

실내 환경 평가 시 미확보 파라미터 예측을 위한 기계학습 모델에 대한 연구

정진형*, 조재현**, 김승훈**, 방소현*, 이상식***

A Study on Machine Learning Model for Predicting Uncollected Parameters in Indoor Environment Evaluation

Jin-Hyoung Jeong*, Jae-Hyun Jo**, Seung-Hun Kim**, So-Hyeon Bang*, Sang-Sik Lee***

요약 본 연구는 수집 파라미터 중 하나가 부족할 경우 다른 파라미터를 통해 부족한 파라미터를 예측하기 위한 기계 학습 모델에 대한 연구로서, 실내 환경 데이터 수집 장치를 통해 시간에 따른 온도·습도·CO2농도·광량에 대한 데이터를 수집하고, 수집한 데이터를 Matlab내 기계학습 회귀분석 기능을 통해 시간·온도·습도·CO2·광량 데이터를 예측하는 회귀모델을 만들었다. 또한 각 파라미터별로 RMSE 값이 가장 적은 3가지 모델을 선정하였으며 이에 대한 검증을 진행했다. 검증을 위해 각 파라미터로 도출된 예측모델에 테스트 데이터를 적용하여 예측치를 구했으며, 실측치와 구해진 예측치 간의 상관계수와 오차 평균을 구한 후 이를 비교하였다.

Abstract This study is about a machine learning model for predicting insufficient parameters through other parameters when one of the collected parameters is insufficient. A regression model was created to predict time, temperature, humidity, CO2, and light quantity data through the machine learning regression analysis function in Matlab. In addition, the three models with the lowest RMSE values for each parameter were selected and verified. For verification, the predicted values were obtained by applying the test data to the prediction model derived from each parameter, and the correlation coefficient and error average between the measured values and the obtained predicted values were obtained and then compared.

Key Words : Indoor environment, Machine learning, Prediction of variables, ICT, IoT

1. 서론

목적에 따라 실내 환경을 측정하고 이를 개선하려는 연구들은 과거부터 현재까지 다양한 방법으로 진행되어왔다. 최근에는 IoT 기술 및 빅데이터, 머신러닝 기법 등을 통해 실내 환경을 측정하고 보다 나은 환경으로 개선하기 위한 연구들이 발표되고 있다[1]. 실내 환경을 평가하기 위한 모델링 방법으로는 온열감 확인을

위한 PMV, 온도, CO2, 소음도, 조도를 활용하여 실내 환경 품질 평가하는 IEQ, 실내 공기 질을 측정하는 IAQ등이 대표적이다[1~7]. 또한 미세먼지[8~10], 라돈[11], 벤젠 및 톨루엔[12] 등 실내 환경을 측정하여 인체 유해 성분을 유추하는 연구들이 발표되고 있다. 이러한 모델에 활용하기 위해서는 온도, 습도, 조도, 소음도, 이산화탄소농도 등의 실내 환경 변수에 대한 기초 데이터 수집이 필요하며 각 모델을 구할 때 각 수

Following are results of a study on the "Leaders in INdustry-university Cooperation +" Project, supported by the Ministry of Education and National Research Foundation of Korea(2021DG033010102).

*Department of Biomedical IT, Catholic Kwandong University

**Department of Bio-medical Engineering, Catholic Kwandong University

***Corresponding Author : Department of Bio-medical Engineering, Catholic Kwandong University (lsskyj@cku.ac.kr)

Received October 05, 2021

Revised October 06, 2021

Accepted October 25, 2021

식에 환경변수 값이 입력되어야, 보다 유효성 있는 데이터 값을 얻을 수 있다. 본 연구에서는 수집 파라미터 중 하나가 부족할 경우 다른 파라미터를 통해 실내 환경 예측을 위한 기계학습 모델에 대한 연구로서 실내 환경 측정기기를 통해 온도, 습도, CO2 농도, 광량의 시계열 데이터를 수집한 후 Matlab 내 기계학습 회귀 분석기능을 활용하여 각 파라미터별 회귀모델을 만들고, 검증하고자 한다.

