ -rays for lyophilized cow milk data set

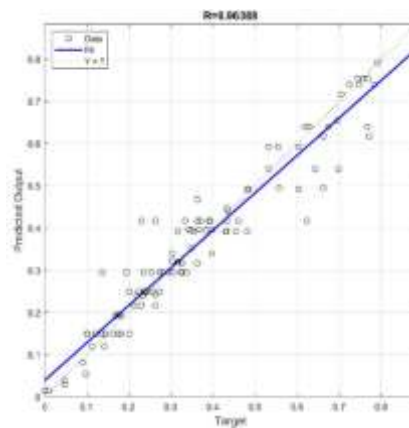


Fig. 3. Scatter plot of BRANN prediction dose of γ -rays for lyophilized cow milk data set. The line in the scatter plot represents the linear regression line

The performance graph of the BRANN model, for lyophilized cow milk and curd respectively, is shown on Fig. 4 and Fig. 9. The proposed model achieves the lowest MSE at an epoch of 24 in a case of milk dataset using and in this point the best performance of the system is realized. In a case of curd dataset using, the proposed model achieves the lowest MSE at an epoch of 17.

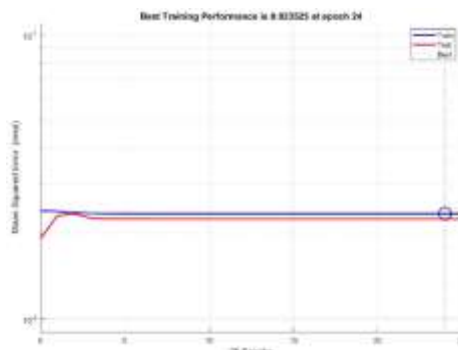


Fig. 4. Performance graph of the BRANN prediction dose of γ -rays for lyophilized cow milk data set

On Fig. 5 and Fig. 10 the error histogram of the BRANN system, respectively for lyophilized cow milk and curd, are shown. The entire error range is divided into 20 bins. The vertical yellow line at the center of the histogram indicates zero error. It can be seen that most frequently the error values are between -0.039 and $+0.039$ in the case of milk dataset. In the case of curd dataset the most frequent error values are even smaller – between -0.022 and $+0.035$.

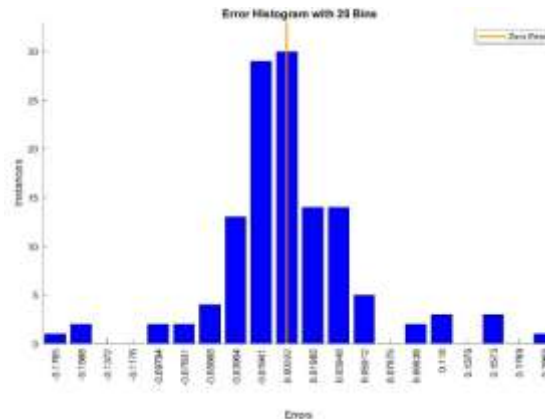


Fig. 5. Error histogram for lyophilized cow milk data set

The boxplots of BRANN prediction for milk dataset and curd dataset is given in Fig. 6 and Fig. 11, respectively. It can be seen that two boxes for target and predicted data are overlapping. In addition, the mean value for both data sets is the same (the red line inside the box). Therefore, the target and predicted data sets are quite similar and it can be concluded that the BRANN model has good performance characteristics. Together with the boxplot, the whiskers show how big is the range between the maximum and the minimum values of each set. The larger ranges indicate a wider distribution, that is, more scattered data which is the case with the predicted data set. The boxplots show also that data sets include outliers. These are the red crosses outside the whiskers.

