Design of an Aquaculture Decision Support Model for Improving Profitability of Land-based Fish Farm Based on Statistical Data

Jaeho Lee[1], Wongi Jeon[1], Juhyoung Sung[1], Kiwon Kwon[1], Yangseob Kim[1], Kyungwon Park[1], Jongho Paik[2], and Sungyoon Cho[1*]
[1] Korea Electronics Technology Institute (KETI)
Seongnam-si, Republic of Korea
[e-mail: {smitelims15, jeonwg, jh.sung, kwonkw, likant, kwpark, sycho}@keti.re.kr]
[2] Department of Software Convergence, Seoul Woman's University
Seoul, Republic of Korea
[e-mail: paikjh@swu.ac.kr]
[*]Corresponding author: Sungyoon Cho

Abstract

As problems such as water pollution and fish species depletion have become serious, a land-based fish farming is receiving a great attention for ensuring stable productivity. In the fish farming, it is important to determine the timing of shipments, as one of key factors to increase net profit on the aquaculture. In this paper, we propose a system for predicting net profit to support decision of timing of shipment using fish farming-related statistical data. The prediction system consists of growth and farm-gate price prediction models, a cost statistics table, and a net profit estimation algorithm. The Gaussian process regression (GPR) model is exploited for weight prediction based on the analysis that represents the characteristics of the weight data of cultured fish under the assumption of Gaussian probability processes. Moreover, the long short-term memory (LSTM) model is applied considering the simple time series characteristics of the farm-gate price data. In the case of GPR model, it allows to cope with data missing problem of the weight data collected from the fish farm in the time and temperature domains. To solve the problem that the data acquired from the fish farm is aperiodic and small in amount, we generate the corresponding data by adopting a data augmentation method based on the Gaussian model. Finally, the estimation method for net profit is proposed by concatenating weight, price, and cost predictions. The performance of the proposed system is analyzed by applying the system to the Korean flounder data.

Keywords: Fish farming, Gaussian process regression (GPR), Long short-term memory (LSTM), Net profit estimation, Shipment decision.

1. Introduction

$\mathbf{F}$isheries, including aquaculture, have been advanced rapidly over the past several decades. The global production of fisheries has increased from 19 million tonnes in 1950 to 179 million tonnes in 2018, representing an average annual growth rate of 3.3% [1]. In particular, fish farming, a branch of aquaculture, accounted for approximately 49% of the total production in 2018. This growth has been fundamentally supported by the adoption of advanced technologies such as artificial intelligence and the Internet of Things. These technologies have been applied to various areas of fish farming, including farm management, water quality monitoring, and fish behavior analysis, contributing to the enhanced efficiency and productivity of fish farming [2-5].

From a viewpoint of a fish farm owner, determining the desirable shipment timing is crucial since it is directly connected to profit. This decision-making process should consider various fish farming related factors (FRF) such as the growth rate of fish, farm-gate prices at the source, and operational costs. Consequently, predicting these FRFs is essential, leading to the development of various prediction models.

Initially, deterministic models such as the thermal growth coefficient (TGC) or the von Bertalanffy-based model were developed for predicting fish growth [6, 7]. However, since fish growth data such as weight and length in each fish farming tank or cage showed a random process characteristic, some probabilistic models such as the Bayesian von Bertalanffy model and Gaussian process regression (GPR) were developed [8, 9]. The Bayesian von Bertalanffy model fits the von Bertalanffy growth equation (VBGE) to length-at-age data utilizing Markov chain Monte Carlo (MCMC) and informative priors. GPR is based on Gaussian process and providing outputs of target variables and variances. To predict fish farm-gate prices at the source, the vector autoregressive (VAR) model was proposed. The VAR model analyzed causality in farm-gate prices. Also, artificial neural network models were proposed to analyze the relationship between amounts of consignment sale and farm-gate prices [10, 11]. In the case of cost prediction, several models were proposed, including comprehensive production cost models using the STELLA model in recirculating aquaculture systems (RAS), capital cost estimation models in fish farming based on statistical data and discount rates, and fish price variation models [12-15].

The above-mentioned models, such as growth prediction models [6-9], price prediction models [10, 11], and cost estimation or fish price variation models [12-15], can predict each FRF corresponding to the growth rate of fish, farm-gate prices, and farming costs. However, they cannot provide guidelines for shipment timing that take the total effect of FRFs into consideration. Generally, all the elements of FRFs are dependent on each other. For instance, net profit for fish farmers is usually determined by the values and costs of fish farming, and these values are affected by farm-gate prices, which are dependent on fish weight and farming costs, both of which are also affected by fish growth rate. Hence, this paper suggests a decision support system designed to integrate FRFs related to farmers' revenue, thereby providing guidelines for fish shipment timing. The proposed system consists of prediction models for individual FRFs and a function for net profit prediction as a shipment decision supporter. This paper describes an exemplary case of the system using data from the farming of Korean flounder, a species of significant importance in Korea.

The remainder of this paper is organized as follows. Section 2 describes the architecture of a decision support system. Section 3 explains the design of two prediction models employed within the system and proposes a net profit prediction algorithm as a decision supporter. Section 4 demonstrates the process of the proposed system using specific fish, and Section 5

concludes the paper.

2. The Structure of a Decision Support System

The main components of the proposed decision support system include a controller, a preprocessor, prediction models, and functions as depicted in **Fig. 1**. The controller collects data from various farms, builds a database, and transfers the collected data to the preprocessor. The collected data includes sensor data (e.g., water temperature, dissolved oxygen levels), growth data (e.g., mean weight, number of fish), and farm-gate price data. Furthermore, by using the controller, users can select the output of a function that they want to predict and the prediction models that are required for the specific function. The preprocessor manipulates the raw data from the data resources and converts it into data that can be used by the prediction models. In the training mode, the converted data is also formatted in the form of training data to train the prediction models. If the amount of training data is insufficient, some data augmentation techniques can be used.

The prediction models are designed to predict the weight growth and farm-gate prices of the cultured fish. In this paper, GPR and long short-term memory (LSTM) are used for growth prediction and farm-gate price, respectively. Note that although other machine learning models including artificial neural networks can be replaced, we will focus on the two models in this paper since GPR can provide the mean and variance of the weight data and LSTM can reflect the time-varying characteristics of the prices.
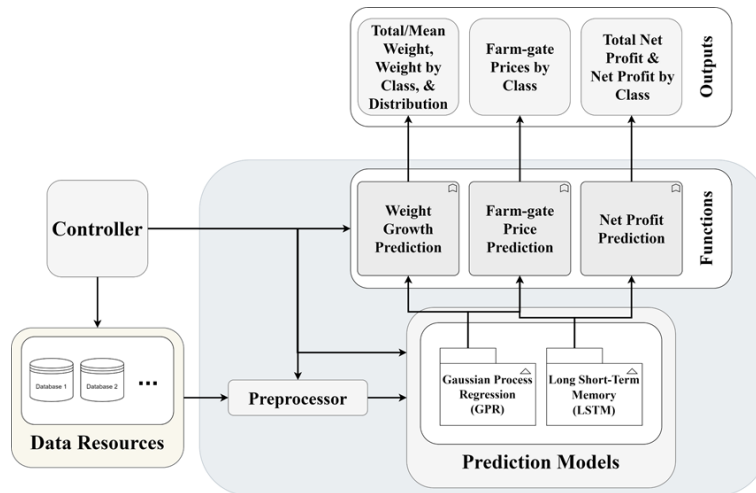

Fig. 1. A block diagram of a decision support system

Finally, the functions provide the predicted outputs that users want to know. The outputs include weight of cultured fish, farm-gate price, and net profit. Weight growth prediction uses the GPR model to provide the total/mean weight, weight by weight class, standard deviation, and a 95% confidence interval of cultured fish. Farm-gate price prediction uses the LSTM model to provide the farm-gate price per weight according to weight classes. Net profit prediction uses both GPR and LSTM models to provide information on when it is advantageous for the shipment. The fish in a specific farm are classified by weight and the net profit for each class and a summation of the classes is estimated using various predicted values from weight of cultured fish, farm-gate prices, and farming costs. Users can know the shipment timing that benefits them by referring to the net profit according to classes or the total net

profit. As a practical example, the system for Korean flounder is described in section 4.3.

