International Journal of Innovative Technology and Exploring Engineering (IJITEE)
ISSN: 2278-3075 (Online), Volume-9 Issue-2, December 2019

Identifying High Significance Input Factors in Strawberry Production using Linear Model

A.B.M. Salman Rahman, Myeongbae Lee, Jangwoo Park, Yongyun Cho, Changsun Shin,

Abstract*: This study is focused on identifying the high significance of input factors in strawberry growth and production using a linear regression model. Greenhouse strawberry cultivation is increasing so fast due to the high demand for strawberry and farmers are also taking different types of technics for greenhouse cultivation to get high productions of strawberry. This study aims to increase the production of strawberries in order to maximize the profits from the cultivation of strawberries and also to fulfill the demand for strawberries. The strawberry data consist of average strawberry productions (AvgSP), electric conductivity (EC), potential of Hydrogen (PH) value, greenhouse inside temperature (Temp), greenhouse inside humidity, CO_2, nutrient solution with water, and supply of water nutrient solution. To find out the relationship among each input factor we use the correlation method and after that based on the correlation we make different types of combination of input factors. In this study, we use the linear regression method to find out the R^2 value and significance factors of different combinations of input factors. For the linear regression model we take average strawberry production as output and different combination of input factors as input. In result and discussion, we concluded the high significance input factors in strawberry growth production.

Keywords: Strawberry Growth Production, Linear Model, Correlation, Strawberry Data.

I. INTRODUCTION

At present, strawberry is regarded as one of the most popular fruits all over the world and strawberry cultivation is considered one of the healthiest economics also. Many of the countries like USA, South Korea, Turkey, Spain, Egypt, Japan, Poland, Germany and etc. yield strawberry in tonnes. Strawberry is consists of various types of nutrition's such as Vitamin C, Iron, Calcium, Magnesium, Protein, Potassium, Carbohydrate, and etc. [1].

Revised Manuscript Received on December 30, 2019.
***** Correspondence Author
Changsun Shin*, Department of Information and Communication Engineering, Sunchon National University, Suncheon-Si, Republic of Korea – 57922. Email: csshin@sunchon.ac.kr
A.B.M. Salman Rahman *, Department of Information and Communication Engineering, Sunchon National University, Suncheon-Si, Republic of Korea – 57922. Email: salman01717@gmail.com
Myeongbae Lee, Department of Information and Communication Engineering, Sunchon National University, Suncheon-Si, Republic of Korea – 57922. Email: lmb@sunchon.ac.kr
Jangwoo Park, Department of Information and Communication Engineering, Sunchon National University, Suncheon-Si, Republic of Korea – 57922. Email: jwpark@sunchon.ac.kr
Yongyun Cho, Department of Information and Communication Engineering, Sunchon National University, Suncheon-Si, Republic of Korea – 57922. Email: yycho@sunchon.ac.kr

This is the reason people are considering to take strawberry fruits in their daily routine life and the demands of strawberry fruits are increasing rapidly.

In year 2006 strawberry productions was 58,41,237 tonnes in total all over the world but in 2014 the production of strawberry was very high and it was 81,14,373 tonnes. The Demand for strawberry is continuously increasing all over the world. Therefore, farmers want to increase strawberry productions day-by-day and this is the high time to help farmers by technological support to increase strawberry productions.

Considering the high demand of strawberry, greenhouse strawberry cultivation is so much familiar nowadays. Farmers has different types of technics to cultivate strawberry in greenhouse. Some technics are used for growing plants with soil, some are used for growing plants with soil and mineral nutrient solutions, and some technics are used for growing strawberry plants without soil only with nutrient solutions flow with water and this technic called hydroponics method. In our study, we got data from hydroponics farm. For hydroponics method farmer have to consider water nutrient solution, inside temperature of greenhouse, inside humidity, electric conductivity (EC), potential of hydrogen (pH) value, and CO_2 [2].

This study focuses on identifying the high significance of input factors in strawberry growth production using linear regression model. The primary purpose of this study's to increase the strawberry production, to increase the profit of strawberry and also to fulfill the demand of strawberry. The strawberry data consists of several categories they are, average strawberry production, EC, PH value, Temperature, Humidity, CO_2, Nutrient Solution, and supply nutrient solution. Using Correlation Method, to find the relationship between each input factors and then based on the correlation result to make the different combination of input factors in linear model. Because, in linear model, average strawberry production taken as output and different combination of factors taken as input. Table 1 shows the linear model result and it contains model formula, R^2 value and significance factors. Depends on the table we concluded that identifying high significance factors in strawberry growth production in the result and discussion.

