Prediction and Sensitivity Analysis by TS Fuzzy Neural Network for Fungal Growth in Food Products

Yu-Hao Chang*, Wen-Hsien Ho*, Hon-Yi Shi*, Jyh-Horng Chou**;***
* Department of Healthcare Administration and Medical Informatics, Kaohsiung Medical University, Kaohsiung, Taiwan, {yu101572012, whho, hshi}@kmu.edu.tw
** Institute of System Information and Control, National Kaohsiung First University of Science and Technology, Kaohsiung, Taiwan, choujh@nkfust.edu.tw
*** Department of Electrical Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan, choujh@cc.kuas.edu.tw

Abstract—A TS fuzzy neural network (TSFNN) was compared with an artificial neural network (ANN) in terms of accuracy in predicting the combined effects of temperature, pH level, sodium chloride level and sodium nitrite level on the growth rate of Leuconostoc mesenteroides. The TSFNN and ANN models were compared in terms of six statistical indices calculated by comparing their prediction results with actual data. The learning-based systems obtained encouraging prediction results. Sensitivity analyses of the four environmental factors showed that temperature and, to a lesser extent, NaCl had the most influence on accuracy in predicting the growth rate of Leuconostoc mesenteroides. The observed effectiveness of TSFNN for modeling microbial kinetic parameters confirms its potential use as a supplemental tool in predictive mycology. Comparisons of the six statistical indices showed that the TSFNN model was better than ANN model in predicting the four kinetic parameters.

I. INTRODUCTION

In recent studies of shelf life in food products, microbiologists have used predictive models to forecast spoilage caused by the growth of micro-organisms. Despite the major technological advances in the food industry in recent years, fungal spoilage of food commodities remains a major cause of economic losses for food producers and an important health concern for regulatory agencies. Therefore, improved understanding of fungal growth in foods, particularly those factors associated with new manufacturing processing and packaging techniques, is urgently needed [1]. Improvements in food quality and safety require the development of appropriate fungal growth prediction tools. For many years, research in predictive microbiology has focused on food-borne pathogens whereas models for predicting growth in filamentous fungi have received relatively less attention [2]. Recently, however, the situation has changed, and the literature now shows a growing number of studies of models for this purpose [3-5].

Leuconostoc mesenteroides is a common spoilage microorganism in cooked meat products. These bacteria can alter food products by fermentation of sugars, which forms lactic acid. An effective tool for predicting shelf life would help to reduce economic losses from deterioration of food. A study by Garcia-Gimeno et al. [6] showed that an artificial neural network (ANN) model was more accurate than response surface methodology (RSM) for predicting Leuconostoc mesenteroides growth given similar environmental conditions.

For describing relationships between different combinations of inputs and outputs such as those that must be determined for accurately predicting growth in spoilage microorganisms, ANN is currently the most widely used technique. A recent literature review shows that the use of TS fuzzy neural network (TSFNN) [7, 8] for such purposes is relatively rare. Since the membership functions in the resulting fuzzy inference system were iteratively adjustable according to a given training set of inputs and outputs, the TSFNN could effectively map input–output relationships according to both human knowledge (in the form of fuzzy if–then rules) and stipulated input–output data pairs [9, 10]. In the current study, TSFNN was used to model the relationship between predicted and actual Leuconostoc mesenteroides growth rates under various conditions. Therefore, this study evaluated the accuracy of TSFNN in predicting Leuconostoc mesenteroides growth rates. The TSFNN and ANN models were then compared in terms of their accuracy in predicting Leuconostoc mesenteroides growth under varying experimental conditions, including temperature, pH, salt, nitrite concentrations. Some experimental results obtained by TSFNN method were also compared with those obtained by ANN methods in an earlier study by Garcia-Gimeno et al. [6]. Finally, sensitivity analyses were performed to identify the environmental factors that had the largest effects on the accuracy of the predictions of Leuconostoc mesenteroides growth rate.

II. METHODS

A. TSFNN Architecture

The TSFNN multilayer feed-forward network of nodes and directional links combines the learning capabilities of an ANN with the reasoning capabilities of fuzzy logic. The ANNs and fuzzy inference systems (FISs) are complementary technologies in the design of adaptive
intelligent systems. The ANNs, which learn from scratch by adjusting the interconnections among neurons, are noted for their generalization capability. That is, a properly trained ANN can correctly match a set of new input data to output data. The FIS is a popular computing framework based on fuzzy set theory, fuzzy if–then rules, and fuzzy reasoning. For a given set of if-then rules, an FIS can perform nonlinear mapping from its input space to its output space [7-10].