3. Design of Prediction Models and a Net Profit Algorithm

The proposed shipment decision assistant system uses various data related to fish farming to support decision-making process from a profit standpoint. To this end, the net profit for each weight class and the summation of all the classes is estimated based on the predicted values of FRFs, which is provided to the user. In this section, we describe two prediction models, GPR for growth and LSTM for farm-gate prices, and propose an algorithm to estimate the net profit using the predicted values. Also, we briefly mention the prediction of farming costs which are used for a net profit estimation.

3.1 Prediction Models

3.1.1 Weight Growth Prediction – GPR

The following three aspects should be considered in selecting a model for weight growth prediction: 1) Not all juvenile fish fed into fish cages (or tanks) grow at the same rate. When analyzing the data collected by sampling at an arbitrary time in a specific farming cage or tank, the weight of fish can be viewed as a random variable with a Gaussian distribution and the set of the random variables that is a function of time can be viewed as a Gaussian random process. Mean and variance at every sampling time is expressed as a function of time and environmental variables such as water temperature or feeding quantity etc. (in this paper, one environmental variable, water temperature, is considered.) 2) Usually, mean and variance of fish weight increases with time, but depending on the fish species or environmental conditions, the increase rate of mean and variance may be mitigated or the mean and variance itself may decrease because large fish can eat small fish from some point in time [16]. 3) Unlike the environment data such as water temperature or dissolved oxygen levels that are automatically collected by sensors, the growth data is collected manually, so it is labor-intensive and can cause damage to the fish with extra costs [17]. Therefore, collecting growth data is not easy, resulting in a small amount of data as well as aperiodic in time.

Considering these aspects, we use GPR as the weight growth prediction model. GPR can customize growth predictions by learning the data collected from each farm. GPR provides standard deviations along with means to understand the distribution of individual weights or serve as the basis for addressing uncertainties. These features allow GPR to provide insights into the variability and distribution of growth of cultured fish to support sorting and shipment timing decisions.

GPR is based on Gaussian processes. A Gaussian process is a collection of random variables which has a joint Gaussian distribution. It can be expressed as the following equation with a training set $\mathcal{D} = \{(\mathbf{x}_i, y_i) \mid i = 0, \dots, n-1\}$ of n observations where $\mathbf{x}_i$ denotes a column input vector of dimension M and y_i indicates a scalar output, and a test set $\mathcal{D}_* = \{(\mathbf{x}_{*i}, y_{*i}) \mid i = 0, \dots, l-1\}$ of l observations where $\mathbf{x}_{*i}$ denotes a column input vector of dimension M and y_{*i} indicates a scalar output [18].

$$f(\boldsymbol{X}) \sim GP(m(\boldsymbol{X}), K(\boldsymbol{X}, \boldsymbol{X}')) \tag{1}$$

where $\boldsymbol{X} = [\mathbf{x}_0 \ \mathbf{x}_1 \ \cdots \ \mathbf{x}_{n-1}]^T$, $m(\boldsymbol{X})$ is the mean function representing the expected value for training $\boldsymbol{X}$, and $K(\boldsymbol{X}, \boldsymbol{X}')$ is the covariance matrix between $\boldsymbol{X}$ and $\boldsymbol{X}'$ which are in a data set.

The covariance matrix measures the similarity between inputs and determines the flexibility of the model.

GPR, as depicted in **Fig. 2**, predicts the distribution of outputs f_* (or $\mathbf{y}_*$) for a new input $\boldsymbol{X}_*$ based on the given training set $\mathcal{D}$. The predictive distribution is expressed as the following conditional probability distribution [18].

$$f_*|\boldsymbol{X}, \mathbf{y}, \boldsymbol{X}_* \sim N(\overline{f_*}, cov(f_*)) \tag{2}$$

$$\overline{f_*} \triangleq E[f_*|\boldsymbol{X}, \mathbf{y}, \boldsymbol{X}_*] = K(\boldsymbol{X}_*, \boldsymbol{X})[K(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 I]^{-1}\mathbf{y} \tag{3}$$

$$cov(f_*) = K(\boldsymbol{X}_*, \boldsymbol{X}_*) - K(\boldsymbol{X}_*, \boldsymbol{X})[K(\boldsymbol{X}, \boldsymbol{X}) + \sigma_n^2 I]^{-1} K(\boldsymbol{X}, \boldsymbol{X}_*) \tag{4}$$

where $\mathbf{y} = [y_0 \ y_1 \ \cdots \ y_{n-1}]^T$, $\boldsymbol{X}_* = [\mathbf{x}_{*0} \ \mathbf{x}_{*1} \ \cdots \ \mathbf{x}_{*l-1}]^T$, $\mathbf{y}_* = [y_{*0} \ y_{*1} \ \cdots \ y_{*l-1}]^T$, $\overline{f_*}$ represents the mean vector of the predicted outputs for a new input X_*, $\mathbf{y}$ is a column vector output consisting of y_i, $cov(f_*)$ denotes the covariance matrix of the predictions, σ_n^2 represents the variance, I denotes the identity matrix, $K(X, X)$ is a covariance matrix (or Gram matrix) whose (i, j)th entry is $K_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$ for $1 \le i, j \le n$.
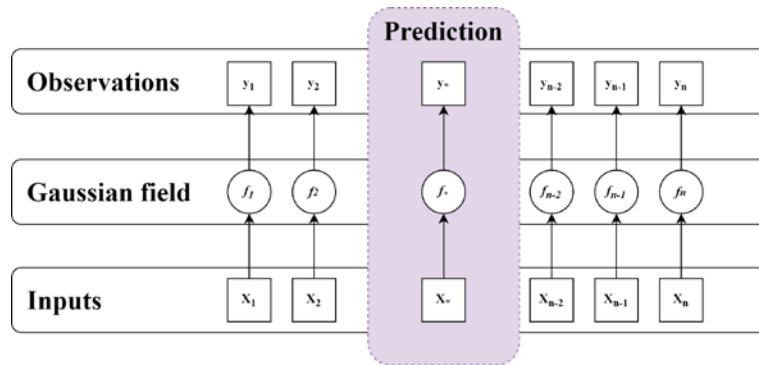

Fig. 2. A concept of Gaussian process regression

For predicting the mean weight and variance of cultured fish, the GPR model utilized Python and the GPy library [19]. The covariance matrix of the designed GPR was simplified using some well-known kernels such as squared exponential (SE), product (linear and white noise kernels), Matérn, and ARD SE kernels [18, 20].