II. RELATED WORKS

Author M.L.A.T.M. Hertog and et al, were published a paper on predicting the quality of strawberries packed in changed atmospheres (cv. ' Elsanta '): an integrated development approach [3].

Narges Banaeian and et al, had been published about the potential and economic analysis of greenhouse strawberry production in Tehran, Iran [4]. Sedgwick and Philip were clearly explained the Pearson's Correlation Coefficient [5]. In statistics, linear regression is a linear approach to modeling the relationship between a scalar dependent variable y and one or more explanatory variables known as X. For more than one explanatory variable, the process is called multiple linear regression [6]. Green Bang and et al, examined the interactive experience room using infrared sensors and user's poses [7]. Jagadeesh Kakarla and et al, had been published the IDMMAC: Interference Aware Multi-Channel MAC WSAN Protocol Distributed [8]. Jin-Hee Lee and Byeong-Seok Shin, had been published a paper about the Delivery of sensors: an effective real-time delivery technique [9]. Ming Xie and et al, has been published a paper about Membrane-based sewage recovery processes: technology, pr oblems, and future path [10]. Dietmar Schwarz and et al, has been published a paper about the influence of nutrient solutions with water concentration and root pathogen (Pythium aphanidermatum) on tomato root growth and morphology [11]. Erik A and et al, has been published a paper about closed soilless growing systems: Dutch greenhouse horticulture sustainable solution [12]. Roland De Marco and et al, has been published a paper about hydroponic nutrient solutions with the flow injection potentiometric and ion-selective cobalt-wire phosphate electrode [13]. K. A. Garrett and et al has been investigated regarding the plant as the corresponding indicator of climate changes indicators [14]. Weitang Song and et al, has been published a paper, tomato fusarium wilt and its chemical control strategies in a hydroponic system [15]. Karen Smeets and et al, has been published numerical verification and critical evaluation of a hydroponic culture system for Arabidopsis thaliana [16]. Huihui Yu and et al, worked for a Chinese solar greenhouse temperature prediction and their worked has been published on the basis of LSSVM optimized by reinforced PSO [17]. Thomas Graham and et al, has been published a paper about Hydroponic tomato response to daily water ozone applications through drip irrigation [18]. Kennedy Were and et al, has been published a paper about A comparative analysis based study among different types regression models to predict and map soil organic carbon stocks across the Afromontane landscape and models are SVM, ANN and RF [19]. A.B.M.Salman Rahman and et al, has published a paper about an analysis study based on a model of linear regression for changes in fruit size in plum diseases [20]. Gianulaca Caruso and et al, has been published a paper about the Impacts of environmental cycles and nutrient solutions on alpine strawberry (Fragbaria vesca L) plant growth, development and fruit quality grown in hydroponics [21].

Based on the related works, this study focuses on identify the high significance input factors in strawberry growth production.

III. MATERIALS AND METHODS

In this study, we included strawberry data over 9 months from September 2015 to May 2016. Data is gained from a strawberry farm in South Korea named Mebangsuliang which consists of average strawberry production, EC value, pH value, inside temperature, inside humidity, amount CO_2, nutrient solution with water flow, and supply of water nutrient solutions. We used different types of sensors to collect environmental data from greenhouse and also used EC and pH sensors to collect EC and pH values.

We used correlation method to find out the relation among every input factors and we create different types of combinations of input factors based on the correlation results. The purpose of using the correlation method to reduce the multiple times for designing different types of combinations of input factors because it is easy to identify the positive correlations factors. Implements the linear regression model to identify the high significance factors with the help of R^2 values.

1.1 Correlation

Correlation is normally known as a measuring system of linear relationship between two different variants. Usually correlations is used when identified response variable is not available. The direction and the level of confidence of linear relationship between two or variables are determined by the correlation method.

The product-moment correlation of Pearson is considered as one of the most common measuring methods among all correlation method types, this method is referred to merely as the correlation or coefficient of correlation. The qualitative characteristic of the correlation are given below:

- +1 indicates a perfect positive correlations
- -1 indicates a perfect negative correlations
- 0 indicates that there is no relationship among the different variables.
- Values between -1to +1 told about the strength of the correlation.