2. 실험 방법

2.1 실내 환경 데이터 수집

실내 환경에 대한 데이터는 ㈜이레이에스에서 서비스 중인 FarmingON2 라는 딸기 재배환경 측정 IoT 기기로서 딸기 하우스 내부의 재배환경을 측정하는 기기의 사양은 표 1과 같다. 데이터 수집은 제품의 관리 서버에 접속하여 진행했다.



그림 1. 실내환경 데이터 수집장치
Fig. 1. Data measurement device of the Ministry of Root Environment

습득한 데이터는 경기도 남양주시 조안면의 한 농가에 대한 데이터로, 높이 2.5m, 면적 300평 (991.74m²) 규모의 딸기 농장에서 21년 5월 1일~31일 한 달 분량이며, 시계열 데이터로서 하루 24시간 동안 분당 온도, 습도, CO₂ 농도, 광량에 대한 변화를 보여준다.

표 1. 실내 환경 데이터 수집 장치 사양
Table 1. Specifications for indoor environment data

collection device

sortation	specification
Supplier	IReis
Time	-40 ~ 124 (Accuracy : ± 0.4°C)
Temperature	0 ~ 100 (Accuracy : ± 3%)
Humidity	-3,000ppm (Accuracy : ± 3%)
CO2 concentration	400~600 Lux (Accuracy : ± 0.5°C)

2.2 각 파라미터별 회귀모델 구성

데이터를 통해 각 파라미터를 예측하는 모델을 Matlab의 기계학습 내 회귀분석 기능을 활용하여 모델링을 진행했다. 모델링 진행을 위해서 입력변수와 예측변수를 설정하는 작업이 필요하며, '온도' 예측을 위해서 입력데이터를 시간, 습도, CO2 농도, 광량 데이터로 설정했으며, '습도' 예측을 위해서는 시간, 온도, CO2 농도, 광량 데이터, 'CO2 농도' 예측을 위해서는 시간, 온도, 습도, 광량 데이터를 '광량' 데이터 예측을 위해서는 시간, 온도, 습도, CO2 농도를 '시간'에 대한 예측을 위해서는 온도, 습도, CO2 농도, 광량을 입력 데이터로 활용하였다. 내용을 정리하면 표 2와 같다.

표 2. 실내 환경 데이터 회귀모델 변수설정
Table 2. Setting variables for indoor environment data regression model.

Predictors	Input variable
Time	Temperature, Humidity, CO2 concentration, The amount of light
Temperature	Time, Humidity, CO2 concentration, The amount of light
Humidity	Time, Temperature, CO2 concentration, The amount of light
CO2 concentration	Time, Temperature, Humidity, The amount of light
The amount of light	Time, Temperature, Humidity, CO2 concentration