$$K(\boldsymbol{X}, \boldsymbol{X}_*) = K_{trend}(\boldsymbol{X}, \boldsymbol{X}_*) + K_{noise}(\boldsymbol{X}, \boldsymbol{X}_*) + K_{Matérn}(\boldsymbol{X}, \boldsymbol{X}_*) + K_{ARD}(\boldsymbol{X}, \boldsymbol{X}_*) \tag{5}$$

where

$$K_{trend}(\boldsymbol{X}, \boldsymbol{X}_*) = K_{trend_{ij}} = k_{trend}(\mathbf{x}_i, \mathbf{x}_{*j}) = \sigma_{trend}^2 \exp\left(-\frac{1}{2l_N^2}|\mathbf{x}_i - \mathbf{x}_{*j}|^2\right) \tag{6}$$

$$K_{noise}(\boldsymbol{X}, \boldsymbol{X}_*) = K_{noise\_ij} = k_{noise}(\mathbf{x}_i, \mathbf{x}_{*j}) = \sigma_{noise}^2 \delta(\mathbf{x}_i, \mathbf{x}_{*j}) \times (\mathbf{x}_i^T \mathbf{x}_{*j}) \tag{7}$$

$$K_{Matérn}(\boldsymbol{X}, \boldsymbol{X}_*) = K_{Matérn\_ij} = k_{Matérn}(\mathbf{x}_i, \mathbf{x}_{*j}) = \sigma_{Matérn}^2 \left(1 + \frac{\sqrt{5}r}{l_m} + \frac{5r^2}{3l_m^2}\right)\exp(-\frac{\sqrt{5}r}{l_m}) \tag{8}$$

$$K_{ARD}(\boldsymbol{X}, \boldsymbol{X}_*) = K_{ARD\_ij} = k_{ARD}(\mathbf{x}_i, \mathbf{x}_{*j}) = \sigma_{ARD}^2 \exp\left(-\frac{1}{2}\sum_{d=1}^{M}\frac{\left(\mathbf{x}_i^{[d]} - \mathbf{x}_{*j}^{[d]}\right)^2}{l_d^2}\right) \tag{9}$$

where σ_{trend}^2, σ_{noise}^2, $\sigma_{Matérn}^2$, and σ_{ARD}^2 represent the variances of K_{trend}, K_{noise}, $K_{Matérn}$, and K_{ARD}, respectively. Also, l_N, l_m, and l_d denote the length scales for K_{trend}, $K_{Matérn}$, and K_{ARD} of dth dimension, respectively. $\delta(\mathbf{x}_i, \mathbf{x}_{*j})$ represents the Kronecker delta, which returns 1 if the input vectors are identical and 0 otherwise, $\mathbf{x}_i{}^T\mathbf{x}_{*j}$ denotes the dot product of the two input vectors. r represents the absolute distance $|\mathbf{x}_i - \mathbf{x}_{*j}|$. $\mathbf{x}_i^{[d]}$ and $\mathbf{x}_{*j}^{[d]}$ represents the values of the dth dimension of the vectors $\mathbf{x}_i$ and $\mathbf{x}_{*j}$ when automatic relevance determination (ARD) is activated. ARD allows for learning a separate length scale parameter, enabling the model to estimate the importance of each dimension. The prediction results of the designed model are described in section 4.2.

To evaluate the performance of the GPR model, mean square error (MSE) and mean standardized log-likelihood (MSLL) are used. The MSE value is obtained by selecting the mean weight predictions for each aquacultivation period as representative values [9]. MSLL can be calculated in a similar manner and is particularly useful for evaluating probabilistic models, making it a criterion for performance evaluation along with MSE [18]. MSE and MSLL are defined as follows [9].

$$MSE(\mathbf{y}_*, \overline{f}_*) = \frac{1}{l}(y_* - \overline{f}_*)^T(y_* - \overline{f}_*) \tag{10}$$

$$MSLL(\mathbf{y}_*, \overline{f}_*) = \frac{1}{l}\sum_{i=1}^{n}(\frac{1}{2}\log(2\pi\sigma_*^2) + \frac{(y_* - \overline{f}_*)^2}{2\sigma_*^2}) \tag{11}$$

where σ_*^2 indicates the predicted variance corresponding to $\mathbf{X}_*$.

3.1.2 Farm Gate Price Prediction – LSTM

The farm-gate prices per unit weight does not apply equally to all fish with different weights, but rather tends to increase as the weight of fish increases. This is because the weight shows a Gaussian distribution, which results in the scarcity of relatively heavier fish favored by consumers. Usually, the farm-gate prices are not randomly determined but are affected by previous price trends, which are more affected by recent prices. Therefore, it is possible to improve the prediction accuracy by using multiple time series data. In this paper, we select a recurrent neural network (RNN) based model that sequentially processes inputs and outputs for price prediction [21]. Specifically, the training of the RNN model utilizes the LSTM structure that can effectively learn temporal dependencies in sequential data for accurate predictions of current or future data, which is widely used in price prediction models for agricultural [22, 23] and livestock products [24, 25].

The past time series data before the forecasting time were processed using a many-to-one structure in which several inputs lead to a single output, as shown in **Fig. 3** [26]. In this study, the LSTM was trained to predict the farm-gate price one month ahead, using the previous five months of price farm-gate information as input. In predicting prices for flounder, deciding sequence length which is about how many monthly prices are considered as an input for LSTM model is a critical factor. This is because prices for flounder are affected by seasons or day of the week able to cause changes in water temperature, farming costs, growth rate etc. To effectively reflect those seasonal effects on the price prediction, several LSTM models were tested with different sequence lengths, where the sequence lengths of 3, 5, 6, and 12 which show MSLE of 15.505, 14.558, 14.846, and 14.741, respectively. Thus, the sequence length of 5 for price prediction is selected. And the price information used as input to LSTM is scaled

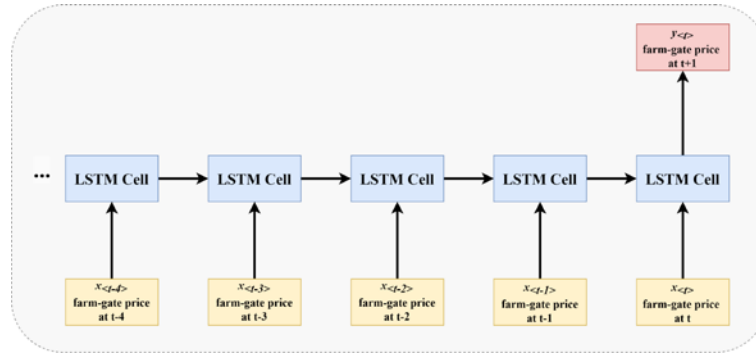by a min-max normalizer to ensure all values range between 0 and 1.

Fig. 3. Structure of many-to-one LSTM

The model was designed to predict farm-gate prices for different weight classes, specifically for 0.5 kg, 1 kg, and 2 kg, taking into account the varying tendencies in price fluctuations associated with each of the weight classes. Although the same LSTM structure is used, the model is designed to adjust weight values according to weight class to reflect the different price trends by class. The prediction outcomes of the designed model are going to be detailed in subsequent sections.

3.2 A Net Profit Algorithm

To predict the net profit of fish farming, the growth prediction model implemented with GPR, farm-gate price prediction model using LSTM, and cost data based on statistics are used. Fundamentally, the growth prediction model uses sample mean weight, aquacultivation period, and water temperature data from the data resource's database as inputs for training, producing means and variances of weight as outputs. The farm-gate price prediction model is trained with unit prices per weight of fish as inputs and estimates unit prices by weight class as outputs. To predict net profit that can assist shipment decision, a statistical technique is used together with cost based on predictive models and statistics. The formula for the net profit algorithm is as follows.

$$C_T(m, T) = c(A, N_S) \times W_T(m, T) + C_s \times N_s \tag{12}$$

where $C_T(m, T)$ represents the total farming cost after m months with the water temperature T in the farms, $c(A, N_S)$ denotes the cost per unit of production, A is the total water surface area of the farms, N_S indicates the total number of juvenile fish stocked in the farms, $W_T(m, T)$ signifies the total weight of fish, and C_s is the cost per juvenile fish.

