

# Development of a Secondary Model for the Growth of *Salmonella enterica* in Food by Applying Artificial Neural Networks and Databases (ComBase and FoodData Central)

Yong Keun Koo, Yong Woon Jung, Dong Hwa Kim, Sang Won Kim, Eun Seol Kim,  
Byeong Jae Park, Seung Ju Lee, and Seung Won Jung\*

Department of Food Science and Biotechnology, Dongguk University

## Abstract

The secondary growth model for *Salmonella* was developed based on the artificial neural network (ANN) with data collected from ComBase and FoodData Central. In addition to the existing secondary model variables (temperature, pH, Na<sup>+</sup>, and water contents), more input variables (sugar, carbohydrate, lipid, and protein contents) were considered. The output variables were microbial growth parameters (lag phase duration [ $\lambda$ ] and maximum growth rate [ $\mu_{\max}$ ]). A commercial ANN program (NeuralWorks Predict) was utilized with training at 80%, validation at 10%, and test data at 10%. ANN models were created using all data and cleansed data. Using the cleansed data, the training/testing root mean square error (RMSE) for  $\mu_{\max}$  improved from 0.14/0.16 to 0.11/0.14, whereas the RMSE for  $\lambda$  was still not acceptable, from 11.94/33.03 to 7.09/4.18. The  $\lambda$  data were divided into two ranges with high and low goodness of fit, whereas the ANN model for each field was built, resulting in an optimally low RMSE.

**Keywords:** artificial neural network, microbial growth, secondary model, *Salmonella*, ComBase

## Introduction

The microbial growth prediction models are categorized into a primary model indicating microbial behavior versus time, a secondary model describing a kinetic parameter versus environmental factors, and a tertiary model that predicts a growth curve under specific environmental conditions by combining the primary and secondary models (Stavropoulou & Bezirtzoglou, 2019). Conventionally, the secondary models have considered temperature, pH, water activity ( $a_w$ ), and salt (Na<sup>+</sup>) (Stavropoulou & Bezirtzoglou, 2019). Quite a few secondary models have been reported. A hyperbola model and an extended hyperbola model were developed to predict the lag phase duration ( $\lambda$ ) as a function of temperature (Boekel, 2008). To predict the maximum growth rate ( $\mu_{\max}$ ) versus temperature, the Ratkowsky square root model and the Zwietering equation were reported (Boekel, 2008). These models describe microbial growth parameters ( $\lambda$ ,  $\mu_{\max}$ ) with respect to one variable. The gamma concept-based model and the cardinal parameter model could predict  $\mu_{\max}$  by considering pH and  $a_w$  as well as

temperature (Boekel, 2008).

However, food components, such as sugars, carbohydrates, lipids, and proteins, also influence growth behavior. Factors affecting the growth of microorganisms are classified as external (e.g., temperature) and internal (e.g., pH, water, salt, sugar, carbohydrate, protein, and fat contents of food) (Busta et al., 2003). Sugars are the most fundamental nutrient for microbial growth because most microorganisms can metabolize the saccharides, such as glucose, to produce energy (Busta et al., 2003). Sugars are also an important factor affecting the growth of microorganisms because if the sugar content is high, the  $a_w$  is lowered and sometimes interferes with the growth of microorganisms. Most foodborne pathogens will first metabolize simple carbohydrates and amino acids, followed by the rest of the nutrients (Busta et al., 2003). These nutrients become vital energy, carbon, and nitrogen resources to produce cellular components through enzyme activity and biochemical reactions for the growth of microorganisms (Busta et al., 2003). These core nutrient contents can determine microbial growth rate, especially at low temperatures, making them more nutrient-dependent, and a low nutrient content affects microbial growth at optimum temperatures by limiting microbial growth (Erkmen and Bozoglu, 2016). Food nutrients also affect microbial death. A high-fat content provides some protection against the death of *Salmonella* species at the killing temperature (Finn et al., 2013). As such, the effect of food components on microorganisms

\*Corresponding author: Seung Won Jung, Department of Food Science and Biotechnology, Dongguk University, 32, Dongguk-ro, Ilsandong-gu, Goyang-si, Gyeonggi-do 10326, Korea  
Tel: +82-31-961-5654; Fax: +82-31-961-5655  
E-mail : [signalco@dongguk.edu](mailto:signalco@dongguk.edu)  
Received November 27, 2023; revised December 12, 2023; accepted December 12, 2023

cannot be ignored (Boekel, 2008). Therefore, it would be desirable to consider the above factors (sugar, carbohydrate, protein, and lipid contents) as well as the conventional ones (temperature, pH,  $a_w$ , and  $\text{Na}^+$ ) in the growth models.

The Pathogen Modeling Program (PMP) is a tertiary model developed by the U.S. Department of Agriculture Agricultural Research Service (USDA ARS) to predict microbial growth based primarily on a microbial culture medium rather than food components. However, as more and more influencing variables are included to encompass the complex and variable nature of food, the food environment, and the food microbiome, it becomes increasingly challenging to predict bacterial behavior (Hiura et al., 2021). To overcome the limitations of existing microbial prediction models, artificial neural network (ANN) techniques were introduced (Hajmeer et al., 1997; Oscar, 2009; Oscar, 2021; Hiura et al., 2021).