We assumed a model:

$$r = \frac{n(\Sigma xy)-(\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2-(\Sigma x)^2][n\Sigma y^2-(\Sigma y)^2]}} \qquad (1)$$

Where, n is considered for the number of performance pairs, Σxy is considered for of the products of paired scores of sum, Σx and Σy are considered for scores of sum. $\Sigma x^2$and y^2 are considered for square scores of sum.

1.2 Linear Regression

Linear regression is a way to determine the relationship between a scalar dependent variable noted as Y and on the other hand one or more independent variables noted as X. In this case, if only one explanatory variable then its referred to as a simple linear regression and if more than one explanatory variable is referred to as multiple linear regression (the concept must be distinguished from multivariate linear regression due to multiple correlated dependent variables instead of a single scalar).

Data sets are modeled with linear predictor functions and are also derived from the data with unknown model parameters in linear regression. Such models are referred to as linear models. Linear regression appears most prominently to refer to a system in which the conditional mean of y given as the value of X is an affine function of X [6]. More specifically, linear regression tends to refer a system in which the median, or some other quintile of the conditional distribution of Y, given X is expressed as a linear function of X. Like most other forms of regression analysis, linear regression are always focused on the conditional distribution of the probability of Y given X rather than the joint distribution of the probability of Y and X, which is field for multiple analysis domain.

We assume a model

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \cdots + \beta pXp + s \qquad (2)$$

Here $\beta 0$ and $\beta 1$ are considered for two unknown constants to represent the slope and intercept and also known as coefficients or parameters, X is represents different types of variable and s is the error term

IV. RESULT AND DISCUSSION

In result and discussion, we used Pearson Correlation to find out the relation among different types parameters of strawberry data and figure 1 shows for Pearson Correlation of factors in strawberry data. In this correlation result, +1 for positive relation, -1 for negative relation, and 0 for no relation among variables or parameters. The numbers of correlation values between two variables and colours shows for identifying the positive or negative correlation. Based on the result of positive correlation to create the different combination of input factors and everything are shown in table 1. Table 1 shows for result of linear model with different combination of input factors, R^2 value and their corresponding significance factors.

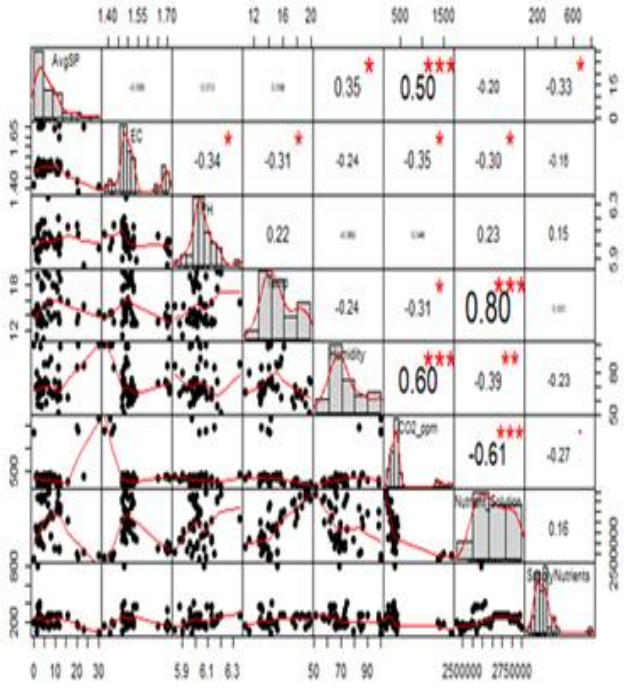

Fig.1. Result shows for Pearson Correlation of Factors.

Table 1 shows for result of linear model with different combination of input factors

Different Combination of Input Factors	R^2 Value	Significance Factors
Nutrient_Solution + Temp + PH + CO2_ppm	0.31	CO2 ppm (*)
Nutrient_Solution + Temp + Humidity + SupplyNutrients	0.27	Temp (*), Nutrient Solution (.)
Nutrient_Solution + Temp + PH + Humidity	0.22	Nutrient Solution (*), Temp (*), Humidity (.)
Nutrient_Solution + Temp + CO2_ppm + SupplyNutrients	*0.34*	CO2 ppm (*)
Nutrient_Solution + Temp + Humidity + CO2_ppm	0.31	CO2 ppm (*)
Nutrient_Solution + Temp + PH + SupplyNutrients	0.23	Nutrient Solution (*), Temp (*), Supply Nutrients (.)