The proportion of fish belonging to each weight class is as follows.

$$p_i(m, T) = \Pr\left(W \in Range(W_i)\right) \tag{13}$$

where i indicates the index of the weight class, and W is a random variable representing the weight of the fish. W_i is the representative value according to the weight class, and $Range(W_i)$ is the weight range of flounder (fish) satisfying W_i. The proportions for each of the weight classes are derived using the mean weight and standard deviation predicted by GPR, along with the cumulative distribution function $F(W)$ based on the Gaussian distribution.

Table 1 provides an example based on Korean flounder, describing the weight classes.

Table 1. Weight class specification about Korean flounder as an exemplary case

Weight class index (i)	Weight class (W_i) (Representative value)	Range of fish weight (W) [unit: g]	$p_i(m, T)$ with cumulative distribution function
0	125 g	$0 \leq W < 250$	$F(250)$
1	500 g	$250 \leq W < 750$	$F(750) - F(250)$
2	1000 g	$750 \leq W < 1250$	$F(1250) - F(750)$
3	1500 g	$1250 \leq W < 1750$	$F(1750) - F(1250)$
4	2000 g	$1750 \leq W < 5000$	$F(5000) - F(1750)$

The shipment value per weight class is given by

$$v_i(m, T) = W_i \times p_i(m, T) \times N_S \times (1 - d(m)) \times PM_i(m) \tag{14}$$

where $PM_i(m)$ and $d(m)$ denote the farm-gate price per weight for weight class i and the mortality rate of the fish, respectively.

The total net profit is given by

$$NP(m, T) = \sum_i v_i(m, T) - C_T(m, T) \tag{15}$$

Finally, the net profit for fish belonging to weight class i is as follows.

$$NP(i, m, T) = v_i(m, T) - p_i(m, T)C_T(m, T) \tag{16}$$

4. Simulations

In this section, the entire operation process of the shipping decision support system proposed in section 3 is described based on Korean flounder data.

4.1 Simulation Assumption

As a use case for the shipment decision support system, Korean flounder from Jeju Island is selected. As mentioned in section 3.1.1, the currently available data is aperiodic and small in quantity, making it difficult to obtain sufficient data for prediction. Therefore, this subsection assumes the following conditions.

1) For the land-based Korean flounder farming in the Jeju region introduced as a use case, it is observed that the survival rate decreases with the length of the aquacultivation period. Additionally, due to the characteristic that the unit farm-gate price varies according to weight classes, farming juveniles up to around 1.0 kg before shipping can increase the farmer's profit [27]. Therefore, in this use case, a prediction period of 18 months after stocking 4 g juveniles

is set to observe the growth up to approximately 1.0 kg and subsequent trends.

2) In flounder farming, seasonal factors typically cause variations in water temperature at the farms. However, it is assumed for this use case that the water temperature remains constant for each case as depicted in **Fig. 4**, and it is assumed that optimized feeding is provided in each case.

3) During the 18-month aquacultivation period or shorter, grading and sorting occurs, and it is assumed for growth data collection that the weight and distribution of the selected individuals are determined by sampling of size 50.

4) Data collected from the farm can vary in collection frequency depending on the data type, and anomalies or missing data may occur due to grading, sorting, shipment decisions, or sensor malfunctions. Thus, preprocessing is necessary [9], but this simulation assumes that such processes have already been completed, and mixed data cycles are converted and standardized to a monthly basis for use.

5) It is assumed that small fish below 250 g does not possess marketable value and thus cannot be sold, and when setting the 18-month observation period, it is assumed that the weight of flounder does not exceed 5000 g.
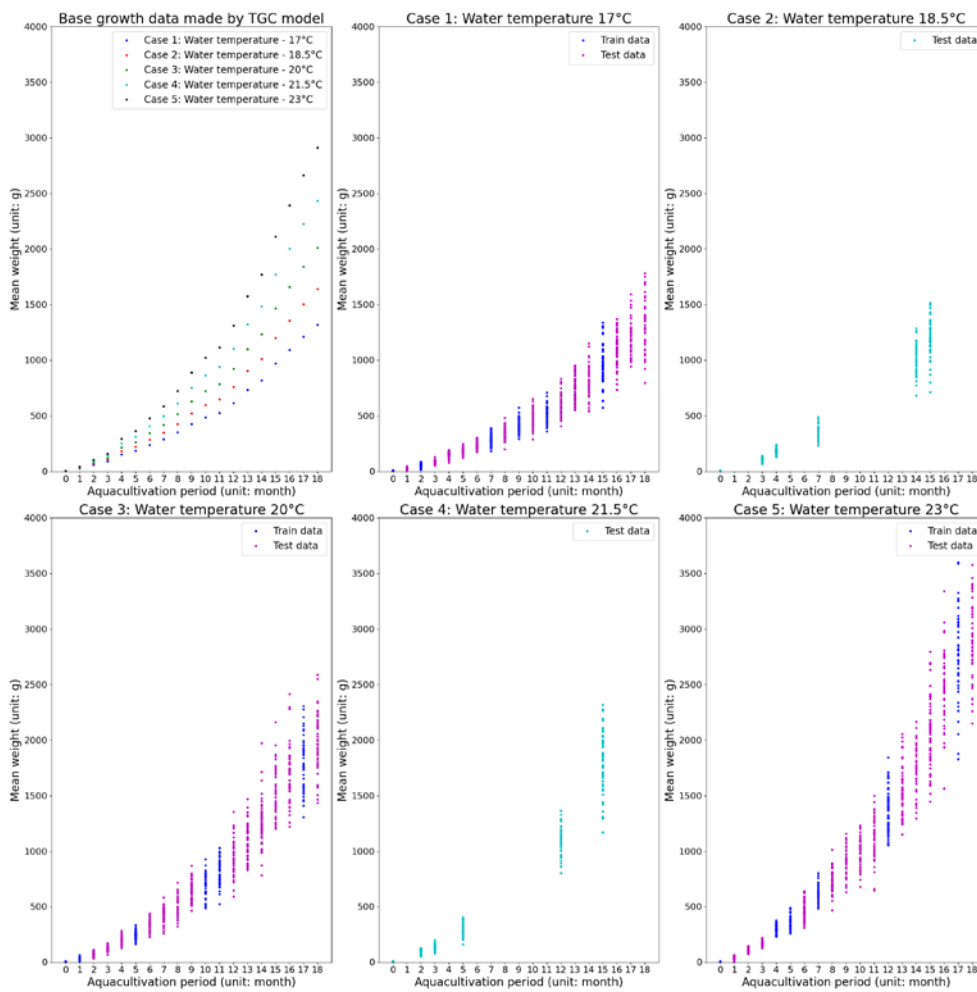

Fig. 4. Baseline data of fish weight and data augmentation according to water temperature conditions and an aquacultivation period

4.2 Simulation Results

4.2.1 Weight Growth Prediction

In this subsection, we perform training and prediction using the model structure described in section 3.1.1 and the growth data for Korean flounder. Due to the lack of sufficient data accumulated in the living lab database for training and testing the GPR growth model, the use case to be discussed later uses the TGC model and mean growth data of Korean flounder to prepare basic data for training and test. The standard deviation relative to the mean weight was obtained using the living lab database.

As can be seen in the first graph of the first row in **Fig. 4**, five sets of mean growth data were generated according to water temperature ranging from 17 to 23℃ and an aquacultivation period and to obtain TGC values [28-30]. The formula for the TGC growth model is as follows [6].

$$W_e = [\sqrt[3]{W_s} + \left(\frac{TGC}{1000} \times T \times \frac{365}{12} \times \Delta m\right)]^3 \tag{17}$$

where W_e represents the weight of the Korean flounder after m months from stock seeding, W_s is the initial weight of the flounder, and TGC stands for the thermal growth coefficient dependent on the water temperature.

