

스마트팜 생육환경 데이터 획득 및 분석

한석호*, 장훈석**

Acquisition and Analysis of Environmental Data for Smart Farm

Seok-Ho Han*, Hoon-Seok Jang**

요약 최근 농촌문제를 해결하기 위해 주목받고 있는 스마트팜은 작물의 생육환경을 최적화하고 효율적인 관리를 통해 작물의 생산성과 품질을 높이는 기술을 의미한다. 이러한 스마트팜에 생육환경 데이터 간 관계를 분석한다면 추가적인 생산성 향상과 작물 관리가 가능할 것이다. 본 논문에서는 온도, 습도, CO₂, 토양온도, 토양습도, 일사, 토양EC, EC, pH 9개의 생육환경 데이터를 획득하고, 이를 분석하는 방법을 제안한다. 데이터 획득은 Main board와 Sensor board 간 RS-485 통신을 통해 획득 후 데이터베이스에 저장하였다. 저장된 데이터는 엑셀 시트 형식으로 내려받아 히스토그램, 데이터 차트, 상관관계 히트맵 분석을 진행하였다. 먼저 히스토그램 분석을 통해 전체, 주간, 야간 데이터의 분포를 파악하였고, 주간과 야간을 구분하여 데이터 차트 분석을 통해 월별로 평균, 중앙값, 최소값, 최대값을 파악하여 월별 데이터 변화 추이를 확인했다. 마지막으로 주간과 야간을 구분하여 상관관계 히트맵 분석을 통해 데이터의 상관관계를 파악하였다. 결과 주간에서는 온도와 토양온도, 토양EC와 EC 간에 매우 강한 양의 상관관계를 보였으며, 야간에는 온도와 토양온도, 토양EC와 EC 간에 매우 강한 양의 상관관계, 온도와 토양EC 간에 강한 음의 관계를 확인할 수 있었다.

Abstract Smart farms, which have been receiving attention as a solution to recent rural problems, refer to technologies that optimize the growing environment of crops and increase the productivity and quality of crops through efficient management. If the relationships between environmental data in smart farms are analyzed, additional productivity enhancement and crop management will be possible. In this paper, we propose a method for acquiring and analyzing nine environmental data, including temperature, humidity, CO₂, soil temperature, soil moisture, insolation, soil EC, EC, and pH. Data acquisition is done through RS-485 communication between the main board and the sensor board and stored in the database after acquisition. The stored data is downloaded in Excel sheet format and analyzed through histograms, data charts, and correlation heatmaps. First, we analyze the distribution of total, day, and night data through histogram analysis, and identify the average, median, minimum, and maximum values by month through data chart analysis separating day and night to see how the data changes by month. Finally, we analyze the correlation of the data through a correlation heatmap analysis separating day and night. The results show a very strong positive correlation between temperature and soil temperature and soil EC and EC during the day, and a very strong positive correlation between temperature and soil temperature and soil EC and EC at night, and a strong negative correlation between temperature and soil EC.

Key Words : Correlation Analysis, Data Acquisition, Data Analysis, Histogram, Smart Farm

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No. 2021-0-00751, Development of multi-dimensional visualization digital twin framework technology for displaying visible and invisible information with lower than 0.5mm precision)

*IT Application Research Center, Korea Electronics Technology Institute (KETI)

**IT Application Research Center, Korea Electronics Technology Institute (KETI)

Received April 26, 2023

Revised May 24, 2023

Accepted June 05, 2023

1. 서론

정보통신기술(ICT)의 발전으로 IoT(Internet of Things) 기반 기술의 발전은 삶의 질을 향상하고 다양한 분야로 활용 범위가 확대되고 있다. 최근 우리나라에서는 농·축산업 인구의 고령화, 인구 감소, 생산 면적의 감소, 투자위축에 따른 소득·수출·성장을 정체 등 지속가능성 위기에 처한 농촌문제의 해결방안으로 스마트팜(Smart Farm)이 큰 주목을 받고 있다 [1,2]. 스마트팜은 작물의 생육 정보와 환경정보를 기반으로 작물의 생육환경을 최적화하고, 적기 처방을 통해 노동력과 에너지, 양분 등을 효율적으로 관리하여 작물의 생산성과 품질을 높이는 농장을 뜻한다. 이러한 스마트팜에서 작물의 생육환경 데이터를 분석하여 특징과 관계를 파악한다면 추가적인 생산성 향상과 작물 관리가 가능할 것이다 [3,4]. 따라서 본 논문에서는 생육환경 데이터를 획득하고 획득한 데이터를 분석하는 연구를 진행하고자 한다.

2. 데이터 획득

2장에서는 생육환경 데이터 획득을 위해 스마트팜 테스트베드 구축 및 데이터 획득을 진행한다.