If the value of R^2 is 1 or close to 1 then the model is goodness of fit. In this case, the R^2 value is below 0.5 sometimes its depends on the data we get this type of values. In significance factors, (*) for importance and (.) for less importance. Figure 2 shows one of the combination of input factors taken in table 1 and the result shows for comparison of actual data and fitted value of strawberry production data. In figure 2, x-axis have number of days and y-axis have average strawberry production (gram/plant). The blue dotted points refers for actual data and red line refers for fitted linear model.

If the red line touch each points (or closely to the points), then it is perfectly fitted in the model. In figure 2, the red line almost touches the blue points (or closely to the points) and therefore it is perfectly fitted in strawberry data.

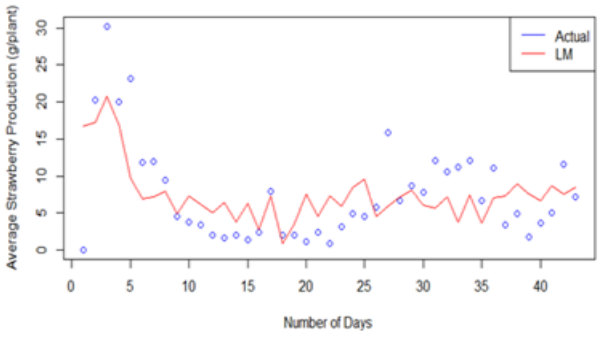

Fig. 2. Result shows for comparison of actual and fitted value of strawberry production data.

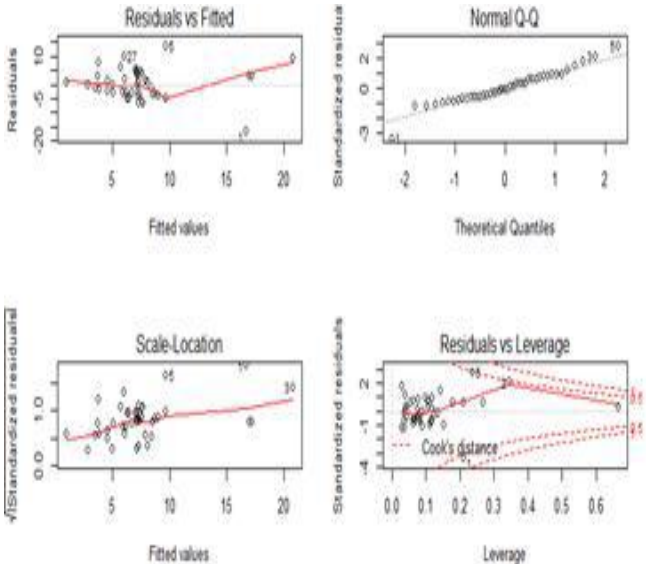

Fig. 3. Result shows for checking the residuals, normal QQ, and leverage plots.

Figure 3 diagram shows Using residuals, normal QQ plots, scale location and leverage plots to check the linear model for fitness or not. For checking residual vs fitted plot, have to check the red line in the plot is linear or non-linear. If the fitted line is non-linear then the model is goodness of fit otherwise the model is not fit. Second, in the normal QQ plot if the data points are closed to the dotted line or on the dotted line then the linear model is fitted otherwise the model is not fitted. Third, in scale spread location plot or location plot, if the data points are widespread and the red line is horizontal, the model is fit otherwise the model is not fit. Finally, the residuals vs leverage plot shows to find the outliers. Therefore, the overall fitness test of the models and the value of R^2 is 0.34.

Depends on the linear model result of table 1, we identified the high significance of input factors they are Nutrient Solution, and Temperature. Finally, these are the high significance input factors in strawberry growth production.

V. CONCLUSION

In conclusion, this study focused on identified the high significance factors in strawberry growth production using linear model. The main goal of this study is to improve the production of strawberries, increase the profit of strawberries and also meet the demand of strawberries. In this study, implemented the correlation method and linear model to identify the high significance factors are Nutrient Solution and Temperature in strawberry data. In future, we will adapt new sophisticated model for predicting strawberry growth production and estimating the demand of strawberry in the market. Future work is to add security features in supplying nutrient water supply and find any attack happened using digital forensics methods[22][23]. Further applying image processing techniques to detect diseases in strawberry growth [24][25][26].