For the training and evaluating of GPR, data augmentation (noise addition) was conducted on the baseline data according to water temperature conditions and an aquacultivation period [31]. The data augmentation was performed using the standard deviation corresponding to the mean weight of each period and a normal distribution. For training and test data for Cases 1, 3, and 5 where the water temperatures are 17, 20, and 23℃, respectively, augmented TGC growth data over the aquacultivation period of 18 months from stock seeding was randomly divided into training data (6 sets, each with 50 samples - 300 total) and test data (13 sets, each with 50 samples - 650 total). Here, we assumed that data from the stock seeding period could be secured and thus was all included in the training data.

Using this training data, we trained and evaluated the growth prediction model for Case 1, 3, and 5. Next, we performed the growth predictions and evaluations for the different water temperature (Cases 2, 4 where the water temperatures in the farms observed are 18.5 and 21.5℃, respectively) based on the model for Cases 1, 3, and 5.

The reason why the augmented data was prepared in this way is that the quantity of growth data is very small and sampling frequency is aperiodic. This suggests 6 growth datasets per year can be acquired from Korea flounder farms in the field. Thus, in this paper, it is assumed that Cases 1, 3, and 5 have 6 growth datasets as training data and 13 datasets as test data and for Cases 2 and 5, 6 growth datasets as test data are acquired.

The optimization of the GPR model was performed using the L-BFGS-B optimizer where L-BFGS-B is a variation of the Limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, designed for solving large-scale optimization problems with bound constraints on variables, resulting in a marginal likelihood of 4758.36. The hyperparameters of the covariance function, as presented in **Table 2**, were as follows: for K_{trend} in (6), σ^2_{trend} and l_N were 17,988,794.25 and 14.87, respectively; for K_{noise} in (7), σ^2_{noise} were 11.92; for for $K_{Matérn}$ in (8), $\sigma^2_{Matérn}$ and $l_{Matérn}$ were 5.56 and 5.58, respectively; and for K_{ARD} in (9), σ^2_{ARD} and l^2_d were 5.56 and 2, respectively.

Table 2. Parameter specification of growth prediction model (GPR)

Parameters	Values
sum.rbf.variance	17988794.25
sum.rbf.lengthscale	14.87
sum.mul.white.variance	11.92
sum.mul.linear.variances	11.92
sum.Mat52.variance	5.56
sum.Mat52.lengthscale	5.85
sum.rbf_1.variance	5.56
sum.rbf_1.lengthscale	(2,)
Gaussian_noise.variance	3.91
Marginal likelihood	4758.36
Optimizer	L-BFGS-B

In **Fig. 5**, we presented training-test data with the mean weight and 95% confidence interval of the growth prediction model, based on the aquacultivation period and water temperature. For Cases 1, 3, and 5, experiments were performed using 6 months of training data and 13 months of test data. For Cases 2 and 4, experiments were carried out in untrained culturing environments based on the GPR model trained from Cases 1, 3, and 5.
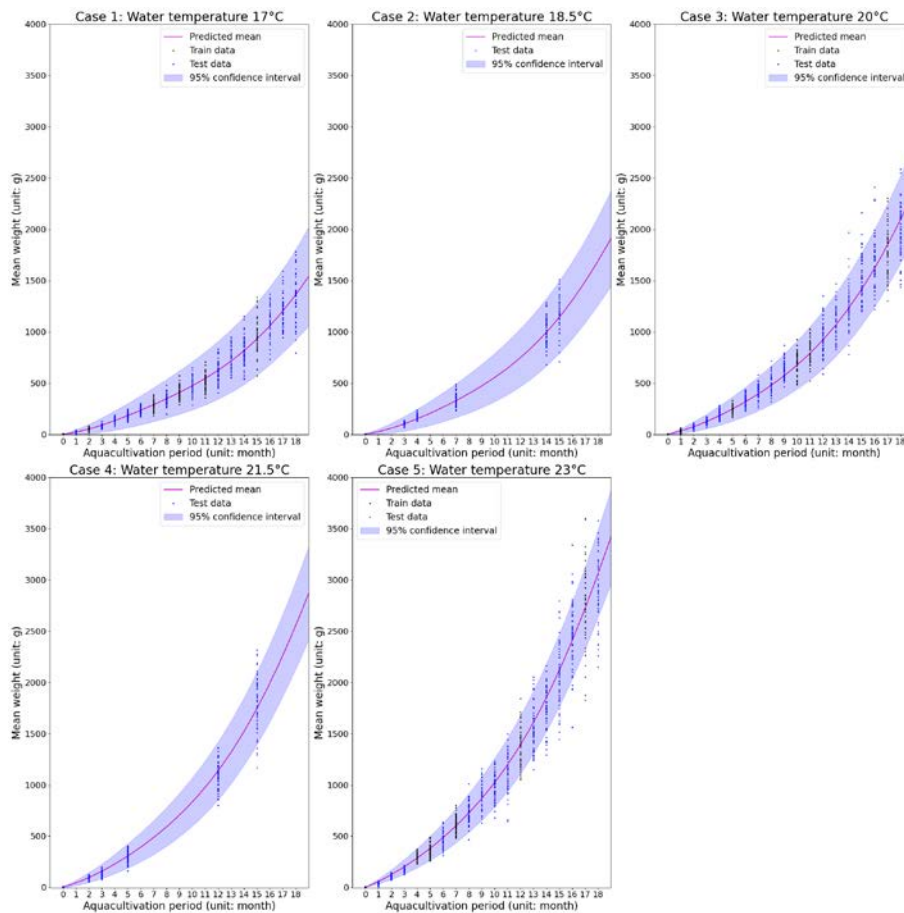
Fig. 5. Weight growth prediction according to water temperature ranging from 17 to 23℃ and an aquacultivation period

According to **Table 3**, for Cases 1, 3, and 5, the MSE values are 17.38, 38.39, and 67.55, respectively, showing an increasing trend with the increase in culturing water temperature. The MSLL values are 5.69, 6.02, and 6.42, respectively, which also display an increasing trend. When examining Cases 2 and 4, the MSE values are 23.46 and 81.77, respectively, and the MSLL values are 5.82 and 6.42, respectively. In Case 2, the MSLL is between that of Cases 1 and 3, whereas in Case 4, the MSLL exceeds that of Cases 3 and 5. While the values of MSE and MSLL can vary depending on the distribution of the growth prediction model and the training and test data, there are no significantly deviating predictions for the test data.

Table 3. Performance of growth prediction model in terms of MSE and MSLL

	Case 1	Case 2	Case 3	Case 4	Case 5
MSE	17.38	23.46	38.39	81.77	67.55
MSLL	5.69	5.82	6.02	6.57	6.42

4.2.2 Farm-gate Price Prediction

In this subsection, we predicted the farm-gate prices using the LSTM model, as described in subsection 3.1.2. As previously mentioned, the model uses five time series data points as inputs to produce a single output value. The hidden layer size of the LSTM is set to 20, with the objective function aiming to minimize the mean squared log error (MSLE) loss. Detailed information about the model is presented in **Table 4**.