2.1 스마트팜 테스트베드 구축

스마트팜 생육환경 데이터를 획득하기 위해 9개의 센서 (온도, 습도, CO₂, 일사, 토양온도, 토양수분, 토양 EC, EC, pH)를 설치하여 테스트베드를 구축하였다. 그림 1은 구축한 스마트팜 테스트베드 사진으로 (a)는 온도, 습도, CO₂, 일사 센서 및 노드 사진, (b)는 토양온도, 토양습도, 토양 EC 센서 및 노드, (c)는 EC, pH 노드, (d)는 EC 센서, (e)는 pH 센서 사진이다.

2.2 생육환경 데이터 획득 과정

생육환경 데이터 획득 과정은 그림 2와 같은 방법으로 진행된다.

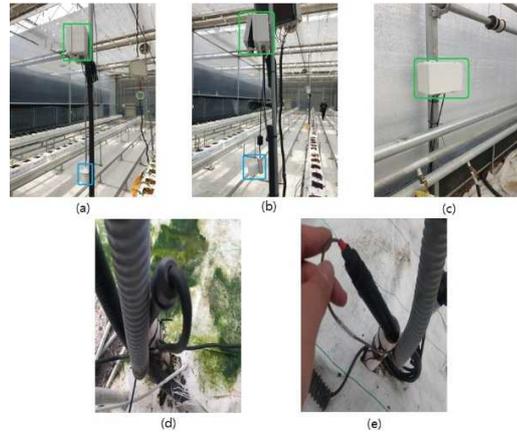


그림 1. 스마트팜 테스트 베드
Fig. 1. Testbed for Smart Farm

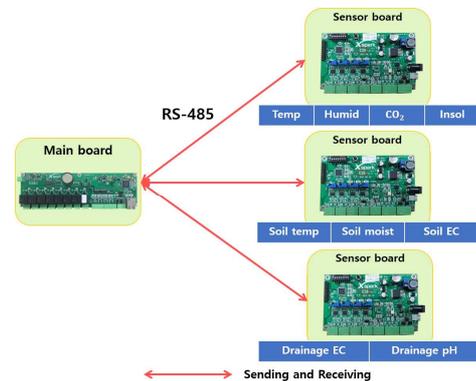


그림 2. 데이터 획득 과정
Fig. 2. Data acquisition process

Main board와 3개의 Sensor board 간 RS-485 통신을 사용하여 데이터를 송수신하는 방식으로 진행된다. 첫 번째 Sensor board에서는 온도, 습도, CO₂, 일사 총 4가지의 데이터를 획득하며, 두 번째 Sensor board에서는 토양온도, 토양수분, 토양 EC 총 3가지의 데이터를 획득하고, 마지막 Sensor board에서는 EC와 pH 2가지의 데이터를 획득한다.

2.3 생육환경 데이터베이스 구축

데이터 저장은 그림 3과 같이 진행된다. 3개의 Sensor board에서 획득한 9가지의 생육환경 데이터를 M

ain board에서 웹 서버로 HTTP REQUEST를 통해 데이터를 전송하게 된다. 웹 서버는 전송받은 데이터를 데이터베이스에 저장하게 되며, 이 과정에서 데이터 저장 성공 여부는 웹 서버에서 보내는 HTTP RESPONSE를 통해 확인할 수 있도록 하였다.

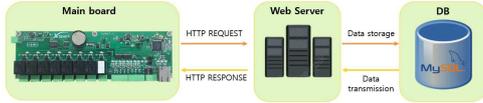


그림 3. 데이터베이스 데이터 저장 과정
Fig. 3. The process of storing data in a database

2.4 생육환경 데이터 전처리

생육환경 데이터는 6월 20일부터 11월 16일까지 1시간 주기로 획득하였다. 저장된 데이터는 데이터베이스에서 엑셀 시트로 내려받았으며, 데이터 전처리를 통해 누락된 데이터를 제외한 약 3만 개의 데이터를 획득할 수 있었다.