회귀분석을 Matlab 회귀학습 기능을 활용했으며, 조밀트리·중간트리·성긴트리로 구성된 회귀트리 모델,

선형 SVM·2차 SVM·3차 SVM·조밀 가우스 SVM·중간 가우스 SVM·성긴가우스 SVM로 구성된 서포트 벡터 머신 모델과 유리 2차·제곱지수·Matern 5/2 지수로 구성된 가우스 과정 회귀모델 가우스과정, 부스팅 트리·배깅 트리로 구성된 트리 앙상블 모델을 활용하여 데이터를 학습모델을 도출한다. 회귀트리 모델은 결정 트리 학습법을 기반으로 있으며 입력변수를 바탕으로 목표 변수의 값을 예측하는 모델을 하양식 기법을 통해 생성한다. 추리 기반 명제가 손상되어도 동작이 잘 되어 안정적이라는 장점이 있지만, 데이터가 특정 변수에 수치 수평적으로 구분되지 못할 경우 분류율이 떨어지고, 트리가 복잡해지는 문제가 발생한다[13]. 서포트 벡터 머신모델은 두 카테고리 중 어느 하나에 속한 데이터의 집합이 주어졌을 때, 주어진 데이터 집합을 바탕으로 새로운 데이터가 어느 카테고리에 속할지 판단하는 비확률적인 선형 분류모델을 만드는 기계학습법으로서 주로 패턴 인식, 자료 분석에 많이 쓰이는 지도학습 모델이다[14]. 트리 앙상블 모델은 목표 변수 예측을 위해 다수의 학습 알고리즘을 결합하여 성능을 향상시키는 앙상블 학습법 중 하나로서 적합한 자료에서 재표본 기법으로 얻은 자료를 대상으로 단일 분류자를 적합한 뒤 이를 결합하여 앙상블을 형성한다[15]. 가우스 과정회귀 모델은 비모수적 특성을 가지고 있는 회귀모델로서 비모수적 특성 때문에 학습시키는 데이터가 많아질수록 모델의 기능이 더 향상되는 장점을 가진 기계학습 모델이다[16,17]. 이후 도출된 모델 중 RMSE 값이 가장 낮은 3가지 모델을 활용하여 시간, 온도, 습도, CO2 농도, 광량 데이터의 예측값을 구하고, 각 값의 상관계수와 오차 평균을 구하여 그 학습모델의 유효성을 검증한다. RMSE값은 평균 제곱근 오차의 약자로서 추정 값 또는 모델이 예측한 값과 실제 환경에서 관찰되는 값의 차이를 다룰 때 흔히 사용되는 척도이다[18]. 회귀모델은 추후 검증을 위해 5월 데이터를 학습데이터로, 6월 1~3일간 수집한 데이터를 테스트 데이터로 활용했다. 다음 그림 2는 제시한 회귀모델 도출과정을 블록다이어그램으로 나타는 그림이다.

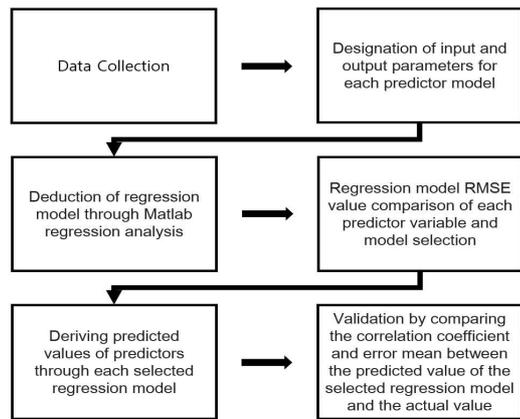


그림 2. 예측변수별 회귀모델 도출과정 블록다이어그램
Fig. 2. Block diagram of the process of deriving a regression model for each predictor variable

3. 실험 결과 및 고찰

회귀모델 제작을 위해 수집된 1달 동안 수집된 시계열 데이터의 평균값을 구했으며 이 값을 앞서 구성한 회귀모델 도출 방법에 적용하여 회귀모델을 구하고 정확도를 검증을 진행했다.

3.1 각 파라미터별 회귀모델

파라미터별 회귀모델을 도출하는 과정을 진행했으며, 진행결과는 다음과 같다.

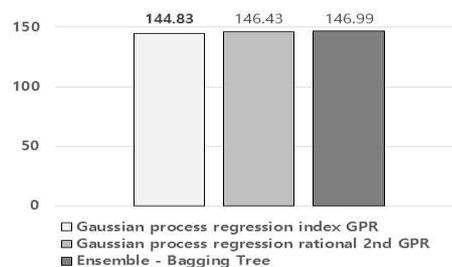


그림 3. 시간 예측모델 및 RMSE 값
Fig. 3. Time prediction model and RMSE value

시간에 대한 회귀모델 결과 RMSE값이 가장 낮았던

모델은 가우스과정 회귀지수 GPR(144.83) 였으며 그 뒤로 가우스과정 회귀지수 유리 2nd GPR(146.43), 앙상블 베깅트리(146.99) 순이었다.