ACKNOWLEDGMENT

This work was carried out with the support of "Cooperative Research Program for Agriculture Science &Technology Development (Project No. PJ01188605)" Rural Development Administration, Republic of Korea and, this research was supported by IPET (Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries) through Advanced Production Technology Development Program, funded by MAFRA (Ministry of Agriculture, Food and Rural Affairs) (No. 315001-5)

REFERENCES

1. Strawberry from Wikipedia, https://en.wikipedia.org/wiki/Strawberry
2. Hydroponics, https://en.wikipedia.org/wiki/ Hydroponics
3. M.L.A.T.M. Hertog, H.A.M. Boerrigter, G.J.P.M. van dan Boogaard, L.M.M. Tijskens, A.C.R. van Schaik, "*Predicting keeping quality of strawberries (cv. `Elsanta') packed under modified atmospheres: an integrated model approach"*, Postharvest Biology and Technology, vol. 15, Issue 1, pp. 1-12, (1999).
4. Narges Banaeian, Mahmoud Omid, Hojat Ahmadi, *"Energy and economic analysis of greenhouse strawberry production in Tehran province of Iran"*, vol. 52, Issue 2, pp. 1020-1025, (2011).
5. Sedgwick, and Philip, *"Pearson's Correlation Coefficient"*, British Medical Journal (online), vol. 345, July (2012).
6. Linear Regression from Wikipedia, https://en.wikipedia.org/wiki/Linear_regression
7. Green Bang, Jinsuk Yang, Kyoungsu Oh and Ilju Ko, *"Interactive Experience Room Using Infrared Sensors and User's Poses"*, JIPS, vol.13, No. 4, pp. 876-892, (2017).
8. Jagadeesh Kakarla, Banshidhar Majhi and Ramesh Babu Battula, *"IDMMAC:Interference Aware Distributed Multi-Channel MAC Protocol for WSAN"*,JIPS, vol. 13, No.5, pp.1229-1242, (2017).
9. Jin-Hee Lee and Byeong-Seok Shin, *"SensDeploy:efficient sensor deployment strategy for real- time localization"*, Human-centric Computing and Information Sciences, vol. (2017)
10. M. Xie, H. K. Shon, S. R. Gray, and M. Elimelech, *"Membrane-based processes for wastewater nutrient recovery: Technology, challenges, and future direction"*, ELSEVIER, Vol. 89, pp. 210-221, 1 Feb. 2016.
11. D Schwarz, and R Grosch, *"Influence of nutrient solutions concentration and a root pathogen (Pythium aphanidermatum) on tomato root growth and morphology"*, ELSEVIER, Vol. 97, Issue 2, pp. 109-120, Jan. 2003.
12. E. A. V. OS, *"Closed soilless growing systems: A sustainable solution for Dutch greenhouse horticulture"*, ELSEVIER, Vol. 39, Issue 5, pp. 105-112, 1999.
13. R. D. Marco, and C. Phan, *"Determination of phosphate in hydroponic nutrient solutions flow injection potentiometry and cobalt-wire phosphate ion-selective electrode"*, ELSEVIER, Vol. 60, Issue 6, pp. 1215-1221, Aug. 2003.
14. K. A. Garrett, M. Nita, E. D. De Wolf, P. D. Esker, L. Gomez-Montano, and A. H. Sparks, *"Plant pathogens as indicators of climate change"*, Climate Changes, pp. 325-338, 2016.
15. W. Song, L. Zhou, C. Yang, X Cao, L. Zhang, and X. Liu, *"Tomato Fusarium wilt and its chemical control strategies in a hydroponic system"*, ELSEVIER, Vol. 23, Issue 3, pp. 243-247, Mar. 2004.
16. K. Smeets, J. Ruytinx, F. Van Belleghem, B. Semane, D. Lin, J. Vangronsvels, and A. Cuypers, *"Critical evaluation and statistical validation of hydroponic culture system for Arabidopsis thalina"*, ELSEVIER, Vol. 46, Issue 2, pp. 212-218, Feb. 2018.
17. H. Yu, Y Chen, S G Hassan, and D Li, *"Prediction of the temperature in a Chinese solar greenhouse based on LSSVM optimized by improved PSO"*. ELSEVIER, Vol. 122, pp. 94-102, Mar. 2016.
18. T. Graham, P. Zhang, E. Woyzbyn, and M. Dixon, *"Response of hydroponic tomato to daily applications of aqueous ozone via drip irrigation"*, ELSEVIER, Vol. 129, Issue 3, pp. 464-471, Jun. 2011.
19. K. Were, D. T. Bui, O. B. Dick, and B. R. Singh, *"A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape"*, ELSEVIER, Vol. 52, pp. 394-403, May 2015.