Temperature	Humidity	CO2	Soil Temperature	Soil Moisture	Insolation	Soil EC	Drainage EC	Drainage pH
24.52	90.54	388.25	25.69	25.6	0	0.55	3.64	7.78
23.81	91.42	402.92	25.25	25.63	0	0.55	3.63	7.8
23.37	91.58	420.17	24.78	25.7	0	0.54	3.58	7.85
23.12	91.53	436.83	24.4	25.75	0	0.55	3.6	7.9
22.98	91.93	440.08	24.11	25.8	0	0.54	3.53	7.94
22.62	92.2	426.64	23.82	25.88	8.17	0.55	3.55	7.97
22.06	91.48	382.25	23.63	25.9	67.56	0.54	3.53	8
25.29	84.48	318.67	24.24	25.95	148.07	0.53	3.49	7.97
28.06	73.19	302.58	25.73	26.05	224.1	0.54	3.46	8.07
30.87	65.44	296	27.46	25.93	299.76	0.54	3.58	8.1
31.71	63	290.33	28.69	25.78	255.92	0.55	3.57	7.95
33.55	59.2	286.42	29.99	25.48	273.14	0.55	3.68	7.89
34.24	56.58	284	30.83	25.29	274.4	0.55	3.75	7.81
35.18	59.24	280	31.7	25.13	296.62	0.55	3.8	7.7
34.86	57.78	277.92	32.1	24.97	327.11	0.56	3.9	7.57
33.61	57.38	277.67	32	25.03	314.77	0.58	3.95	7.5

그림 4. 전처리 후 데이터
Fig. 4. Preprocessed data

3. 데이터 분석

3장에서는 획득한 생육환경 데이터 분석을 진행한다. 분석은 히스토그램 분석, 데이터 차트 분석, 상관계수 히트맵 분석 순으로 진행된다.

3.1 히스토그램 분석

9개의 생육환경 데이터의 분포를 파악하기 위해 히스토그램을 사용하여 시각화하였다.

Fig 5 데이터 히스토그램에서 데이터는 좌측 위부터 온도, 습도, CO2, 토양온도, 토양수분, 일사, 토양EC, EC, pH 순서이다. 데이터 분포를 살펴보면 온도는 20~30℃에서 데이터가 많은 분포를 보였고, 습도는 80~100%에 데이터가 많은 분포를 보였다. CO2는 200~350ppm에 데이터가 고르게 분포된 것을 확인할 수 있었으며, 토양온도는 온도와 비슷하게 20~30℃에서 데이터가 많이 분포된 것을 확인할 수 있었다. 토양수분은 26~28%, 일사는 0~100w/m²에서 많은 분포를 보였으며, 토양 EC는 0.5~0.6dS/m과 0.8~0.9dS/m, EC는 3.0~3.5dS/m, 4.5~5.5d/Sm 2개의 범위에 분포된 것을 확인할 수 있었으며 pH는 6.0~7.5pH에 많이 분포된 것을 확인할 수 있었다.

자세한 분석을 위해, 일사 데이터가 자주 측정된 06:00~18:00 시간대를 주간으로, 그 외 시간대를 야간으로 설정하여 주간과 야간 데이터를 구분하여 Fig 6 주간 야간 데이터 비교 히스토그램을 시각화하였다. 온도의 경우 주간이 야간에 비해 25~30℃에 데이터가 많이 분포되어 상대적으로 야간이 온도가 더 낮은 것을 확인할 수 있었고, 습도는 야간에 90~100%에 많이 분포되어있는 야간에 비해 주간은 60~80%에도 데이터가 분포된 것을 확인할 수 있었다. CO2는 주간의 경우 150~300ppm, 야간은 200~350ppm에 분포된 것을 확인할 수 있었다. 토양온도는 주간, 야간 모두 온도와 비슷하게 분포된 것을 확인할 수 있었다.

