Journal of Knowledge Information Technology and Systems

ISSN 1975-7700

http://www.kkits.or.kr

# An Analysis Study Based on Linear Regression Model for Changes of Fruit Size over Plum Diseases

**A.B.M.Salman Rahman, Vasanth Ragu, Myeongbae Lee, Jangwoo Park, Yongyun Cho, Meonghun Lee[2], Changsun Shin[*]**

[1]*Department of Information & Communication Engineering, Sunchon National University,*
[2]*Dept. of Agricultural Eng., National Institute of Agricultural Sciences,*

## ABSTRACT

There are different types of diseases occur in plants and fruits due to the changes of not only the climate, weather, and seasons, but also the environmental factors like temperature, humidity, and rainfall etc. Diseases can change the structure of plants, fruits and also crops. The influence of plants and fruits disease affects our agriculture industry and agriculture sector. Diseases can interrupt our plant's growth, fruits growth, production growth, and also makes an effect on economic growth all over the world. Farmers have a lot of experience in detecting problems and also they can identify the diseases in various plants. They can easily take care or actions for environmental factorial diseases but sometimes it's not working. Technologies can support farmers to make their methods more reliable to improve farmer's crops and fruits quality for a good production. Therefore in this study, we analyse different types of plum plant diseases and then use the linear regression model to know the condition of plum fruit size using plum growth, plum length and plum width. In result and discussion, we identify the environment factorial diseases in plum and also make a decision by verifying the plum fruits size after diseases occurred in Plum Plants.



**KEYWORDS:** Linear regression, Plum diseases analysis, Plum growth analysis, Plum width analysis, Plum length analysis.

**ARTICLE INFO:** Received 3 August 2018, Revised 15 September 2018, Accepted 12 October 2018.

[*]Corresponding author is with the Department of Information & Communication Engineering, Sunchon National University, Suncheon-si, Republic of Korea,57922. E-Mail address: csshin@sunchon.ac.kr

## 1. Introduction

More than 50 to 60 percent of the people's depends on farming in most of the agriculture-based country. Farmers have a different types of diversity to select suitable Fruits, Crops, and Vegetables. However, the plantation of these ways for crops yield and quality produce is good but nowadays many kinds of diseases are presented around the world and a disease can be changed easily the structure or the function of a human, animal, and plant due to the fact of climate, weather, seasons and significance of environmental (temperature, humidity, rainfall, etc.) changed. In a word, a disease is the disorder of structure or function in a human, animal, and plant.

There are various enormously diseases present around the world. The termination of plants disease affects our agriculture industry and also agriculture sectors like plant growth, production growth, and economic growth. Farmers have lots of experience in detecting problems and identify the diseases in various plants but sometimes it's not working. It's high time, to improve farmer's method with technological support to make a huge production for our society.

This study deals to, analysis of plant disease and uses the linear regression model to analysis plum fruits size (using plum growth, plum length and plum width). Plant disease diagnosis is also an art of science. Mostly, different type's diseases occurred in plum plants like bacterial diseases, fungal diseases, miscellaneous diseases, nematodes, parasitic, phytoplasma, virus and virus-like diseases, etc. The common plum plant diseases anthracnose, bacterial canker, bacterial spot, black knot, brown rot, cherry leaf spot, crown gall, cystospora canker, peach leaf curl, plum leaf spot, plum pox, plum pockets, powdery mildew, rhizopus rot, peach scab, prunus stem pitting, and rusty spot. At first we identify, which types of diseases occur in plum plant for environmental factors and then using the linear regression model to analysis plum data for plum fruits size (using plum growth, plum length and plum width). In section 2 mentions about related works and linear regression model, in section 3 mention about materials and methods, in section 4 about result and discussion, and finally section 5 suggested the conclusion of this paper.