Two main advantages of ANN are that their application does not require a priori knowledge of the process, and they are effective for predicting networks with multiple input and output factors of discrete nonlinear data (Bourquin et al., 1998). Stangierski et al. (2019) developed an ANN that used five variables to model the overall quality of Gouda cheese during storage and showed that the model had better prediction performance than a multiple regression model.

To build an ANN, it is critical to use many high-quality data. Hiura et al. (2021) developed the eXtreme gradient boosting tree (XGBoost)-ANN for predicting the growth and survival of *Listeria monocytogenes* in food by applying machine learning classification using ComBase-based viable count data. ComBase is a formatted database (DB) containing a large amount of data on microbial responses to environmental conditions. Eight input variables selected from ComBase that included numerical data types (temperature, pH,  $a_w$ , time, and initial cell number) and categorical data types (food category, food name, and whether initial cell number or not) were used to develop the XGBoost-ANN model. This research demonstrated that the accumulated data in a DB could be useful for predicting bacterial population behavior (Hiura et al., 2021). It is expected that the predictions will be further improved if numerical data of food component contents, rather than categorical data for food, are used directly in ANN modeling.

The data required to develop a microorganism prediction model using ANN can be collected from online DBs, such as ComBase, FoodData Central maintained by the USDA, and the Food Composition Database (FCDB) maintained by the Korea Ministry of Food and Drug Safety (MFDS). These DBs provide information on the content of components,

such as moisture, carbohydrate, protein, fat, and minerals, in agricultural and livestock products, aquatic products, and processed foods (MFDS, 2020; USDA, 2020).

In this study, a secondary model for predicting the growth parameters of *Salmonella enterica* was developed using ANN. Dependent variables were  $\lambda$  and  $\mu_{\max}$ . More independent variables (sugar, carbohydrate, lipid, and protein contents of food) were considered in addition to the conventional secondary model variables (temperature, pH, and  $\text{Na}^+$  and water contents). All data were retrieved from ComBase and FoodData Central and cleansed to increase the ANN model prediction accuracy. The ANN models were built, and their prediction accuracies were evaluated.

## Materials and Methods

### Data collection

The *Salmonella* growth parameter data were collected from ComBase (<https://www.combase.cc>), and food components data were obtained from FoodData Central (<https://fdc.nal.usda.gov/index.html>). In ComBase, the microbial growth environments are either foods or microbial media. For this study, only the growth data in food were collected. The growth data were provided in the forms of growth parameters ( $\lambda$  and  $\mu_{\max}$ ) or growth curves. For the growth curves, growth parameters were estimated using ComBase's online DMFit fitting tool (<https://browser.combase.cc/DMFit.aspx>) because the input variables in ANN modeling are growth parameters, not growth curves. Online DMFit is a first-order model-based fitting tool, such as the Baranyi and Roberts model or the trilinear model. The growth curves in the form of images were digitized using a plot digitizer 2.6.9 software. The food component data corresponding to microbial growth data in ComBase were collected from FoodData Central.

### Data cleansing

To improve the prediction accuracy of ANN, data cleansing was performed to remove data with missing values and outliers among data of output variables (Ridzuan et al., 2019). Data that included missing values were removed first. Data containing outliers were removed by selecting the values determined arbitrarily during the data collection process and those with a z-score greater than 3.0 and less than -3.0. z-score was calculated as follows:

$$Z_i = \frac{x_i - \mu}{\sigma}$$

where  $Z_i$  is the z-score of  $x_i$ ,  $x_i$  is the output value of a single data, and  $\mu$  and  $\sigma$  are the mean and standard deviation of all data, respectively.

### ANN modeling

The ANN model was built using NeuralWorks Predict (NeuralWare, Carnegie, PA, USA), an Excel add-on program. The ANN was based on a backpropagation algorithm, which performs learning on a multilayer feed-forward neural network (multilayer perceptron). Dependent variables were set as the input variables for ANN: temperature ( $^{\circ}\text{C}$ ), water content (%), pH, and  $\text{Na}^+$ , sugar, carbohydrate, lipid, and protein contents (%) of food. Independent variables were set as the output variables:  $\lambda$  (h) and  $\mu_{\max}$  (1/h). Data were split into training (80%), validation (10%), and test (10%) data. Using the program, level options should be chosen for “noise level”, “data transformation level”, “variable selection level”, and “neural network search level”. By adjusting each level, the criteria for

ANN learning and construction can be changed selectively. To determine the optimum levels, the root mean square error (RMSE) values were compared. RMSE values were calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (R_i - P_i)^2}{n}}$$

where  $n$  is the number of data,  $R_i$  is the actual value of a single data, and  $P_i$  is the predicted value of a single data.

### Statistical analysis

The z-score and RMSE were used for data cleansing and prediction accuracy, respectively.