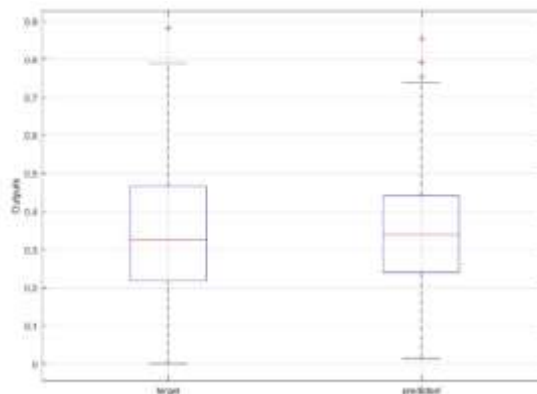


Fig. 6. Boxplots for lyophilized cow milk data set

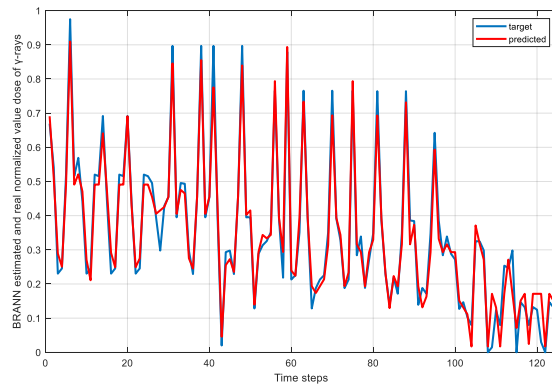


Fig. 7. Normalized values of the predicted and actual dose of γ -rays for curd data set

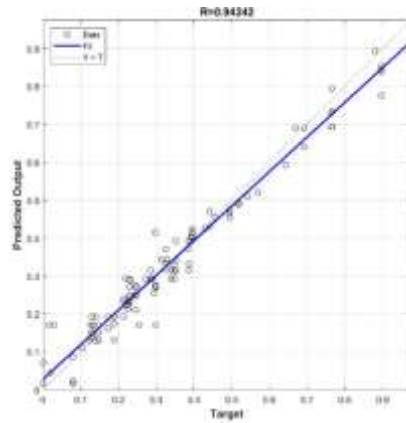


Fig. 8. Scatter plot of BRANN prediction dose of γ -rays for curd data set. The line in the scatter plot represents the linear regression line

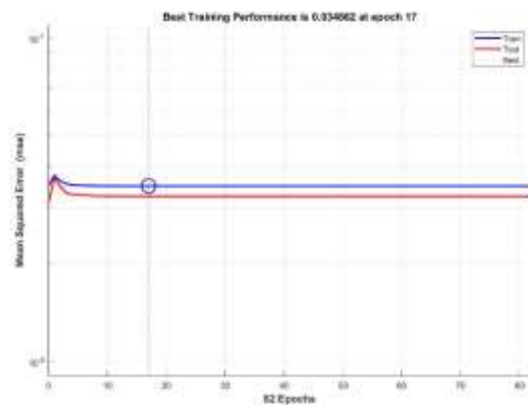


Fig. 9. Performance graph of the BRANN prediction dose of γ -rays for curd data set

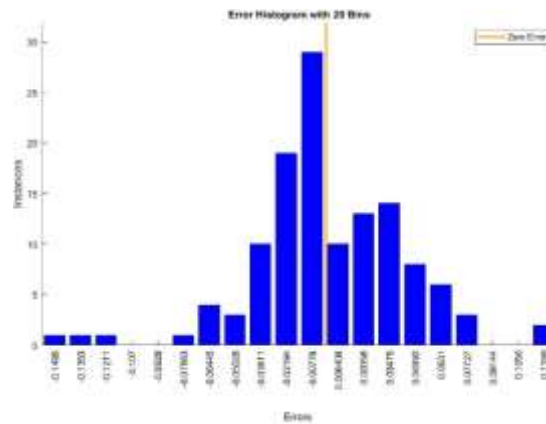


Fig. 10. Error histogram for curd data set

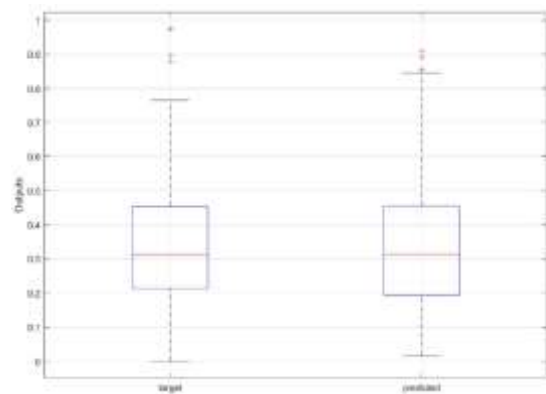


Fig. 11. Boxplots for curd data set

4. Conclusions

In this paper a Bayesian Regularized Artificial Neural Network (BRANN) was developed as a potential tool for data based modelling of biological data. In general, BRANNs are conventional backpropagation networks with an additional term added to the objective function and Bayesian-based optimization of the regularization parameters. The BRANN is not sensitive to overtraining and therefore is suitable for small data sets. In this study BRANN was used for prediction of the dose in gamma irradiated dairy products. The efficiency of the proposed BRANN model is proved by testing two dairy products – lyophilized cow milk and curd. The presented results have shown that the model is quite effective that has been proven at further by relevant statistical measures.