## 2. Related Works

In statistics, linear regression is a linear approach to modelling the relationship between a scalar response and one or more explanatory variables [1]. Gareth James and et al has been published a book an introduction to statistical learning with applications in R [2]. Sindhuja Sankaran and et al had been published a review of advanced techniques for detection plant diseases [3]. A. K. Mahlein and et al examined the development of spectral indices for detecting and identifying plant diseases [4]. J. E. Yuen and G. Hughes were published the Bayesian analysis of plant disease prediction [5]. Michael P. S. Brown and et al examined the knowledge-based analysis of microarray gene expression data by using support vector machines [6]. Pamela K. Anderson and et

al analysed the emerging infectious diseases of plants: pathogen pollution, climate change and agro technology drivers [7]. K. A. Garrett and et al had been published the plant pathogens as indicators of climate change [8]. K.A Garrett and et al has published the climate change effects on plant diseases: genomes to ecosystems [9]. S. Chakraborty and A. C. newton has been published Climate Change, plant diseases and food security: an overview [10]. Jayamala K. Patil and Raj Kumar has published advances in image processing for diction of plant diseases [11]. Vijai Singh and A.K Misra has published detection of plant leaf diseases using image segmentation, soft computing techniques, and information processing in agriculture [12]. Ilaria Perto and et al has published identificator: A web-based tool for visual plant disease identification, a proof of concept with a case study on strawberry [13]. S Chakraborty and et al analysed potential impact on plant diseases for the reason of climate changes [14]. K. A. Garrett and et al has been investigated the plant pathogens as the corresponding indicators of climate changes [15]. Fay Newbery and et al examined by modelling impacts of climate change on arable crop diseases: progress, challenges and applications [16]. Marie Launay and et al has been published the climatic indicators for crop infection risk: Application to climate change impacts on five major foliar fungal diseases in Northern France [17].

## 2.1 Linear Regression

Linear regression is an approach for modelling the relationship between a scalar dependent variable y and one or more independent variable denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear regression(This term should be distinguished from multivariate linear regression, where multiple correlated dependent variables are predicted, rather than a single scalar variable).

In linear regression, data are modelled using linear predictor functions, and unknown model parameters are estimated from the data. Such models are called linear models. Most commonly, linear regression refers to a model in which the conditional mean of y given the value of X is an affine function of X. [1] Less commonly, linear regression could refer to a model 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 all forms of regression analysis, linear regression focuses on the conditional probability distribution of y given X, rather than on the joint probability distribution of y and X, which is the domain of multivariate analysis.

We assume a model

$$Y = \beta_0 + \beta_1 X + s \qquad (1)$$

Here $\beta_0$ and $\beta_1$ are two unknown constants that represent the intercept and slope, also known as coefficients or parameters, and s is the error term.

## 3. Materials and Methods

In this study, we used the plum data in the year 2015 to May 2016, which gained from HaeRyong in Korea. In this plum data, there are various types of categories data including with date like, Mean Temperature, Minimum Temperature, Maximum Temperature, Rainfall, Wind speed, Humidity, Solar Power, Growth, Plum Width, Plum Length, Plum Size, and also the diseases are X0, X1, X2, X3, X4, X5, X6, and X7. The series of X0, X1, ···X7 and the corresponding name of the diseases are bacterial canker, anthracnose, powdery mildew, black leaf spot, depression and diseases, peach seed, sclerotinia sclerotiorum (drop), and plum scab.