## Results and Discussion

### Data for ANN

ANN modeling has the advantage of excellent predictability

<i>Salmonella</i> growth data mining
ComBase: <ul style="list-style-type: none"> <li>- 158 tables and 117 graphs (growth parameters (<math>\mu_{\max}</math> and <math>\lambda</math>) vs. temperature)</li> <li>- 5 tables and 28 graphs (microbial growth concentration vs. time)</li> </ul>
Data classification for food
Food category (number of cases): Food name <ul style="list-style-type: none"> <li>- Chicken (193): Chicken breast (raw), chicken tenderloin (raw), chicken breast (autoclaved), and chicken thigh (autoclaved)</li> <li>- Beef (70): Ground beef (95% lean), ground beef (90% lean, 10% fat), minced beef, beef tissue (lean), beef tissue (fatty), and beef rip eye</li> <li>- Vegetables (9): Iceberg lettuce and mung bean sprout</li> <li>- Eggs (22): Egg yolk (raw), egg yolk (sugared, frozen, and pasteurized), and egg whole (raw, frozen, and pasteurized)</li> <li>- Milk (4): Whole milk (low fat milk 1:1)</li> <li>- Cheese (9): Queso fresco type soft cheese and brie cheese</li> <li>- Pork (1): Bacon (back quarter and rindless)</li> </ul>
Food component data mining
FoodData Central: <ul style="list-style-type: none"> <li>- Chemical compositions (carbohydrate, lipid, protein, sugar, <math>\text{Na}^+</math>, moisture contents) of the above foods</li> </ul>

Fig. 1. Data preparation for *Salmonella* growth and food compositions provided by public databases.

because it can include many variables. In this study, the ANN has a feature that includes food components. Two main DBs were available to access *Salmonella* growth data online: ComBase and MediaDB (<https://mediadb.systemsbio.net>, System Biology) (Fig. 1). According to environment variables, *Salmonella* data were 8,463 in ComBase, including research paper and agency-funded data, and 70 in MediaDB. ComBase includes both food and microbial medium environments for microbial growth, whereas MediaDB contains only the microbial medium environment. In this study, only food-based data were initially collected from ComBase because *Salmonella* growth in microbial media cannot represent the growth in foods with different components and where bacteria other than *Salmonella* may also be present. Then, only data from research papers were selected because the papers provided detailed information on target food and experimental conditions, unlike agency-funded data. The total number of collected data sets was 308 (Fig. 1). The form of the microbial growth data varied: 158 tables and 117 graphs showed microbial growth parameters ( $\lambda$  and  $\mu_{\max}$ ) versus temperature, and 5 tables and 28 graphs showed microbial growth curve data versus time. Data from 33 microbial growth curves were regressed with DMFit using the Baranyi model to calculate growth parameters. In addition, the different units of growth parameters between the reference sources were unified. The  $\lambda$  were easily unified in time (h or day) in all data. Most of the  $\mu_{\max}$  were expressed in units of 1/h, but some in doubling time. In the case of  $\ln(N/N_0)/\text{time}$  instead of  $\log_{10}(N/N_0)/\text{time}$ ,  $\mu_{\max}$  was divided by 2.303 due to,  $\log_{10}(x) = \frac{\ln(x)}{2.303}$ . When the growth rate was expressed as 1/doubling time, it was multiplied by  $\log_{10}(2)$  due to  $\log_{10}(2) = \mu_{\max} \times [\text{doubling time}]$ .

In ComBase, *Salmonella* strains were very diverse, such as *Salmonella* spp., including *S. enterica* serovars Typhimurium, Enteritidis, Newington, and Derby. All these were used as data for ANN modeling. According to Lamas et al. (2018), 99% of human salmonellosis is caused by more than 1,500 serotypes of *S. enterica* subsp. *enterica*. Oscar (2000) compared the growth rates of 11 strains of *S. enterica* in sterile ground chicken breast burgers. The coefficient of variation (standard deviation/mean) was 9.4% and 5.7% for  $\lambda$  and growth rate, respectively. This indicates that there is no problem using the data of various serotypes of *Salmonella* in ANN modeling.

ComBase contains information about food category, food name, temperature, pH,  $\text{Na}^+$ ,  $a_w$ , and reference source. The target food categories for *Salmonella* growth were chicken (193 cases), beef (70 cases), eggs (22 cases), cheese (9 cases),

vegetables (9 cases), milk (4 cases), and pork (1 case). Food component contents for the food category and food name (e.g., chicken and raw chicken breast) were obtained from FoodData Central. Temperature and pH provided by ComBase were directly used.  $\text{Na}^+$  content and  $a_w$  were missing in ComBase in many cases.  $\text{Na}^+$  content was obtained from FoodData Central, and  $a_w$  was replaced with water content. The  $a_w$  is associated with water, salt, and sugars (Ziegler et al., 1987). As the  $\text{Na}^+$  content was considered and sugars were also included in the food components, it is deemed that there would be no major problem even if water content was used instead of  $a_w$ .