The TSFNN uses fuzzy if–then rules involving premise and consequent parts of a Sugeno-type fuzzy inference system [7]. Figure 1 shows how the description of this system can be simplified as an inference system of inputs m and n and output f [7]. The corresponding TSFNN architecture is also shown. The five-layer system TSFNN architecture includes a fuzzification layer (Layer 1), a production layer (Layer 2), a normalization layer (Layer 3), a de-fuzzification layer (Layer 4), and a total output layer (Layer 5) [7].

B. Sensitivity Analysis of TSFNN Output

Sensitivity analysis was performed with TSFNN learning disabled so that network weights would not be affected [27]. The first input varies between its mean plus or minus a user-defined number of standard deviations whereas all other inputs are fixed at their respective means. The TSFNN output is computed and recorded as the percent change above and below the mean channel output. This process is repeated for each input variable [11].

In this so-called ‘sensitivity’ analysis to determine the relative importance of input variables for determining output, 0 indicates a variable that does not affect prediction, and 1.0 indicates a field that completely dominates the prediction [11].

III. Results

First, the TSFNN was trained using 30 data sets [6] selected from the 58 data sets obtained in the experiments. After training was completed, another 28 data sets were then used to verify its accuracy in predicting growth rates [6]. The prediction results are further analyzed and discussed below.

Figure 2 shows the fuzzy rule architecture of an TSFNN with Gaussian membership function. The four inputs (temperature, pH, NaCl, and NaNO2) and one output (growth rate) of the TSFNN model were designed using MATLAB Fuzzy Logic Toolbox [12]. Each input divides three Gaussian membership functions. Therefore, the architecture in Figure 2 includes 81 fuzzy rules. The TSFNN was trained using 30 sets of experimental data in 100 learning cycles. For each input in the architecture, the Gaussian membership function can be divided into small, medium and large areas.

Figures 3 and 4 show the prediction results for the training and testing data sets obtained by the TSFNN model, respectively. The mean absolute percentage errors (MAPEs) obtained are 0.18 for the training data set and 0.27 for the testing data set. The root mean square errors (RMSEs) obtained are 0.001 for the training data set and 0.002 for the testing data set. The standard error of prediction percentages (SEPs) obtained are 0.58 for the training data set and 0.67 for the testing data set.

<table>
<thead>
<tr>
<th>Statistical index</th>
<th>Model</th>
<th>Data set</th>
<th>Training</th>
<th>Testing</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean absolute percentage error (%) (MAPE)</td>
<td>ANN</td>
<td>3.06</td>
<td>15.69</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>0.18</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Root mean square error (RMSE)</td>
<td>ANN</td>
<td>0.019</td>
<td>0.067</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>0.001</td>
<td>0.002</td>
<td></td>
</tr>
<tr>
<td>Standard error of prediction percentage (%) (SEP)</td>
<td>ANN</td>
<td>8.45</td>
<td>23.27</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>0.58</td>
<td>0.67</td>
<td></td>
</tr>
<tr>
<td>Bias factor (Bf)</td>
<td>ANN</td>
<td>0.99</td>
<td>0.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Accuracy factor (Af)</td>
<td>ANN</td>
<td>1.03</td>
<td>1.17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>Absolute fraction of variance ($R^2$)</td>
<td>ANN</td>
<td>0.9937</td>
<td>0.9539</td>
<td></td>
</tr>
<tr>
<td></td>
<td>TSFNN</td>
<td>1.0000</td>
<td>1.0000</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Framework of TS fuzzy neural network.

Figure 2. Fuzzy rule architecture of the Gaussian membership function.

Table 1. Comparison of performance indices between ANN and TSFNN models.
Figure 3. Comparison of actual growth rates for *Leuconostoc mesenteroides* and growth rates predicted by TSFNN model and by ANN model.

Figure 4. Comparison of residual values (predicted–observed growth rate) of *Leuconostoc mesenteroides* obtained by TSFNN model and by ANN model.

Comparisons of performance indices for predictions of the 30 patterns on which the models were trained showed that the TSFNN was more accurate than the ANN. Specifically, the TSFNN model was superior in terms of MAPE, RMSE and SEP (Table 1). For the 28 testing data sets, comparisons of model performance again confirmed the superior performance of the TSFNN model, considering the biological variability associated with the experiment. Additionally, measurements of bias and accuracy in the TSFNN model approached unity in all experiments, which indicated good agreement between observations and predictions. The performance difference between the two examined models was also graphically depicted in plots of bias (observed vs. predicted growth rate) and residual bias for all data sets. Generally, the predictions derived by TSFNN model was closer to the line of equity compared to the ANN model (Figure 3), which indicated the better fit of the TSFNN model. In both the TSFNN and ANN models, residuals were also symmetrically distributed around 0 with no systematic tendency to appear on the positive or negative sides of the graph (Figure 4). The narrower spread of residual values obtained by the TSFNN model in all data sets also indicated their superior prediction performance.