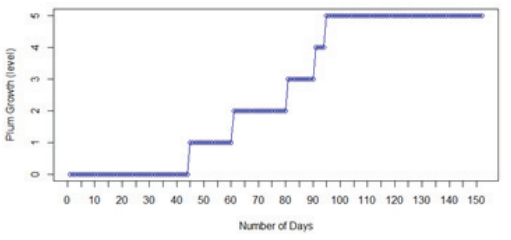


Figure 1. The diagram shows for Plum growth

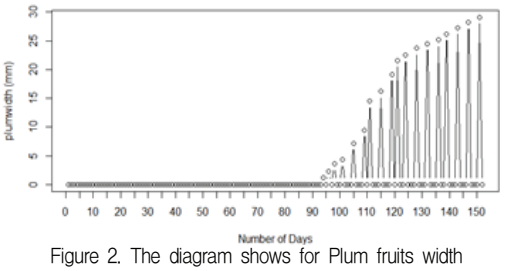


Figure 2. The diagram shows for Plum fruits width

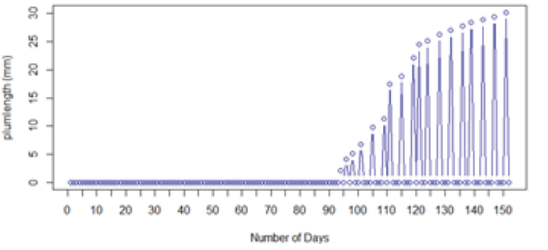


Figure 3. The diagram shows for Plum fruits length

First, we analyse how many types of diseases occurred in plum plants on the base of the environmental factor, and then we use Plum growth, Plum width, and Plum length to check the effect of diseases in plum fruits size by the linear regression model. Form <figure 1> we can know about plum growth, <figure 2> shows plum width and <figure 3> shows plum length. Finally, we make a decision is there any effect of plum diseases on increasing plum fruits size or not.

## 4. Result and Discussion

In this section, we analysis data to identify diseases to know about the types of diseases occurred in plum plants for environmental factors. Use the linear regression model to analysis plum fruits size. After all, we discussed about the effect of plum plants diseases on plum fruits growth.

### 4.1 Diseases Analysis

Nowadays different types of diseases occurred in plants so, here we find out which types of environment factorial diseases have occurred in the plum plant. After analysing

data, we find out there are three types of diseases occurred in plum plants for environmental factors and the three types of diseases are X2 = Powdery Mildew, X6 = sclerotinia sclerotiorum, X7 = Screening of scab. Figure 4 shows for identifying the various type's diseases. In Figure 4, X- axis for date and Y-axis shows for diseases, here 0 means no diseases occurred in plum plants and 1 means diseases occurred in the plum plant. In the figure, blue, red and green dot shows the diseases with the occurring date. Here blue dot shows for X6 diseases, red dot shows for X2 diseases, and green dot X7 diseases.

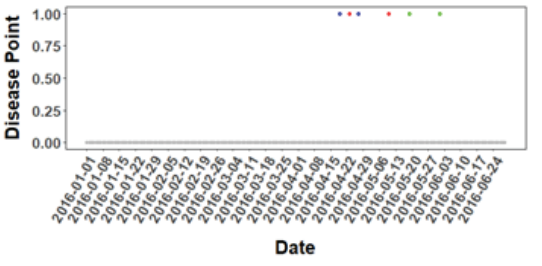


Figure 4. The diagram shows for identifying the different types of diseases.

## 4.2 Linear Regression Analysis

In this study, the main purpose of using the linear regression model to check plum fruits size increasing or not. The linear regression model, analysis data to identify plum fruits size increasing or not after occurring diseases in plum plants by using Plum Growth, Plum length, Plum Width stepwise variable selection and also checked data is fitted or not.

### 4.2.1 Analysing Plum Growth

<Figure 5> shows, X-axis for the number of plum fruits and Y-axis shows the level of plum fruits growth. In curve blue line shows for original data and read line for predicting data. <Figure 5> shows that plums are increasing or not. We divided growth in level and the levels are 0 to 1 1st, 1to2 2nd, 2 to 3 3rd, 3 to 4 4th and 4 to 5 5th. Here 1st level is 0 to 6 mm, 2nd level is 6 to 12 mm, 3rd level is 12 to 18 mm, 4th level is 18 to 24 mm and 5th level is 24 to 30mm. The purpose of levels are easily understandable and it also shows for level by level plum growth. Using linear regression, we see that plum fruits are continuously increasing for a certain period.