### Data cleansing

The data cleansing process for missing values and outliers followed the option selection to improve the ANN prediction accuracy. If it belonged to missing values or outliers regardless of  $\lambda$  and  $\mu_{\max}$ , the case was excluded because a case is composed of a set of  $\lambda$  and  $\mu_{\max}$ .  $\mu_{\max}$  had no missing values, and I had 44 missing values out of 308. An outlier was judged in two ways. First, the outliers were the cases where  $\lambda$  seemed to be ignored as zero because the measurement interval was day. Four out of 264 cases ( $= 308 - 44$ ) were outliers of this type. Second, the case where either the z-score for  $\lambda$  or  $\mu_{\max}$  was beyond  $\pm 3.0$  was an outlier (Shiffler, 1988; Smiti, 2020). Five out of 260 ( $= 264 - 4$ ) cases for  $\lambda$  were outliers of this type, whereas the  $\mu_{\max}$  had no outliers. Finally, 53 ( $= 44 + 4 + 5$ ) data with missing values and outliers were excluded, and 255 data ( $= 260 - 5$ ) were used for ANN modeling with cleansed data.

### ANN model

The ANN modeling procedure is summarized in Fig. 2. The user's option of NeuralWorks Predict was 400:  $4 \times 4 \times 5 \times 5$  levels for noise level (option A), data transformation level (option B), variable selection level (option C), and neural network search level (option D), respectively. Option C was fixed to "no variable selection", and option D considered only 4 levels except for "no network search". By selecting "no variable selection", the undesirable case of automatically selecting all or only some of the input variables was excluded. "No network search" was excluded because if this option is selected, a neural net of algorithms will not be trained, and only the result of the variable selection is presented. Therefore, the number of possible final option cases was 64 ( $4 \times 4 \times 1 \times 4$ ). To select the optimal option, an ANN model was built for each option using all data of 308 data sets. The difference between the actual and predicted growth parameters from each ANN

Training set/validation set/test set (80:10:10 %)
308 data sets: - Independent variables: Temperature, pH, carbohydrate, lipid, protein, sugar, Na <sup>+</sup> , and moisture contents - Dependent variables: $\mu_{max}$ and $\lambda$
Optimization of user's options of the ANN program (Neuralworks Predict)
Option: - Levels of noise - Data transformation - Variable selection - Neural network search Level determination at the lowest RMSE
Data cleansing
Data cleansing: Delete missing value and outlier
ANN modeling
ANN 1 (308): All data
ANN 2 (255): Cleansed data

Fig. 2. ANN modeling procedure of *Salmonella* growth secondary model.

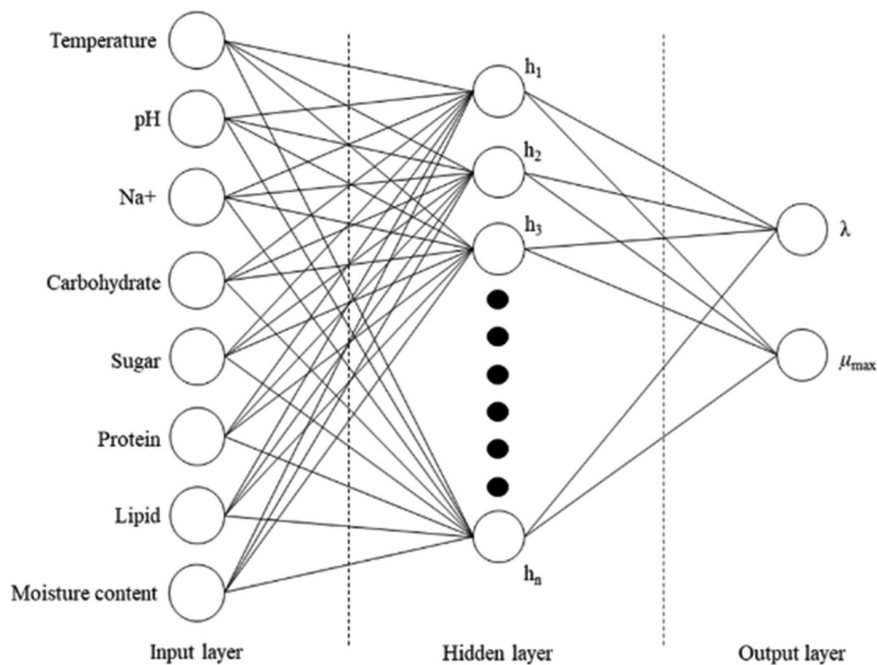


Fig. 3. Multilayer perceptron ANN (ANN 1, n = 13; ANN 2, n = 3).

model was evaluated in terms of RMSE. The lowest and highest RMSE ( $\lambda$  and  $\mu_{\max}$ ) and their sum were 13.760, 0.141, and 13.901 at the 2-2-1-4 level and 35.450, 0.125, and 35.575 at the 2-2-1-5 level for A, B, C, and D, respectively. Therefore, the option at the 2-2-1-4 level for A, B, C, and D was adopted in ANN modeling.