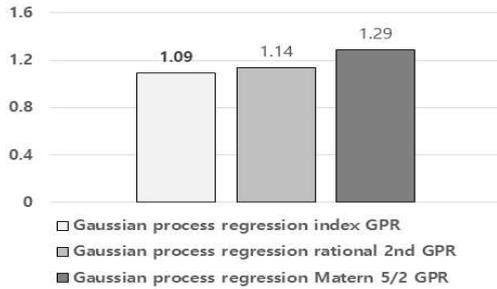


그림 4. 온도 예측모델 및 각 RMSE
Fig. 4. Temperature prediction model and each RMSE

온도는 가우스과정 회귀지수 GPR(1.09), 가우스과정 회귀지수 유리 2nd(1.14), 가우스과정 회귀 지수 Matern 5/2 GPR(1.29) 순이었다.

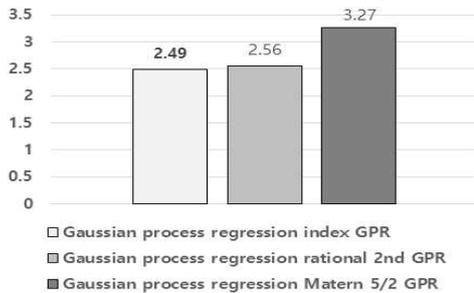


그림 5. 습도 예측모델 및 각 RMSE
Fig. 5. Humidity prediction model and each RMSE

습도는 가우스 과정 회귀지수 GPR(2.49), 가우스과정 회귀지수 유리 2nd(2.56), 가우스과정 회귀지수 Matern 5/2 GPR(3.27) 순이었다.

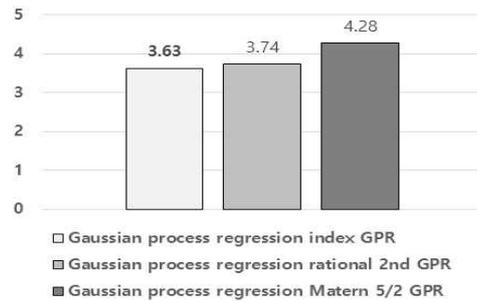


그림 6. CO2 예측모델 및 각 RMSE
Fig. 6. CO2 prediction model and each RMSE

CO2 농도는 가우스 과정 회귀지수 GPR(3.63), 가우스과정 회귀지수 유리 2nd(3.74), Matern 5/2 GPR(4.28) 순이었다.

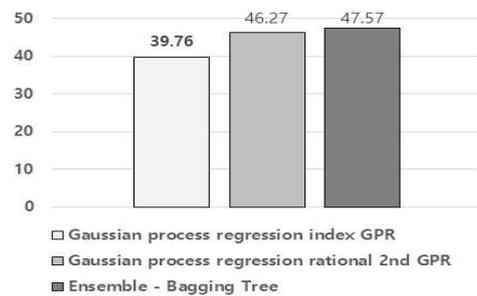


그림 7. 광량변수 예측모델 및 각 RMSE
Fig. 7. Light quantity prediction model and each RMSE

광량은 가우스 과정 회귀지수 GPR(39.76), 가우스과정 회귀지수 유리 2nd(46.27), 앙상블 베깅트리 (47.57) 순 이었다. 5개의 파라미터 모두 가우스 과정 회귀 GPR 모델의 RMSE값이 가장 낮았다.

3.2 도출된 회귀모델의 정확도 검증

도출된 회귀모델의 정확도 검증을 위해서, 회귀모델에 테스트 데이터를 입력 값으로 주었을 때 나온 예측치 데이터와 테스트 데이터 간의 상관계수와 오차 평균 및 표준편차를 구하고 이를 비교했다. 시간, 온도, 습도, CO2, 광량 모든 파라미터에서 가우스과정 회귀

지수 GPR 모델의 RMSE 값이 가장 낮은 결과를 보였으며 이는 가장 정확도가 높은 모델임을 나타낸다.