Acknowledgments: The research work reported in the paper is funded by DN 06/5 16.12.2016 project at the Fund for Scientific Research at the Ministry of Education, Youth and Science, Bulgaria.

References

1. Demuth, H. B., M. H. Beale, O. De Jess. M. T. Hagan. Oklahoma State University, Stillwater, Oklahoma, Neural Network Design, 2014.
2. Burden, F., D. Winkler. Bayesian Regularization of Neural Networks. – In Artificial Neural Networks, Humana Press, 2008, pp. 23-42.
3. Topuz, A. Predicting Moisture Content of Agricultural Products Using Artificial Neural Networks. – Advances in Engineering Software, Vol. **41**, 2010, Issue 3, pp. 464-470.
4. Sun, Q., M. Zhang, A. S. Mujumdar. Recent Developments of Artificial Intelligence in Drying of Fresh Food: A Review. – Critical Reviews in Food Science and Nutrition, Vol. **59**, 2019, Issue 14, pp. 2258-2275.
5. Román, R. C., O. G. Hernández, U. A. Urtubia. Prediction of Problematic Wine Fermentations Using Artificial Neural Networks. – Bioprocess and Biosystems Engineering, Vol. **34**, 2011, Issue 9, pp.1057-1065.
6. Lamrini, B., G. Della Valle, I. C. Trelea, N. Perrot, G. Trystram. A New Method for Dynamic Modelling of Bread Dough Kneading Based on Artificial Neural Network. – Food Control, Vol. **26**, 2012, Issue 2, pp. 512-524.
7. Al-Mahasneh, M., M. Aljarrah, T. Rababah, M. Alu'datt. Application of Hybrid Neural Fuzzy System (ANFIS) in Food Processing and Technology. – Food Engineering Reviews, Vol. **8**, 2016, Issue 3, pp. 351-366.
8. Todorov, Y., M. Doneva, P. Metodieva, I. Nacheva. An Intelligent Approach to Formulate the Contents of Novel Functional Food. – In: Proc. of IEEE International Symposium on Innovations in Intelligent Systems and Applications (INISTA), IEEE, 2014, pp. 98-103.
9. Todorov, Y., I. Nacheva, P. Metodieva, M. Doneva, T. Tsvetkov. Soft Computing Applications in Food Technology. – Bulgarian Journal of Agricultural Science, Vol. **19**, 2013, Issue 3, pp. 503-507.
10. Xiong, Hui Yuan, Y. Barash, B. J. Frey. Bayesian Prediction of Tissue-Regulated Splicing Using RNA Sequence and Cellular Context. – Bioinformatics, Vol. **27**, 2011, Issue 18, pp. 2554-2562.
11. Hajmeer, M. N., I. A. Basheer. A Hybrid Bayesian – Neural Network Approach for Probabilistic Modeling of Bacterial Growth/No-Growth Interface. – International Journal of Food Microbiology, Vol. **82**, 2003, Issue 3, pp. 233-243.
12. Corney, D. Designing Food with Bayesian Belief Networks. – In: I. C. Parmee, Ed. Evolutionary Design and Manufacture. London, Springer, 2000, pp. 83-94.
13. Barker, G. C. Application of Bayesian Belief Network Models to Food-Safety Science. – In: Frontis. 2004, pp. 117-128.
14. Forsee, F. D., M. T. Hagan. Gauss-Newton Approximation to Bayesian Learning. – In: Proc. of IEEE International Conference on Neural Networks (ICNN'97), Vol. **3**, 1997, pp. 1930-1935.
15. Harder, M. N. C., V. Arthur, L. N. C. Harder. Gamma Radiation Effect on Allergenic Food. – In: Allergen. London, UK, IntechOpen, 2017, pp. 139-150.
16. Odueke, O., K. Farag, R. Baines, S. Chadd. Irradiation Applications in Dairy Products: A Review. – Food and Bioprocess Technology, Vol. **9**, 2016, Issue 5, pp. 751-767.
17. Lee, J. W., J. H. Kim, H. S. Yook, K. O. Kang, S. Y. Lee, H. J. Hwang, M. W. Byun. Effects of Gamma Radiation on the Allergenic and Antigenic Properties of Milk Proteins. – J. of Food Protection, Vol. **64**, 2001, Issue 2, pp. 272-276.

Received: 21.11.2019; Second Version: 08.04.2020; Third Version: 11.05.2020; Accepted: 21.05.2020