Two groups of all data and cleansed data were applied to build the ANN models. The number of data for ANN should generally be very large, but the data amount in this study was insufficient. The reason was that the data that could be referred from ComBase were limited due to the difficulty of testing food poisoning bacteria in food. The data were split into 80% training, 10% validation, and 10% test data. A training (including validation)/testing splitting ratio of 80:20 is commonly used, but there are other ratios, such as 70:30, 60:40, and 50:50 (Vrigazova, 2021; Joseph, 2022). In ANN modeling, when the number of data is small, it is common to reduce the number of training data to prevent overfitting. However, if the number of data is too small, there is a possibility of underfitting, so there is a limit to reducing it (López et al.,

2022). With the train validation split, the hyperparameters (the number of hidden layers, nodes, and epochs) are determined optimally (Yoo, 2019). Train/test split is a model validation process that allows to simulate how the model would perform with new data. The number of hidden layers and nodes of each ANN was automatically determined by the program (Fig. 3).

#### Accuracy of ANN model

The goodness of fit was described in terms of RMSE. The training/testing RMSE for  $\mu_{\max}$  were 0.14/0.16 and 0.11/0.14 for all data and cleansed data, respectively. This indicates that data cleansing improved the prediction accuracy (Fig. 4). Ölmez and Aran (2005) found that the RMSE was 0.27 in the  $\mu_{\max}$  prediction of *Bacillus cereus* using temperature (8, 15, 26, and 32°C), pH (5.3, 5.8, 6.3, 6.8, and 7.3), and concentrations of sodium lactate (0, 200, 400, and 600 mM) and sodium chloride (85, 342, and 600 mM). Lee et al. (2021) reported an RMSE of  $\mu_{\max}$  of 0.1-0.35 in the prediction model of *Salmonella* species growth in cakes. In the ANN-based predictive model for *Salmonella* growth developed in the current study, the RMSE