시간의 경우 상관계수가 제일 높은 모델은 가우스과정 회귀 유리 2차 GPR 로 0.65, 오차 평균값이 가장 낮은 모델은 RMSE 값이 가장 낮았던 가우스 과정 회귀 지수 GPR 모델로 0.17 ± 0.31 이었다. 그림 8, 9는 각각 시간변수에 대한 상관계수 및 오차 평균에 관한 그래프이다.



그림 8. 시간 예측모델의 상관계수
Fig. 8. Correlation coefficient of time prediction model

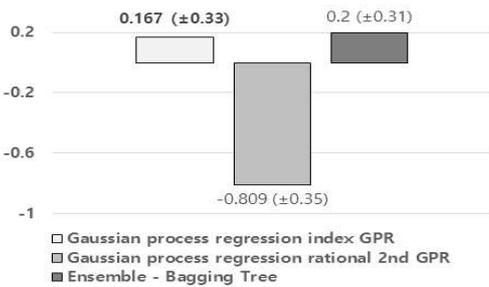


그림 9. 시간 예측모델의 오차 평균
Fig. 9. Error mean of temporal prediction model

온도의 경우 3가지 모델 모두 상관계수 0.93으로 높은 상관관계를 나타냈으며, 오차 평균값은 가우스과정 회귀 지수 Matern 5/2 GPR 모델로 -5.58 ± 6.3 이었다. 그림 10, 11은 각각 온도변수에 대한 상관계수, 오차 평균 대한 그래프이다.

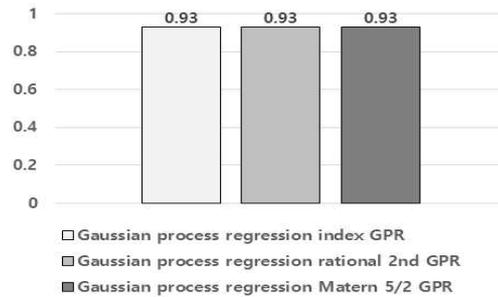


그림 10. 온도 예측모델의 상관계수
Fig. 10. Correlation coefficient of temperature prediction model

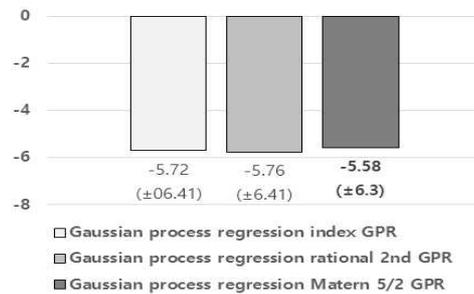


그림 11. 온도 예측모델의 오차 평균
Fig. 11. Error average of temperature prediction model

습도의 경우도 3가지 모델 모두 상관계수 0.92으로 높은 상관관계를 나타냈으며, 오차 평균값은 가우스과정 회귀 유리 2차 GPR 6.31 ± 20.21 모델이었다. 그림 12, 13은 각각 습도변수에 대한 상관계수, 오차 평균 대한 그래프이다.

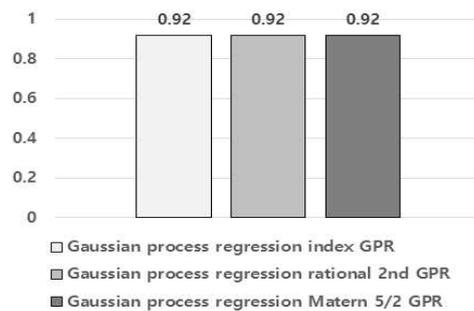


그림 12. 습도 예측모델의 상관계수
Fig. 12. Correlation coefficient of humidity prediction model