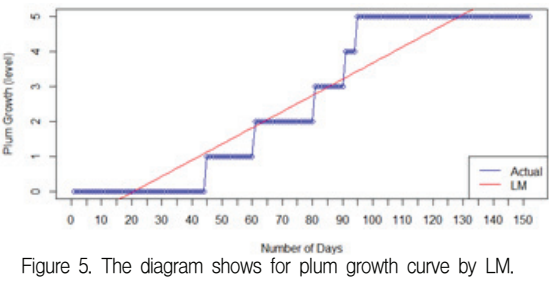


Figure 5. The diagram shows for plum growth curve by LM.

For checking the liner model goodness of fit or not for plum fruits growth data by using residuals, normal QQ plot, scale-location and leverage plots. In the residual vs. fitted plot, checking the red line is linear or non-linear. If the fitted red line is non-linear it means the model is good for this data. The second one is normal am, QQ plot and the data points are close to the dotted line or on

the dotted line then the linear model is fitted and good. The third one is scale location plot or spread location plot, if the data points are widely spread and the red line shows horizontal then the model is good and fit. Finally, the residuals vs leverage plot shows for to find the outliers.

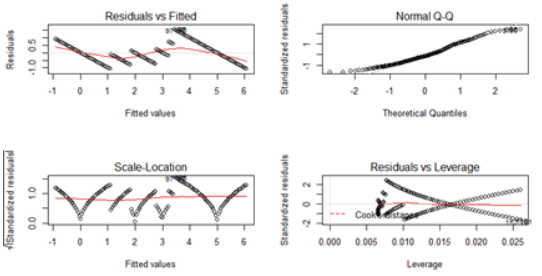


Figure 6. The diagram shows for LM fitted for plum growth.

<Figure 6> shows residual vs. fitted plot, normal QQ plot, scale location plot and residuals vs leverage plot for plum growth. In residual vs. fitted plot shows that read line is non-linear, normal QQplot shows that data points are close to the dotted line. Scale location plot or spread location plot data points are widely spread and the red line shows for horizontal. The residuals vs leverage plot show for to find the outliers. Therefore, the overall checking the model is the goodness of fit and the $R^2$ value is 0.90.

### 4.2.2 Analysing Plum Length

<Figure 7> shows X-axis for the number of plum fruits and Y-axis shows plum length. The value of plum length shows in millimetre. In curve blue line shows for original data and read line for predicting data. Using linear regression, we can easily identify that plum fruits are continuously increasing for a certain period.

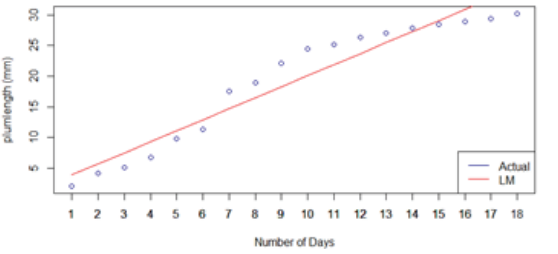


Figure 7. The diagram shows for Plum lenth curve by LM.

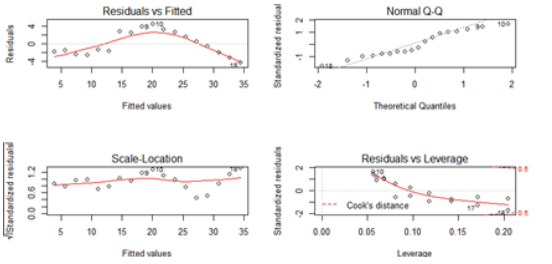


Figure 8. The diagram shows for LM fitted for plum length.

<Figure 8> shows residual vs. fitted plot, normal QQplot, scale location plot and residuals vs leverage plot for plum length. In residual vs. fitted plot shows that read line is non-linear, normal QQplot shows that data points are close to the dotted line. Scale location plot or spread location plot data points are widely spread and the red line shows for horizontal. The residuals vs leverage plot show for to find the outliers. Therefore, the overall checking the model is

the goodness of fit and the $R^2$ value is 0.92.