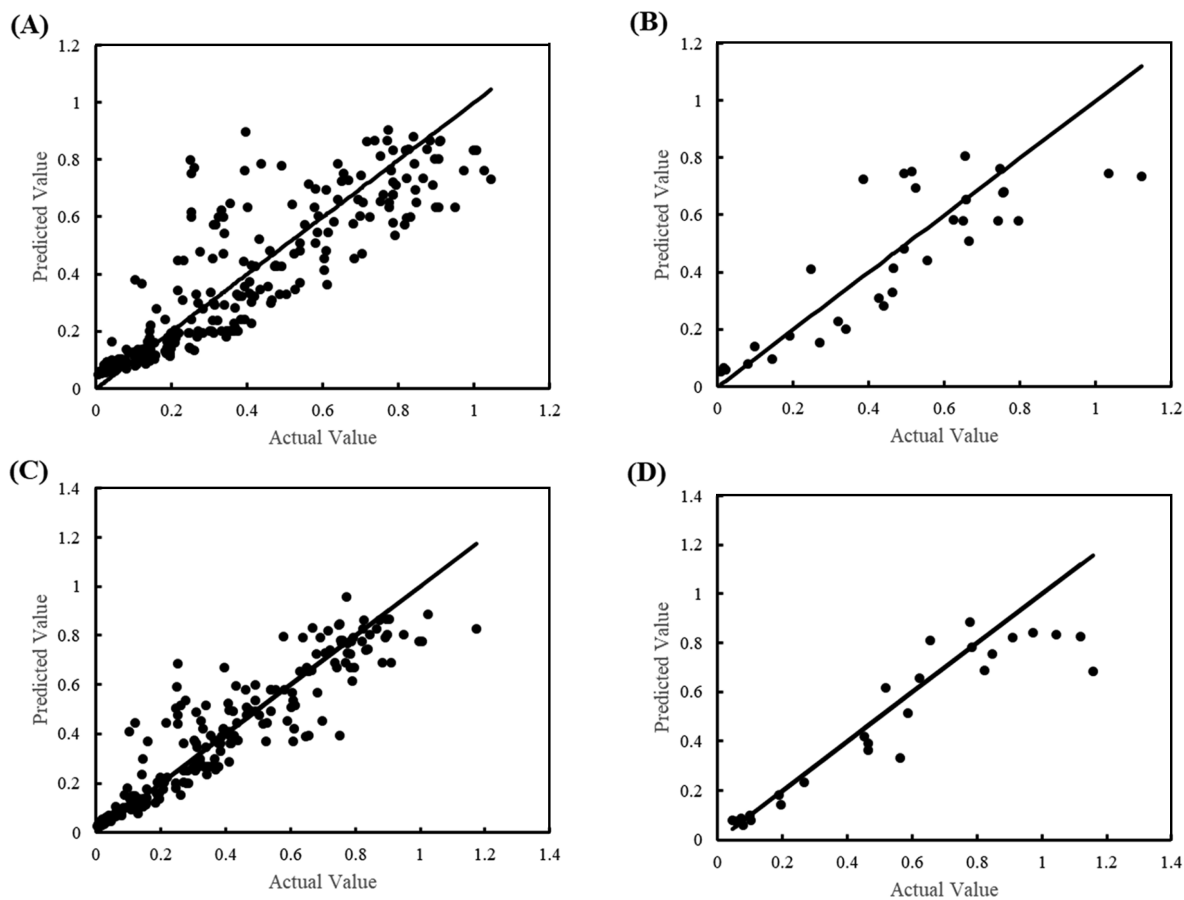
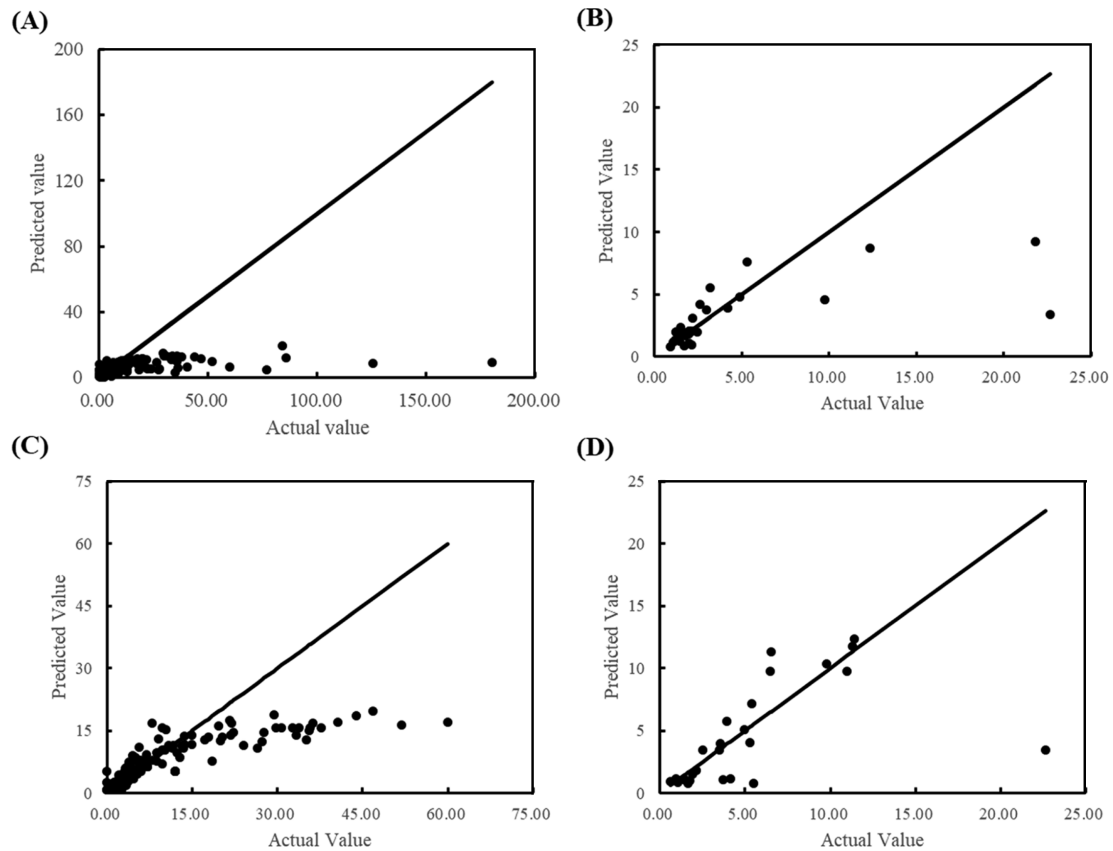


Fig. 4. Comparison of real  $\mu_{\max}$  and predicted  $\mu_{\max}$  based on ANN 1 and ANN 2. (A) The training set of ANN 1. (B) The test set of ANN 1. (C) The training set of ANN 2. (D) The test set of ANN 2.



**Fig. 5.** Comparison of real  $\lambda$  and predicted  $\lambda$  based on ANN 1 and ANN 2. (A) The training set of ANN 1. (B) The test set of ANN 1. (C) The training set of ANN 2. (D) The test set of ANN 2.

of  $\mu_{\max}$  ranged from 0.11 to 0.16, being good in predictive performance compared to the above references. However, there was a tendency for overfitting that although the RMSE of training data is small enough, that of test data is still high. Overfitting occurs when the ANN model is fitted even to noise data (López et al., 2022).

The training/test set RMSE of  $\lambda$  were 11.94/33.03 and 7.09/4.18 for all data and cleansed data, respectively. Although the prediction accuracy was improved by data cleansing, it still showed underfitting. Due to the nature of ANN, it is possible to increase the prediction accuracy for training data by further training with epoch number increased. However, with further training, an overfitting problem occurs, and the final model is completed by stopping training at the optimal number of epochs. In the case of  $\lambda$ , it was judged that the completed model could not be used as a predictive model as it is because the degree of fitting was low. Conversely, as shown in Fig. 5, the goodness of fit was high in a small range of  $\lambda$  but low in a range of larger  $\lambda$ . Focusing on this point, rather than one universal model, models for each range were developed by dividing it into two. As shown in Fig. 6, it was found that the

fitting with all cleansed data was improved by dividing it into two models. The RMSE of each model was 0.325 and 6.022 for the  $\lambda$  ranges from 0 to 2.60 h and 2.61 to 60 h, respectively.