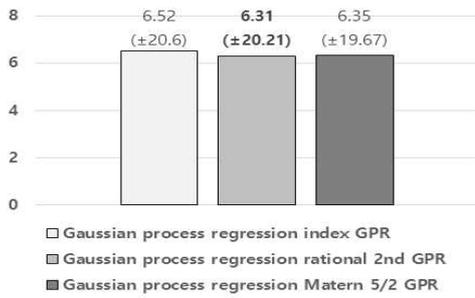


그림 13. 습도 예측모델의 오차 평균
Fig. 13. Error average of humidity prediction model

CO2의 경우 가우스과정 회귀 유리 2차 GPR 모델이 상관계수가 0.53으로 가장 높고, 오차 평균값도 -18.53 ± 19.47 로 가장 낮았다. 그림 14, 15는 각각 습도 변수에 대한 상관계수, 오차 평균 대한 그래프이다.

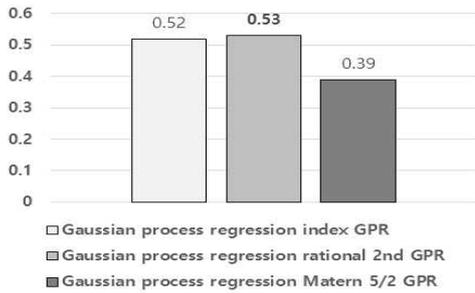


그림 14. CO2 예측모델의 상관계수
Fig. 14. Correlation coefficient of time prediction model



그림 15. CO2 예측모델의 오차 평균
Fig. 15. Error mean of temporal prediction model

광량의 경우 앙상블-배깅트리 모델이 상관계수

0.92, 데이터 오차 평균 -6.08 ± 98.9 로 값이 가장 낮았다. 그림 16, 17은 각각 습도 변수에 대한 상관계수, 오차 평균 대한 그래프이다.

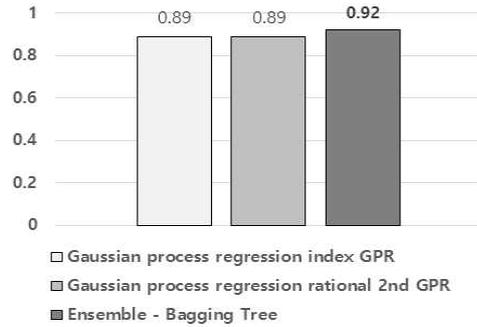


그림 16. 광량 예측모델의 상관계수
Fig. 16. Correlation coefficient of light quantity prediction model

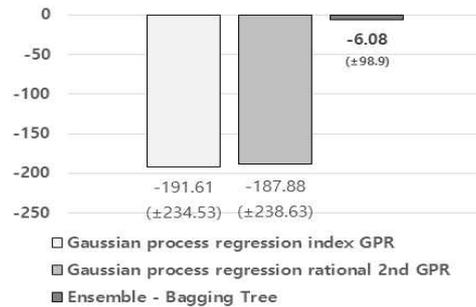


그림 17. 광량 예측모델의 오차 평균
Fig. 17. Error average of light quantity prediction model

정확도 검증 결과 회귀모델의 RMSE 값이 가장 낮아 성능이 좋을 것으로 예상되었던 가우스과정 회귀 유리 GPR 함수는 시간의 오차 평균에서만 정확도가 가장 좋았다. 이를 통해 RMSE 값이 낮다고 해서 성능이 가장 좋다고 볼 수는 없다는 것을 확인 할 수 있었다.

4. 결론

본 연구는 실내 환경 측정 장치를 통해 경기도 남양주시 조안면의 한 농가에서 21년 5월 한 달 동안 수집한 온도, 습도, CO2농도, 광량 데이터를 활용하

여 각 파라미터를 예측하는 예측모델을 만드는 연구로서, 수집 데이터를 Matlab 내 기계학습 회귀분석 기능을 활용하여 시간, 온도, 습도, CO₂, 광량 데이터를 예측하는 회귀모델을 만들었다. 또한 만들어진 모델의 검증을 위해 6월 1~3에 해당되는 테스트 데이터를 수집하고 이를 각 파라미터로 도출된 예측모델에 적용하여 상관계수와 실측치와 예측치의 오차 평균을 구하는 연구를 진행했다.