### 4.2.3 Analysing Plum Width

<Figure 9> shows X-axis for the number of plum fruits and Y-axis shows plum width. The values of plum width show in millimetre. In curve blue line shows for original data and read line for predicting data. Using linear regression, we can easily identify that plum fruits are continuously increasing for a certain period.

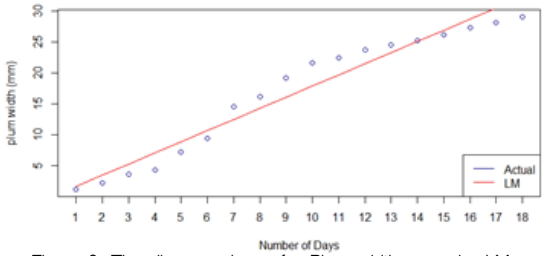


Figure 9. The diagram shows for Plum width curve by LM

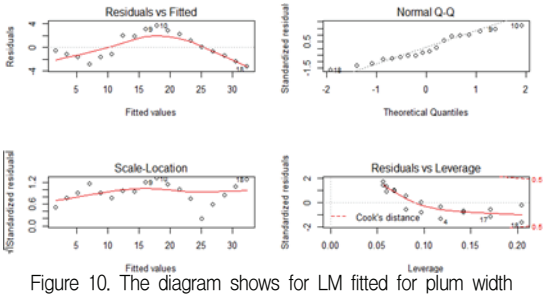


Figure 10. The diagram shows for LM fitted for plum width

<Figure 10> shows residual vs. fitted plot, normal QQ plot, scale location plot and residuals vs leverage plot for plum width. In residual vs. fitted plot shows that read line is non-linear, normal QQ plot shows that data points are close to the dotted line. Scale location plot or spread location plot data points are widely spread and the red line shows for horizontal. The residuals vs leverage plot show for to find the outliers. Therefore, the overall checking the model is the goodness of fit and the $R^2$ value is 0.94.

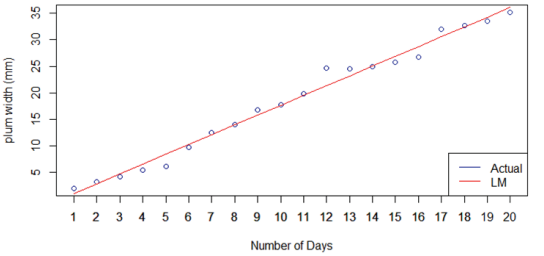


Figure 11. Diagram shows for new data plum Width curve by LM

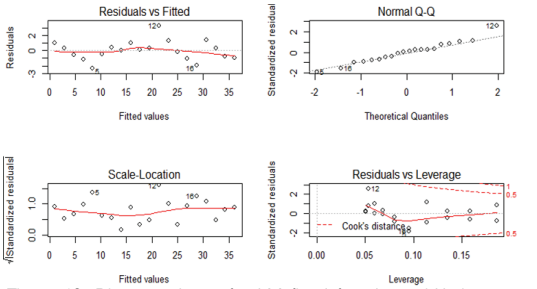


Figure 12. Diagram shows for LM fitted for plum width in new data.

<Figure 11> diagram shows for new data plum width curve by implemented the linear model. <Figure 12> diagram shows for LM fitted for plum width in new data. The purpose of new data (2018) is testing the linear model for suited same result or not. In new data, we get same type of results and so, based on these results the linear regression model is best fitted in plum data.

### 4.3 Discussion

All results and analyse provided us an acuteness between environment factorial plant diseases and fruits size. Those results about the environment factorial diseases and also about the status of plum growth, plum length and plum width. From results, we identified that three type's diseases are accrued in plum plants for environmental factor and then after analysing plum growth, length and width we see that plum size are also continuously increasing for a certain period. we get a good value for plum growth (0.90), length (0.92), width (0.94) and new data width (0.98). If the value is close to 1 means the model is good and fitted. So, the linear regression model is also fitted for analysing this plum data. After analysing all diseases, and also plum growth, plum length, plum width, we reached this incantation that environmental factorial diseases did't made a huge effect on plum fruits size.