The complexity of developing a model to predict  $\lambda$  is discussed as follows.  $\mu_{\max}$  is a reproducible parameter for the same environment and microorganism, whereas  $\lambda$  varies depending on the microbial cell history, even for food with the same environmental parameters (Ross et al., 2014).  $\mu_{\max}$  can be reproducibly obtained experimentally, but  $\lambda$  cannot (Baty et al., 2004). Because the log phase with  $\mu_{\max}$  on the microbial growth curve is after bacteria have already started to grow,  $\mu_{\max}$  can be obtained when bacteria are in a normal state. Conversely, because the lag phase with  $\lambda$  is before bacteria have reached a steady state, even the same bacteria may be affected by the age of bacteria or the history of the previous environment when testing the microbial growth curve. Therefore, the variation of  $\lambda$  seems inevitable, and thus  $\lambda$  might be difficult to accurately predict with one model.

In conclusion, secondary models of *Salmonella* growth in food were developed by using ANN. Only food data were selected among the microbial growth data in food and

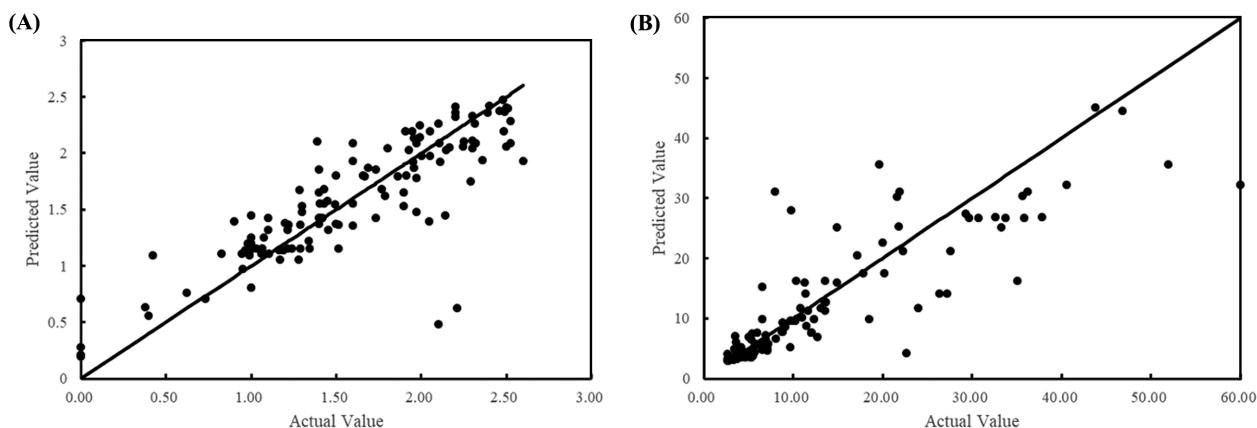


Fig. 6. Comparison of real  $\lambda$  and predicted  $\lambda$  based on ANN 2. (A) The group with  $\lambda$  of 0 to 2.60. (B) The group with  $\lambda$  of 2.61 to 6.

microbial medium from ComBase. According to the advantage of ANN, which can include many independent variables, data on food components (sugar, carbohydrate, lipid, and protein contents) were added to the independent variables in addition to the conventional variables (temperature, pH,  $\text{Na}^+$ , and water contents). In ComBase, for  $\mu_{\max}$  and  $\lambda$ , which are dependent variables, the data are presented not only as a number but also as a microbial growth curve, and in the latter case, the numbers were obtained through regression analysis. The prediction accuracy of the ANN model was high in the case of  $\mu_{\max}$  but very low in the case of  $\lambda$ . In the latter case, the prediction could be improved by dividing  $\lambda$  into two ranges and building an ANN model for each. In this study, the total number of *Salmonella* growth data obtained from ComBase to date was rather insufficient for use in ANN. Continuous development of the ANN models is required by applying more data as it is accumulated in the future.

### Acknowledgment

This work was supported by a grant (21163MFDS501) from the Ministry of Food and Drug Safety (MFDS), funded by the Korean government in 2022; Korea Institute of Planning and Evaluation for Technology in food, Agriculture, Forestry (IPET) through High value-added Food Technology Development Program, funded by ministry of Agriculture, Food and Rural Affairs (MAFRA) (121017-03).

### References

Baty F, Delignette-Muller ML. 2004. Estimating the bacterial lag time: which model, which precision?. *Int. J. Food Microbiol.* 91: 261-277.