In the TSFNN model, the sensitivity values for temperature, NaCl, pH and NaCO$_2$ were 0.88, 0.13, 0.07 and 0.04. In terms of effect on the growth rate of *Leuconostoc mesenteroides*, the most influential (sensitive) parameters in both the training and testing data sets were temperature and, to a lesser extent, NaCl.

IV. DISCUSSION

Table 1 compares the statistical data for the two models, and Figures 3 and 4 compare graphical plots of the data. Compared to the ANN model, the TSFNN model showed better agreement with the experimental observations in both the training and testing data sets. The better data fit obtained by the TSFNN -based model in the comparisons with experimental data was confirmed by its Absolute fraction of variance ($R^2$) value of 1.0000 (versus 0.9539–0.9937 in the ANN model).

Notably, although $R^2$ is a common criterion for comparing statistical models [12], it assumes a normally distributed error that is independent of the mean value. However, since error distributions are unknown when predicting microbial growth, this term must be used cautiously, particularly in nonlinear regression models [13]. Hence, RMSE values were used for further comparisons of model performance. The TSFNN model showed better RMSE values (0.001 for the training data set and 0.002 for the testing data set) compared to the ANN model (0.019 for training data set and 0.067 for the testing data set) (Table 1). The RMSE, which is calculated by comparing desired and actual output values, which are then averaged across all data, index provides an estimate of goodness of fit in statistical models and indicates the long-term consistency of a model [13]. The lower RMSE values in the TSFNN model in comparison with the ANN model confirmed its superior prediction performance.

Like accuracy factor (Af), MAPE indicates the average deviation from the observed value [14]. In this study, the MAPE results were in good agreement with the Af values estimated for all data sets. In the training data set, for instance, the Af values obtained by the ANN and TSFNN models were 1.03 and 1.00, respectively (Table 1), and their average deviations in predicted and actual growth rates were 3% and 0%, respectively. These values were highly consistent with the MAPE values of 3.06% and 0.18% obtained by the ANN and TSFNN models, respectively, in the training data set. The ANN and TSFNN models obtained Af values of 1.17 and 1.00, respectively (Table 1), and their average deviations in predicted and actual growth rates were 17% and 0%, respectively. Again, these values were closely approximated the MAPE values of 15.69% and 0.27% obtained by the ANN and TSFNN models, respectively, in the testing data set. The relevant figures again confirmed the better performance of the TSFNN model.

The accuracy factor is similar to the bias factor (Bf) statistic, which was also introduced by Ross [14]. In this case, a Bf value greater than 1 indicates that the model overestimates growth rate and is thus a ‘fail-dangerous’ model whereas a value less than 1 indicates that the model underestimates growth rate and is thus a ‘fail-safe’ model. For example, the accuracy of the growth rate estimates was confirmed by Bf values of 1.00 in the
TSFNN model and by $B_r$ values of 0.99 in the ANN model (Table 1). These data indicate that the TSFNN model provides a relatively more accurate estimate of growth rate compared to the ANN model, which tends to underestimate growth rate.

The SEP index indicates the relative deviation in mean prediction values. A notable advantage of the index is that its calculation is independent of the magnitude of the measurements [6]. For all data sets (Table 1), the SEP indices were again better in the TSFNN model (range, 0.58-0.67) than in the ANN model (range, 8.45-23.27).

The TSFNN and ANN models offer an interesting option for defining the sensitivity and relative importance of inputs. The sensitivity analysis in this experiment revealed the important effects of temperature on Leuconostoc mesenteroides growth rates. In earlier works, Zurera-Cosano et al. [14] obtained similar results in RSM models of Leuconostoc mesenteroides growth whereas Panagou and Kodogiannis [5] reported similar results in an ANN model of Monascus ruber growth.

V. CONCLUSION

This study confirmed that, compared to ANN model, TSFNN architectures provide better accuracy in predicting the growth rate of Leuconostoc mesenteroides based on input data for temperature, pH, NaCl and NaNO2. The statistical indices and graphic plots confirmed the superior performance of the TSFNN model in both training and testing data sets. Sensitivity analyses of the TSFNN model revealed that the most influential (sensitive) factors in the growth rate of Leuconostoc mesenteroides was temperature and, to a lesser extent, NaCl. As TSFNN is mainly applied in predictive microbiology, the findings that TSFNN model is effective for predicting the kinetic parameters of fungi indicate their good potential use as an alternative to ANNs in this field.

ACKNOWLEDGMENT

This work was supported by grant NSC 101-2320-B-037-022 from the National Science Council, Taiwan, Republic of China.

REFERENCES