Table 4. Parameter specification of Farm-gate prediction model (LSTM)

Parameters	Values
Input data time step	5
Input data feature size	1
Output data length	1
Output data feature size	1
Number of training epochs	500
Learning rate	0.005
Dropout rate	0.3
Batch size	64
LSTM hidden layer size	20
Loss function	MSLE
Optimizer	Adam
MSLE	14.558

For the training and test of the LSTM model, data from Korea maritime institute fisheries observation center was used [32], with a total of 168 months of price data for 0.5 kg, 1.0 kg, and 2.0 kg from January 2008 to December 2021. To test the performance of the price prediction model, 70% of the total price data was used for training the model, and the remaining 30% was used as test data to evaluate the prediction model.

Fig. 6 illustrates the loss function during the training and the results of the model evaluation for the 2 kg farm-gate price prediction model. As can be seen in **Fig. 6 (a)**, the training loss decreases as the model training progresses. In **Fig. 6 (b)**, actual farm-gate prices for 2.0 kg flounder are compared with predictions from the LSTM model. To evaluate LSTM performance, the VAR model, commonly used for price forecasting, was also used [10]. As the VAR model follows the multivariate time series forecasting, it accounts for the

interrelationship between 1.0 kg and 2.0 kg flounder prices in this simulation. Both models were trained on the same data and forecasted prices for 18 months starting from September 2019. The green, orange, and blue lines represent the actual farm-gate prices, the predicted prices from the LSTM model, and the predicted prices from the VAR model, respectively. The comparison of performance results indicates that the farm-gate prices predicted by the LSTM model closely match the actual values. In this simulation, the LSTM model captures temporal dependencies in sequential data, resulting in accurate predictions for future events. The model achieved MSLE of 0.005, compared to 0.0139 for the VAR model.
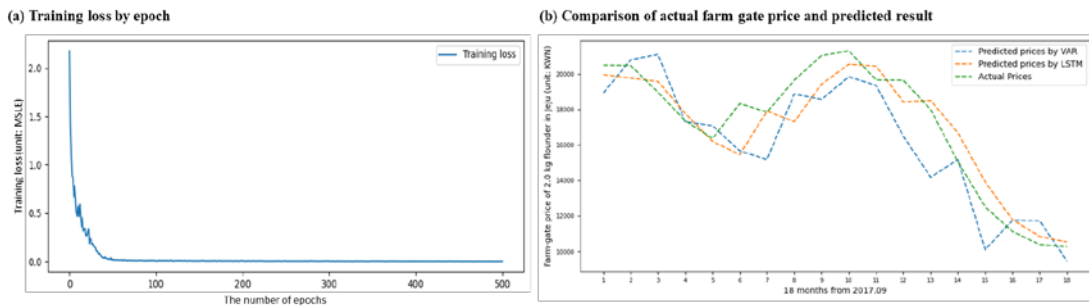
Fig. 6. Training loss and prediction performance of LSTM for farm gate price for 2.0 kg flounder class: (a) Training loss by epoch (b) Comparison of actual farm gate price and predicted prices for 2.0 kg flounder by VAR and LSTM

4.2.3 Net Profit Prediction

To predict net profit using the aforementioned prediction models, the profitability of land-based Korean flounder farming was assessed under the optimal condition of having water surface areas of 6,611.57 m^2 in fish farms with average water temperature of 20℃, utilizing cost statistics per kg of production as shown in **Table 5**. It is assumed that in September 2023, 30,000 juvenile fish weighing 4 g each were stocked. The cost per juvenile fish for stock seeding is 350 KWN, the unit production is 6,835.37 KWN (KWON: Korean currency unit) [27, 33].

Table 5. Parameter specification for net profit prediction

Water surface area of farm [unit: m^2]	$A = 6{,}611.57$
Water temperature [unit: ℃]	$T = 20$
Number of stock seeding [unit: qty]	$N_S = 30{,}000$
Cost of stock seeding [unit: $(\frac{KWN}{\# fish})$]	$C_s = 350$
Unit production cost [unit: $(\frac{KWN}{kg})$]	$c(A, N_S) = 6{,}835.37$

Based on the weight growth and farm-gate price prediction models, along with farming cost statistics, the estimated net profit is depicted in **Fig. 7**, where the corresponding figure focuses on the range of farming weights up to 2kg. Examining the first row in **Fig. 7**, the total

weight (kg) of flounder cultivated over time (months) shows a monotonic increase. Excluding the weight class corresponding to 2.0 kg, the total weight of each weight class increases, reaches a peak, and then decreases over time. This pattern illustrates that as the weight of the cultured individuals increases, the proportion of weight classes accordingly changes.
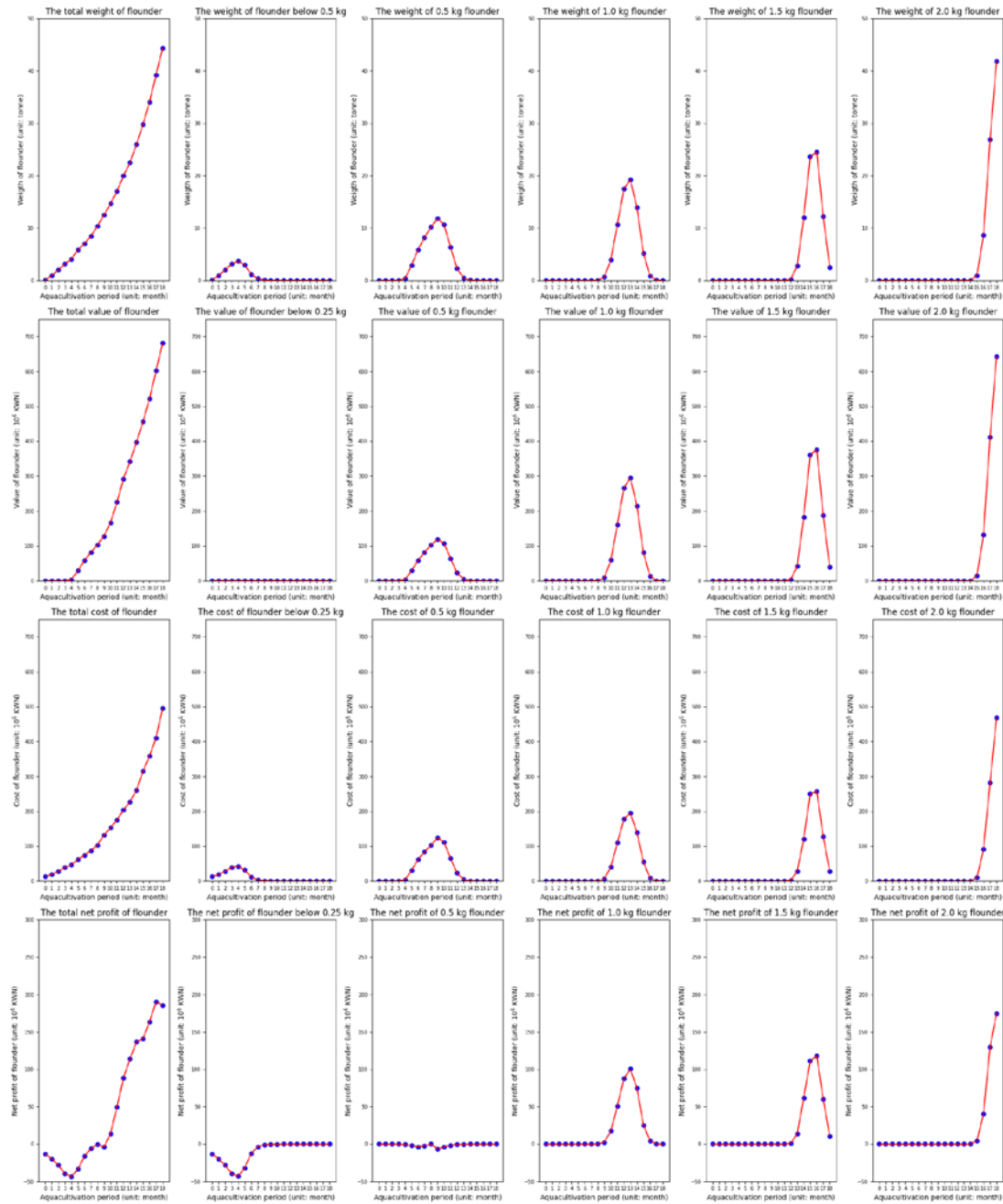

Fig. 7. Net profit inferences from predicted weight of fish, farm gate prices, and farming costs

The second row represents the shipment value (KWN) of flounder over the aquacultivation period (months). As the farming time progresses, the total shipment value increases in proportion to the distribution of total weight at each weight class, except for the case of weights below 0.25 kg in second column in second row of **Fig. 7**. This is because we follow the fifth assumption of section 4.1, which assumes that small fish below 0.25 kg have no marketable value so that the corresponding value remains at zero. However, the shipment values for the remaining weight classes show a pattern similar to the estimated weights per weight class.