- Boekel MAJS van. 2008. *Kinetic Modeling of Reactions in Foods*. 1st ed. CRC Press, Boca Raton, Florida, USA.
- Bourquin J, Schmidli H, Van HP, Leuenberger, H. 1998. Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenic study on a solid dosage form. *Eur. J. Pharm. Sci.* 7: 5-16.
- Busta FF, Bernard DT, Gravani RB, Hall P, Pierson MD, Prince G, Schaffner DW, Swanson KMJ, Woodward B, Yiannas F. 2003. Evaluation and definition of potentially hazardous foods. *CRFSFS*. 2: 1-109.
- Erkmen O, Bozoglu TF. 2016. *Food Microbiology: Principles into Practice*. 1st ed. John Wiley & Sons, Chichester, West Sussex, UK, pp. 91-106.
- Finn S, Condell O, McClure P, Amézquita A, Fanning S. 2013. Mechanisms of survival, responses, and sources of *Salmonella* in low-moisture environments. *Front. Microbiol.* 4: 331.
- Hajmeer MN, Basheer IA, Najjar YM. 1997. Computational neural networks for predictive microbiology II. Application to microbial growth. *Int. J. Food Microbiol.* 34: 51-66.
- Hiura S, Koseki S, Koyama K. 2021. Prediction of population behavior of *Listeria monocytogenes* in food using machine learning and a microbial growth and survival database. *Sci. Rep.* 11: 10613.
- Joseph VR. 2022. Optimal ratio for data splitting. *Stat. Anal. Data Min.* 15: 531-538.
- Lamas A, Miranda JM, Regal P, Vázquez B, Franco CM, Cepeda AA. 2018. Comprehensive review of non-enterica subspecies of *Salmonella enterica*. *Microbiol. Res.* 206: 6073.
- Lee H, Park JH, Park YK, Kim HJ. 2021. Mathematical modeling for the growth of *Salmonella* spp. and *Staphylococcus aureus* in cake at fluctuating temperatures. *Appl. Sci.* 11: 2475.
- López OAM, Lóopez AM, Crossa J. 2022. *Multivariate Statistical Machine Learning Methods for Genomic Prediction*. Springer, Berlin, Germany.



- MFDS. 2020. Food Composition Databases. Available from: <http://www.foodsafetykorea.go.kr/fcdb/>. Accessed Jun. 18. 2022.
- Oscar TP. 2021. Development and validation of a neural network model for growth of *Salmonella* Newport from chicken on cucumber for use in risk assessment. *J. Food Process. Preserv.* 45: e15819.
- Oscar TP. 2009. General regression neural network and Monte Carlo simulation model for survival and growth of *Salmonella* on raw chicken skin as a function of serotype, temperature, and time for use in risk assessment. *J. Food Prot.* 72: 2078-2087.
- Oscar TP. 2000. Variation of lag time and specific growth rate among 11 strains of *Salmonella* inoculated onto sterile ground chicken breast burgers and incubated at 25C. *J. Food Saf.* 20: 225-236.
- Ölmez HK, Aran N. 2005. Modeling the growth kinetics of *Bacillus cereus* as a function of temperature, pH, sodium lactate and sodium chloride concentrations. *Int. J. Food Microbiol.* 98: 135-143.
- Ridzuan F, Zainon WMNW. 2019. A review on data cleansing methods for big data. *Procedia Comput. Sci.* 161: 731-738.
- Ross T, McMeekin TA, Baranyi J. 2014. Predictive microbiology and food safety. *Encyclopedia of Food Microbiology*. 2nd ed. Batt CA, Tortorello ML (ed). Academic Press, Cambridge, Massachusetts, USA, pp. 59-68.
- Shiffler RE. 1988. Maximum Z scores and outliers. *Am. Stat.* 42: 7980.
- Smiti A. 2020. A critical overview of outlier detection methods. *Comput. Sci. Rev.* 38: 100306.
- Stavropoulou E, Bezirtzoglou E. 2019. Predictive modeling of microbial behavior in food. *Foods.* 8: 654.
- Stangierski J, Weiss D, Kaczmarek A. 2019. Multiple regression models and Artificial Neural Network (ANN) as prediction tools of changes in overall quality during the storage of spreadable processed Gouda cheese. *Eur. Food Res. Technol.* 245: 2539-2547.
- USDA. FoodData Central. 2020. Available from: <http://fdc.nal.usda.gov/>. Accessed Aug. 26. 2021.
- Vrigazova B. 2021. The proportion for splitting data into training and test set for the bootstrap in classification problems. *Bus. Syst. Res.* 12: 228-242.
- Yoo Y. 2019. Hyperparameter optimization of deep neural network using univariate dynamic encoding algorithm for searches. *Knowl.-Based Syst.* 178: 74-83.
- Ziegler GR, Rizvi SSH, Acton JC. 1987. Relationship of water content to textural characteristics, water activity, and thermal conductivity of some commercial sausages. *J. Food Sci.* 52: 901-905.

### Author Information

**Yong Keun Koo:** Dongguk University, Student  
**Yong Woon Jung:** Dongguk University, Student  
**Dong Hwa Kim:** Dongguk University, Student  
**Sang Won Kim:** Dongguk University, Student  
**Eun Seol Kim:** Dongguk University, Student  
**Byeong Jae Park:** Dongguk University, Student  
**Seung Ju Lee:** Dongguk University, Professor  
**Seung Won Jung:** Dongguk University, Research Professor