## 5. Conclusion

This paper shows different types of environmental factorial disease in the plum plant. Diseases can really make a great harm to plants and crops. Farmers need to know about the types of harmful diseases in plants, crops, and fruits. Because of, for high production farmers need to know the diseases types and also the effect of diseases. In this study, we analysis and identified the number of disease in plum plants and also checked the plum growth, plum width, and plum length. For high production, we have to know that if the diseases occur in plum plants then the level of plum growth, size of plum width and size of plum length, are increasing continuously or not. Sometimes, plum growth, size of plum width and size of plum length are not increased and for that reason need to check the nutrition supply, and also necessary to check the environmental factors. In another, if the diseases occur in plum plants, but it still increases the level of plum growth, increases the size of plum width and increases the size of plum length then easily conclude these types of diseases won't affect in plum plants or fruits. After analysing the results of environment factorial diseases and also plum growth, plum length, and plum width, we see that environmental factorial diseases do not make a huge effect to increase plum fruits size. In future, it will be extending to strawberry diseases and find out the reasons for the cause of strawberry disease.

## References

[1] LinearRegression, *https://en.wikipedia.org/wiki/Linear_regression,* 16 Feb.2017.

[2] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learing with applications in R*, Springer New York Heidelberg Dordrecht London, 2013.

[3] S. Sankaran, A. Mishra, R. Ehsani, and C. Davis, *A review of advanced techniques for detecting plant diseases*, Computers and Electronics in Agriculture, Vol. 72, No. 1, pp. 1-13, 2010.

[4] A. K Mahlein, T. Rumpf, P. Welke, H. W. Dehne, L. Plumer, U. Steiner, and E. C. Oerke, *Development of spectral indices for detecting and identifying plant diseases*, Remote Sensing of Environment, Vol. 128, pp. 21-30, Jan. 2013.

[5] J. E. Yuen, and G. Hughes, *Bayesian analysis of plant disease prediction*, Plant Pathology, Vol. 51, Issue 4, pp. 407-412, 2002.

[6] M. P. S. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares Jr., and D. Haussler, *Knowledge-based analysis of microarray gene expression data by using support vector machines*, National Academy of Sciences, Vol. 97, No. 1, pp. 262-267, 2000.

[7] P. K. Anderson, A. A. Cunningham, N. G. Patel, F. J. Morales, P. R. Epstein, and P. Daszak, *Emerging infectious diseases of plants: pathogen pollution, climate changes and agrotechnology drivers*, Trends in Ecology & Evolution, Vol. 19, Issure 10, pp. 535-544, Oct. 2004.

[8] K. A. Garrett, M. Nita, E. D. De Wolf, P.D. Esker, L. Gomez-Montano, and A. H. Sparks, *Chapter 21-plant pathogens as indicators of climate change*, Climate Change, pp. 325-338, 2016.

[9] K. A. Garrett, S. P. Dendy, E. E. Frank, M. N. Rouse, and S. E. Travers, *Climate change effects on plant disease: Genomes to ecosystems,* Annual Review of Phytopathology, Vol. 44, pp. 489-509, Sep. 2006.

[10] S. Chakraborty, and A. C. newton, *Climate change, plant diseases and food security: an overview,* Plant Pathology, Vol. 60, pp. 2-14, Feb. 2011.

[11] J. K. Patil, and R. Kumar, *Advances in image processing for diction of plant diseases,* Advanced Bioinformatics Applications and Research ISSN 0976-2604, Vol. 2, pp. 135-141, 2011.

[12] V. Singh, and A. K Misra, *Detection of plant leaf diseases using image segmentation and soft computing techniques,* information processing in agriculture, Vol. 4, pp. 41-49, 2016.