그 결과 도출된 회귀 모델은 CO₂ 농도와 시간을 제외한 파라미터는 높은 상관계수를 나타내며 오차 값이 크지 않다는 것을 확인했다. 또한 RMSE 값과 회귀모델의 성능은 반비례 하지 않는 것도 확인할 수 있었다. 본 연구를 바탕으로 CO₂ 농도와 시간 변수 회귀모델의 상관관계를 높이는 방안과 평균 오차를 전반적으로 줄이기 위한 연구를 진행할 예정이다.

REFERENCES

- [1] Francesco Salamone^{1*}, Lorenzo Belussi¹, Cristian Currò², Ludovico Danza¹, Matteo Ghellere¹, Giulia Guazzi¹, Bruno Lenzi², Valentino Megale² and Italo Meroni¹, "Integrated Method for Personal Thermal Comfort Assessment and Optimization through Users' Feedback, IoT and Machine Learning: A Case Study††", MDPI Journals Sensors Volume 18 Issue 5, 12 April 2018
- [2] JiyoungSeo, AnseopChoi, MinkiSung, "Recommendation of indoor luminous environment for occupants using big data analysis based on machine learning", Building and Environment, Volume 198, July 2021, 107835
- [3] JohannaKallioa, JaakkoTervonena, PauliRäsänen, RikuMäkynenb, JaniKoivusaaria, JohannesPeltola, "Forecasting office indoor CO₂ concentration using machine learning with a one-year dataset, Building and Environment, Volume 187, January 2021, 107409
- [4] L.T.Wong, K.W.Mui, P.S.Hui, "A multivariate -logistic model for acceptance of indoor environmental quality (IEQ) in offices", Building and Environment, Volume 43, Issue 1, January 2008, Pages 1-6
- [5] Boram Lee*, Daeyeop Lee**, Hyunkyung Ban*, Sewon Lee*, KyooSang Kim***, Kiyong Lee*†, "Evaluation of Annual Indoor Environment Quality in Hospitals using Various Comfort-related Factors", J Environ Health Sci. 2017; 43(3): 214-222
- [6] Andrei ClaudiuCosma, RahulSimha "Machine learning method for real-time non-invasive prediction of individual thermal preference in transient conditions", Building and Environment, Volume 148, 15 January 2019, Pages 372-383
- [7] Boram Lee, "Evaluation of Overall Comfort in Hospitals in Summer using Various Comforts Related Factors", Department of Environmental Health Graduate School of Public Health Seoul National University
- [8] Jung Jaeyoun, Jeong Insoo, Bao Wei "A Study on the Improvement of Indoor Environment in the Underground Parking Lot of Apartments in Jeonju City", The Journal of the Korean Rural Architecture Association, Volume 22, No.1, No. 76, pp.1-12
- [9] Young Jae Choi, Eun Ji Choi, Hye Un Cho, Jin Woo Moon, "Development of an Indoor Particulate Matter (PM_{2.5}) Prediction Model for Improving School Indoor Air Quality Environment", Korea Institute of Ecological Architecture and Environment, 2021.2, 35-40 (6 pages)
- [10] Kim Seong-Hoon, Han Gi-Tae, "A Study of Indoor Air Pollution Prediction based on LSTM Model using Fine Dust Sensor Data in Kitchen Environment", Korea Institute Of Communication Sciences, 2020.2, 1127-1128(2 pages)
- [11] Deog-Gyeong Yoon*, Jai-Won Chung**, "Prediction Model of Indoor Radon Concentration using Indoor and Outdoor Temperature, Humidity and Indoor Radon Concentration", Journal of KIIT. Vol. 18, No. 12, pp. 31-37, Dec. 31, 2020.
- [12] Kim, Da-young, Kim, Him-chan, Kim, Hyung-keun, Kim, Tae-yeon, "Health Risk Assessment by a Numerical Estimation