The third row illustrates the farming costs (KWN) over time (months). With time, the total farming cost shows an increasing trend, with costs by weight class exhibiting a pattern similar to the estimated weights per weight class.

The fourth row presents estimates for net profit (KWN) over the aquacultivation period (months). Overall, due to the interplay of growth, farm-gate prices, farming costs, and survival rates, the break-even point for net profit is represented at 8 months of cultivation. After this, the net profit gradually increases with a slower rate observed at 14 and 15 months and the maximum net profit during the 18-month cultivation period appears at 17 months. However, net profits by weight classes vary according to the weight of flounder, farm-gate prices, and farming costs. For example, as shown in the second column of fifth row in **Fig. 7**, the weight class below 0. 25 kg is difficult to trade in the market, resulting in a net loss where total net profit is negative due to increasing costs during the aquacultivation period. Additionally, as shown in the third column of fifth row in **Fig. 7**, the weight class of 0.5 kg also incurs a net loss, because the farming costs are relatively high compared to the farm-gate price. However, the net profit gradually increases in other weight classes above 0.5 kg.

5. Conclusion

In this paper, we designed a prediction system to assist the shipment decisions for fish farmers. The prediction system consists of a growth prediction model, a farm-gate price prediction model, a cost statistics table, and a net profit estimation algorithm. The GPR model was used for weight growth prediction based on the analysis that the characteristics of the weight data are Gaussian probability processes, and the LSTM model was used in consideration of the simple time series characteristics of the farm-gate price data. In the case of the GPR model, it is possible to cope with a data missing problem of the weight data collected from the fish farm in the time and temperature domains. Also, to solve the problem that the data acquired from the fish farms was aperiodic and small in amount, a data augmentation method based on the Gaussian model was used. After explaining the function of each prediction model, an estimation method for net profit using weight, price, and cost was proposed. The performance was analyzed by applying the proposed system to the Korean flounder data. Farmers can determine the timing of shipments of cultured fish so that they can improve their profit by referring to the net profit prediction.

References

[1] Food and Agriculture Organization of the United Nations, Part 1 World review, The State of World Fisheries and Aquaculture 2022, Rome, Italy: FAO, 2022. Article(CrossRefLink)

[2] M. Chiu, W. Yan, S. A. Bhat, and N. Huang, "Development of smart aquaculture farm management system using IoT and AI-based surrogate models," *Journal of Agriculture and Food Research*, vol.9, Sep. 2022. Article (CrossRef Link)

[3] K. Tsai, L. Chen, L. Yang, H. Shiu, and H. Chen, "IoT based smart aquaculture system with automatic aerating and water quality monitoring," *Journal of Internet Technology*, vol.23, no.1, pp.179-186, Jan. 2022. Article (CrossRef Link)

[4] K. P. Rasheed Abdul Haq and V. P. Harigovindan, "Water Quality Prediction for Smart Aquaculture Using Hybrid Deep Learning Models," *IEEE Access*, vol.10, pp.60078-60098, Jun. 2022. Article (CrossRef Link)

[5] L. Du, Z. Lu, and D. Li, "Broodstock breeding behaviour recognition based on Resnet50-LSTM with CBAM attention mechanism," *Computers and Electronics in Agriculture*, vol.202, 2022. Article (CrossRef Link)

[6] M. Jobling, "The thermal growth coefficient (TGC) model of fish growth: a cautionary note," *Aquaculture Research*, vol.34, no.7, pp.581-584, May 2003. Article (CrossRef Link)

[7] M. S. Chambers, L. A. Sidhu, B. O'Neil, and N. Sibanda, "Flexible von Bertalanffy growth models incorporating Bayesian splines," *Ecological Modelling*, vol.355, pp.1-11, Jul. 2017. Article (CrossRef Link)

[8] K. I. Siegfried and B. Sansó, "Two Bayesian methods for estimating parameters of the von Bertalanffy growth equation," *Environmental Biology of Fishes*, vol.77, pp.301-308, Aug. 2006. Article (CrossRef Link)

[9] J. Kim, E. Park, S. Cho, K. Kwon, and Y. M. Ko, "Probabilistic Modeling of Fish Growth in Smart Aquaculture Systems," *KSII Transactions on Internet and Information Systems*, vol.17, no.8, pp.2259-2277, Aug. 2023. Article (CrossRef Link)

[10] J. Son, J. Nam, "A Leading Price Estimation of Jeju Flounder Producer Prices by Fish Weight and a Dynamic Influence Analysis of Market Price Impulse," *Journal of Fisheries and Marine Sciences Education*, vol.28, no.1, pp.198-210, 2016. Article(CrossRefLink)

[11] K. Hwang, J. Choi, and T. Oh, "Forecasting common mackerel auction price by artificial neural network in Busan Cooperative Fish Market before introducing TAC system in Korea," *Journal of the Korean Society of Fisheries and Ocean Technology*, vol.48, no.1, pp.72-81, Feb. 2012. Article (CrossRef Link)

[12] T. M. Losordo and P. W. Westerman, "An Analysis of Biological, Economic, and Engineering Factors Affecting the Cost of Fish Production in Recirculating Aquaculture Systems," *Journal of the World Aquaculture Society*, vol.25, no.2, pp.193-203, Jun. 1994. Article (CrossRef Link)

[13] S. R. Campo and S. Zuniga-Jara, "Reviewing capital cost estimations in aquaculture," *Aquaculture Economics & Management*, vol.22, no.1, pp.72-93, 2018. Article (CrossRef Link)

[14] R. E. Dahl and A. Oglend, "Fish Price Volatility," *Marine Resource Economics*, vol.29, no.4, pp.305-322, Dec. 2014. Article (CrossRef Link)

[15] P. Deb, M. M. Dey, and P. Surathkal, "Fish price volatility dynamics in Bangladesh," *Aquaculture Economics & Management*, vol.26, no.4, pp.462-482, 2022. Article (CrossRef Link)

[16] K. Kim, H. Hwang, H. Kim, K. Kim, J. Do, and M. Jung, "Hatchery production," *Flounder Aquaculture Standard Manual*, Busan, Republic of Korea: NIFS, ch.3, sec.2, pp.36-38, 2016. [Online]. Available: https://www.nifs.go.kr/

[17] Y. Eh, "Production planning in fish farm," *The Journal of Fisheries Business Administration*, vol.46, no.3, pp.129-141, Dec. 2015. Article (CrossRef Link)

[18] C. Rasmussen and C. Williams, Gaussian Processes for Machine Learning, Gaussian processes for Machine Learning, MIT Press, 2006. Article (CrossRefLink)