[13] I. Pertot, T. Kuflik, I. Gordon, S. Freeman, and Y. Elad, *Identificator: A web-based tool for visual plant disease identification, a proof of concept with a case study on strawberry,* Computers and Electronics in Agriculture, Vol. 84, pp. 144-154, 2012.

[14] S. Chakraborty, A. V. Tiedemann, P. S. Teng, Climate change: potential impact on plant diseases, Environmental Pollution, Vol. 108, Issue 3, pp. 317-326, Jun. 2000.

[15] 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.

[16] F. Newbery, A. Qi, B. and DL Fitt, *Modelling impacts of climate change on arable crop diseases: progress, challenges and applications,* Current Opinion in Plant Biology, Vol. 32, pp. 101-109, Aug. 2016.

[17] M. Launay, J. Caubel, G. Bourgeois, F.

Huard, I. G. de Cortazar-Atauri, M-O. Bancal, and N. Brisson, *Climatic indicators for crop infection risk: Application to climate changes impacts on five major foliar fungal diseases in Northern France,* Agriculture, Ecosystems & Environment, Vol. 197, pp. 147-158, Dec. 2014.

## 과실 크기에 따른 매실병의 선형회귀모델 기반 분석 연구

**A.B.M. 살만 라호만[1], 바산트 라구[1], 이명배[1], 박장우[2], 조용윤[2], 이명훈[3], 신창선*[2]**

[1]*순천대학교 정보통신공학과 대학원생*
[2]*순천대학교 정보통신공학과 교수*
[3]*국립농업과학원 스마트팜개발과 연구사*

**요 약**

기후와 날씨, 계절과 온도, 습도 및 강우량과 같은 환경 인자들의 변화에 따라 나무와 과실에 다양한 종류의 병들이 발생한다. 이러한 병들은 나무나 과실의 구조 변화를 야기하여 농사 전반에 영향을 미친다. 농부는 많은 경험을 통하여 작물로부터 문제점을 찾아내고 규명할 수 있다. 환경 및 생장 데이터의 분석은 과실의 질을 향상시키데 있어서의 더욱 신뢰성있는 방법을 농부에게 제공할 수 있다. 본 연구에서는 여러 유형의 매실병들은 분석하고, 매실의 상태를 파악하기 위해 매실 크기와 같은 생장 정보를 이용한 선형회귀 모델 제안한다. 또한 매실병의 환경적 요소들을 규명하고 매실 크기에 따른 매실나무에 발생하는 병들을 검증한다.

**Acknowledgment**

This research was supported by IPET (Korea Institute of Planning and Evaluation for Technology in Food, Agriculture, Forestry and Fisheries) through Agri-Bio-industry Technology Development Program, funded by MAFRA (Ministry of Agriculture, Food and Rural Affairs) (No. 315018-3).

**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

**Vasanth Ragu** completed Master of Computer Application (MCA) degree from India. He is pursuing a Doctorate degree in information and Communication Engineering at Sunchon National University in South Korea. His area of interest includes Machine Learning Algorithms, Ubiquitous Computing, and Big Data Processing.

*E-mail address*: vasanth4224@scnu.ac.kr

**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. He area of interest includes Advanced Agriculture Technology, IT Convergence, Cloud and Ubiquitous Computing

*E-mail address*: lmb@sunchon.ac.kr

**Jangwoo Park** received the B.S., M.S. and Ph.D. 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.

*Email:* jwpark@sunchon.ac

**Yongyun Cho** received the Ph.D. 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

**Meong Hun Lee** is a researcher of National Institute of Agricultural Sciences, Republic of Korea. He received his M.S. (2006) and Ph.D. (2011) degrees from Sunchon National University. His work includes a wide variety of wireless communication, big data analytics, and applied researches. His research interests are in the fields of mobile and wireless networks, ICT convergence, and standards.

*E-mail address*: jblee@jeongbo.ac.kr

**Changsun Shin** received the Ph.D. 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