20. A.B.M.Salman Rahman, Vasanth Ragu, Myeongbae Lee, Jangwoo Park, Yongyun Cho, Meonghun Lee, Changsun Shin, *"An Analysis Study Based on Linear Regression Model for Changes of Fruit Size over Plum Diseases"*, Journal of Knowledge Information Technology and Systems, Vol. 13, Issue 5, pp. 509-519, Oct. 2018.
21. G. Caruso, G. Villari, G. Melchionna, and S. Conti, *"Effects of cultural cycles and nutrient solutions on plant growth, yield and fruit quality of alpine strawberry (Fragaria vesca L.) grown in hydroponics"*, ELSEVIER, Vol. 129, Issue 3, pp. 479-485,27 Jun. 2011.
22. Senthilkumar, K., and Sathishkumar Easwaramoorthy. "*A Survey on Cyber Security awareness among college students in Tamil Nadu.*" In Materials Science and Engineering Conference Series, vol. 263, no. 4, p. 042043. 2017.
23. Thamburasa, Sankar, Sathishkumar Easwaramoorthy, Karrothu Aravind, S. Bharath Bhushan, and Usha Moorthy. *"Digital forensic analysis of cloud storage data in IDrive and Mega cloud drive."* In 2016 International Conference on Inventive Computation Technologies (ICICT), vol. 3, pp. 1-6. IEEE, 2016.
24. Easwaramoorthy, Sathishkumar, F. Sophia, and A. Prathik. *"Biometric Authentication using finger nails."* In 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), pp. 1-6. IEEE, 2016.
25. Easwaramoorthy, Sathishkumar, Usha Moorthy, Chunduru Anil Kumar, S. Bharath Bhushan, and Vishnupriya Sadagopan. *"Content Based Image Retrieval with Enhanced Privacy in Cloud Using Apache Spark."* In *International Conference on Data Science Analytics and Applications*, pp. 114-128. Springer, Singapore, 2017.
26. Sravankumar, B., Chunduru Anilkumar, Sathishkumar Easwaramoorthy, Somula Ramasubbareddy, and K. Govinda. *"Iterative Sharpening of Digital Images."* In Information Systems Design and Intelligent Applications, pp. 53-62. Springer, Singapore, 2019.

AUTHORS PROFILE

A B M Salman Rahman received the Bachelor degree on Electronics and Telecommunication Engineering from Southeast University, Dhaka, Bangladesh. Currently pursuing Masters leading Doctorate degree in information and Communication Engineering at Sunchon National University in South Korea. His area of interest includes Forecast Model, Ubiquitous Computing, and Big Data Processing**.**
E-mail address: salman01717@gmail.com

Myeongbae Lee completed Bachelor degree in Computer Engineering from Korea. He received Master degree on Computer Science in South Korea. And currently pursuing Doctorate degree in the Information and Communication Engineering. His area of interest includes Advanced Agriculture Technology, IT Convergence, Cloud and Ubiquitous Computing**.**
E-mail address: lmb@scnu.ac.kr

Jangwoo Park received the BS, MS and PhD degrees in Electronic engineering from Hanyang University, Seoul, Korea in 1987, 1989 and 1993, respectively. In 1995, he joined the faculty member of the Sunchon National University, where he is currently a professor in the Department of Information & Communication engineering. His research focuses on Localization and SoC and system designs and RFID/USN technologies
E-mail address: jwpark@sunchon.ac.kr

Yongyun Cho received the PhD degree in computer engineering at Soongsil University. Currently, he is an assistant professor of the Department of Information & communication engineering in Sunchon National University. His main research interests include System Software, Embedded Software and Ubiquitous Computing.
E-mail address: yycho@sunchon.ac.kr

Changsun Shin received the PhD degree in Computer Engineering at Wonkwang University. Currently, he is a professor of the Dept. of Information & Communication Engineering in Sunchon National University. His researching interests include Distributed Computing, IoT, Machine Learning, and Agriculture/ICT Convergence
E-mail address: csshin@sunchon.ac.kr