[19] GPy, GPy: A gaussian process framework in python. [Online]. Available: http://github.com/SheffieldML/GPy

[20] GPy, GPy - A Gaussian Process (GP) framework in Python, Read the Docs, 2020. [Online]. Available: https://gpy.readthedocs.io/en/deploy/index.html

[21] S. A. Zargar, "Introduction to Sequence Learning Models: RNN, LSTM, GRU," Apr. 2021. Article (CrossRef Link)

[22] R. C. Staudemeyer and E. R. Morris, "Understanding LSTM -- a tutorial into Long Short-Term Memory Recurrent Neural Networks," *arXiv:1909.09586*, Sep. 2019. Article (CrossRef Link)

[23] K. Kurumatani, "Time series forecasting of agricultural product prices based on recurrent neural networks and its evaluation method," *SN Applied Sciences*, vol.2, Jul. 2020. Article (CrossRef Link)

[24] T. Zhang and Z. Tang, "Agricultural commodity futures prices prediction based on a new hybrid forecasting model combining quadratic decomposition technology and LSTM model," *Frontiers in Sustainable Food Systems*, vol.8, Feb. 2024. Article (CrossRef Link)

[25] J. Chen, L. Lin, and X. Li, "Pork Price Prediction Using Bi-RNN-LSTM Artificial Neural Network," in *Proc. of 5th International Conference on Artificial Intelligence and Big Data (ICAIBD)*, pp.168-172, 2022. Article (CrossRef Link)

[26] Y. J. Lee, K. S. Ko, D. H. Hwang, S.A. Lee, and J. P. Cho, "A Study on the Prediction Model of Chicken Price Using a Multi-Variable LSTM Deep Learning Network," *The Journal of Korean Institute of Communication and Information Sciences*, vol.47, no.12, pp.2058-2064, Dec. 2022. Article (CrossRef Link)

[27] K. Kim, H. Hwang, H. Kim, K. Kim, J. Do, and M. Jung, "Economical analysis of aquaculture," *Flounder Aquaculture Standard Manual*, Busan, Republic of Korea: NIFS, ch.8, pp.91-100, 2016. [Online]. Available: https://www.nifs.go.kr/

[28] Y. Lee, "Productivity comparison to the management of olive flounder, Paralichthys olivaceus aqua-farms in Jeju and Wando," *M.S. thesis, Dept. Proliferation Science, Jeju Univ.*, 2019. Article (CrossRef Link)

[29] K. Kim, H. Hwang, H. Kim, K. Kim, J. Do, and M. Jung, "Rearing," *Flounder Aquaculture Standard Manual*, Busan, Republic of Korea: NIFS, ch.4, sec.5, pp.52-54, 2016. [Online]. Available: https://www.nifs.go.kr/

[30] D. Park, M. Son, M. Park, et al., "General status and ecology," Standard Manual of Olive Flounder Culture, Busan, Republic of Korea: NIFS, ch. 1, pp. 1-7, 2006.

[31] K. Song, T. Park, and J. Chang, "Novel Data Augmentation Employing Multivariate Gaussian Distribution for Neural Network-Based Blood Pressure Estimation," *Applied Sciences*, vol.11, no.9, Apr. 2021. Article (CrossRef Link)

[32] KMI, Observation Statistics of Korean Halibuts. [Online]. Available: https://www.foc.re.kr/

[33] Y. Yang, "Analysis of the management status and economic viability," in Analysis of the Management Status and Economic Viability of Flatfish Aquaculture in the Jeju Area, Jeju, Republic of Korea, JDI, ch.4, pp. 52-76, 2010.

Jaeho Lee received the B.S. and B.Eng. degrees in Manufacturing Systems and Design Engineering (MSDE) from Seoul National University of Science and Technology, Seoul, Korea, and Northumbria University, Newcastle, United Kingdom, in 2023, respectively. He has been an assistant researcher in Korea Electronics Technology Institute (KETI), Seoul, Korea, since 2023. His research interests include artificial intelligence, robotics, and control.

Wongi Jeon received the B.S., M.S., and Ph.D. degrees in Electronic Engineering from Chung-Ang University, Seoul, Korea, in 1994, 1996, and 1999, respectively. In 2001, he joined in Korea Electronics Technology Institute (KETI), Korea, where he is presently a chief researcher. His research interests are in the area of intelligent signal processing and wireless digital communications, especially for multicarrier transmission systems.

Juhyoung Sung received the B.S. degree (as a valedictorian) in Computer and Communication Engineering (currently Computer Science) from Korea University, Seoul, South Korea, in 2015; the M.S. degree in Electrical Engineering from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 2019. From 2015 to 2017, he was with the ROK Army as a lieutenant working on C4I systems and information security. He joined Samsung Electronics, Suwon, South Korea, in 2019, as a member of engineer in charge of LTE/NR MAC scheduling algorithm. Since 2021, he has been working as a senior researcher at Korea Electronics Technology Institute (KETI), Seoul, South Korea. His current research interests include wireless communication systems optimization, signal processing and reinforcement learning.

Kiwon Kwon received B.S. and M.S. degrees in Computer Engineering from Kwangwoon University, Korea, in 1997 and 1999, respectively. He also received the Ph.D. degree in the school of Electrical & Electronics Engineering from Chung-Ang University, Korea, in 2011. In 1999, he joined in Korea Electronics Technology Institute (KETI), Korea, where he is currently the director of the Smart Network Research Center. His research interests are in the area of advanced communication system, digital twin, ICT convergence in the oceans and fisheries field.

Yangseob Kim received his Ph.D. degree in Computer Science from Chung-Ang University, Seoul, Korea in 2008. Currently, he is a principal researcher at smart network research center, Korea Electronics Technology Institute (KETI). His research interests include data management and data analysis.
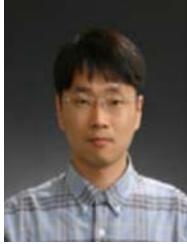
Kyungwon Park received the B.S. and M.S. degrees in Electrical Engineering from Chung-Ang University, Seoul, Korea, in 1999 and 2001, respectively, and Ph.D. degrees in the school of electrical and electronic Engineering from Chung-Ang University in 2005. Since 2005, he has been working at Korea Electronics Technology Institute (KETI), Seongnam-si, Gyeonggi-do, Korea, where he is presently a chief researcher of Smart Network Research Center. His research interests are in the areas of wireless/wired communications system design and multiple antenna systems including multiple-input multiple-output systems and smart antenna systems.

Jongho Paik received the B.S., M.S., and Ph.D. degrees in the school of Electrical and Electronic Engineering from Chung-Ang University, Seoul, Korea, in 1994, 1997, and 2007, respectively. He was a director with Advanced Mobile Research Center at Korea Electronics Technology Institute (KETI) by 2011. Since 2011, he is currently an associate professor in the department of Software Convergence, Seoul Women's University, Seoul. His research interests are in the areas of web-based service platform design, ML-based big-data analysis, UAV monitoring system, immersive video communications system design, and system architecture for realizing advanced digital communications systems and for advanced mobile broadcasting networks as well.

Sungyoon Cho received the B.S., M.S. and Ph.D. degrees in Electrical and Electronic Engineering from Yonsei University, Seoul, Korea, in 2006, 2008, and 2013, respectively. From 2013 to 2020, he was with Samsung Electronics, Korea, as a Staff Engineer to research cellular communication systems and develop 4G and 5G modem chipset. Since 2020, he has been with Korea Electronics Technology Institute (KETI) as a principal researcher to develop the advanced technologies for embedded network. His research interests are in the fundamental aspects of wireless communication and signal processing, and AI development for practical application.