Model for Indoor Air Quality Evaluation of Residential Buildings", Journal of Korean Institute of Architectural Sustainable Environment and Building Systems 10(1), 2016.2, 7-14(8 pages)

- [13] Wikipedia, "Decision Tree Learning"
- [14] Wikipedia, "support vector machine"
- [15] Ah Hyoun Kim, Ji Hyun Kim, Hyun Joong Kim, "The guideline for choosing the right-size of tree for boosting algorithm", Journal of the Korean Data And Information Science Society 23(5), 2012.10, 949-959 (11 pages)
- [16] PoryaGhasemia, MasoudKarbasi, Alireza Zamani Nouric, MahdiSarai Tabrizia, Hazi Mohammad Azamathullad, "Application of Gaussian process regression to forecast multi-step ahead SPEI drought index", Alexandria Engineering Journal, Volume 60, Issue 6, December 2021, Pages 5375-5392
- [17] Marcus M. Noack1*, Gregory S. Doerk2, Ruipeng Li3, Jason K. Streit4, RichardA.Vaia4, KevinG.Yager2* & Masafumi Fukuto3*, "Autonomous materials discovery driven byGaussian process regression with inhomogeneous measurement noise and anisotropic kernels", Scientific Reports, (2020) 10:17663
- [18] Wikipedia, "Root Mean Square Deviation: RMSD"

저자약력

정진형 (Jin-Hyoung Jeong) [정회원]



<관심분야>

- 2012년 2월 : 가톨릭관동대학교 의료공학과 졸업 (학사)
 - 2014년 2월 : 가톨릭관동대학교 일반대학원 졸업 (공학석사)
 - 2017년 8월: 가톨릭관동대학교 일반대학원 졸업 (공학박사)
 - 2017년 9월 ~ 2021.02: 가톨릭관동대학교 초빙교수
 - 2021년 03월 ~ 현재: 가톨릭관동대학교 의료IT학과 조교수
- 의료 시스템, 데이터 분석, 통신, 인공지능

조재현 (Jae-Hyun Jo) [정회원]



<관심분야>

- 2015년 2월 : 가톨릭관동대학교 의료공학과 졸업 (학사)
- 2017년 2월 : 가톨릭관동대학교 일반대학원 졸업 (공학석사)
- 2020년 8월 : 가톨릭관동대학교 일반대학원 졸업 (공학박사)
- 2020년 9월 ~ 2021년 8월 : 가톨릭관동대학교 시간강사
- 2021년 9월 ~ 현재 : 가톨릭관동대학교 초빙교수

의용메카트로닉스, 디지털 신호처리, 영상처리,

김승훈 (Seung-Hun Kim) [정회원]



<관심분야>

- 2018년 2월 : 가톨릭관동대학교 의료공학과 졸업 (학사)
- 2020년 2월 : 가톨릭관동대학교 의료공학과 졸업 (공학석사)
- 2020년 3월 ~ 현재 : 가톨릭관동대학교 의료공학과 박사과정

의용메카트로닉스, IoT, u-Health, 의료영상처리

방소현 (So-Hyeon Bang) [정회원]



<관심분야>

- 2021년 2월 : 가톨릭관동대학교 의료IT학과 졸업 (학사)
- 2021년 2월 : 가톨릭관동대학교 컴퓨터공학과 석사과정

인공지능, 소프트웨어, 데이터분석

이상식 (Sang-Sik Lee) [중신회원]



<관심분야>

- 1993-2000년 LG전선(주)
- 1996-2000년 성균관대학교 박사
- 2001-2004년 (주)미도테크
- 2004-2010년 성균관대학교 연구교수
- 2011-현재 가톨릭관동대학교 의료공학과 교수

의용메카트로닉스, u-Health, 생체역학, 의용전기전자