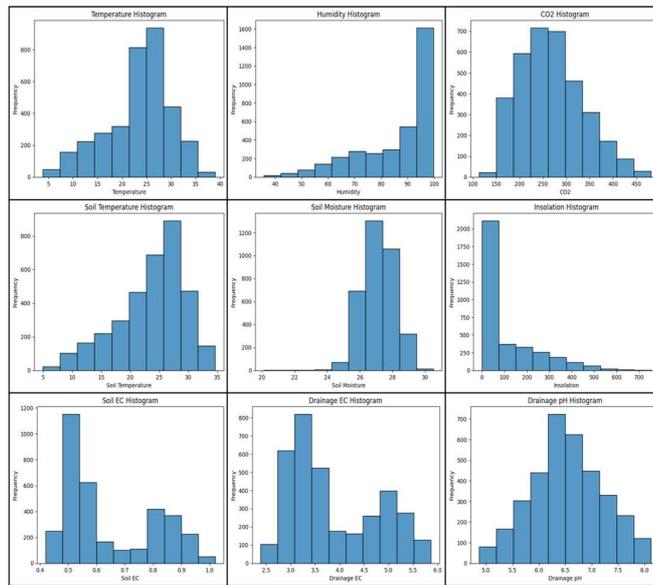


그림 5. 데이터 히스토그램

Fig. 5. Data histogram

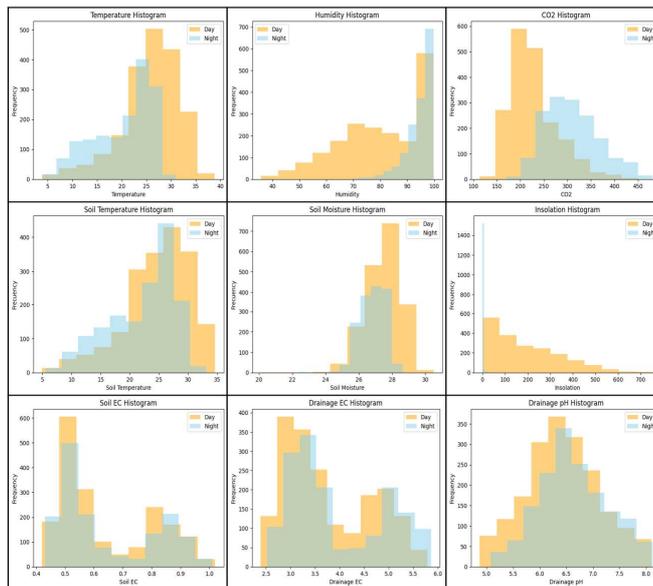


그림 6. 주간 및 야간 데이터 비교 히스토그램

Fig. 6. Histogram for day and night data

토양수분은 주간의 경우 25~30%에 매우 많은 분포가 있었으며, 야간은 25~28%에 많은 분포가 있는 것을 확인할 수 있었다. 일사는 야간의 경우 데이터가 대부분 0w/m²에 수렴했고 주간의 경우 0~700w/m²까지 분포된 것을 확인할 수 있었다. 토양 EC, EC, pH는 주간 야간과 관계없이 비슷하게 분포된 것을 확인할 수 있었다.

3.2 데이터 차트 분석

히스토그램을 이용하여 데이터 분포를 확인한 후, 주간과 야간을 구분하여 월별 평균, 중앙값, 최솟값, 최댓값을 계산하고 이를 차트로 시각화하여 데이터 변화 추이를 분석하였다. 차트에서 파란색 평균, 주황색은 중앙값, 회색은 최솟값, 노란색은 최댓값을 나타낸다.

3.2.1 온도

온도 데이터 분석 결과, 주간과 야간 평균은 6월부터 8월까지 일정하게 유지하다 이후 9월부터 11월까지 하락하는 것을 확인할 수 있었다. 최솟값과 최댓값을 비교해보면 주간이 야간에 비해 일교차가 큰 것을 확인할 수 있었다.

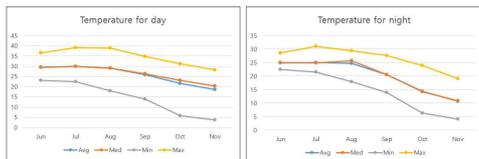


그림 7. 주야간 온도 데이터 차트
Fig. 7. Data chart for day and night temperature

3.2.2 습도

습도 데이터 분석 결과, 온도 데이터와 달리 주간과 야간의 평균과 중앙값이 월별로 일정하게 유지되는 것을 확인할 수 있었다. 주간의 월별 평균 습도는 74%~83%, 야간의 월별 평균 습도는 89%~97%로 야간이 주간에 비해 높은 평균 습도를 유지하였으며, 최솟값과 최댓값 간격 역시 야간이 주간보다 더 안정적인 습도를 유지하는 것을 확인할 수 있었다.

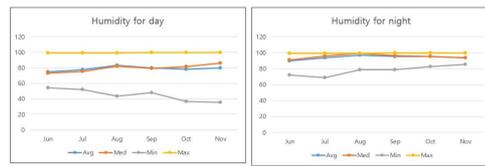


그림 8. 주야간 습도 데이터 차트
Fig. 8. Data chart for day and night humidity

3.2.3 CO2

CO2 데이터 분석 결과, 주간과 야간 평균과 중앙값이 6월부터 11월까지 서서히 하락하는 모습을 확인할 수 있었으며, 야간이 주간에 비해 높은 중앙값과 최솟값, CO2 농도를 유지하는 것을 확인할 수 있었다. 최댓값은 8월까지의 야간이 더 높았지만, 9월부터는 미세한 차이로 주간이 더 높은 것을 확인할 수 있었다.

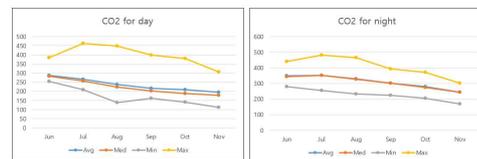


그림 9. 주야간 CO2 데이터 차트
Fig. 9. Data chart for day and night CO2

3.2.4 토양온도

토양온도 데이터 분석 결과, 온도 데이터와 비슷한 모습을 보이는 것을 확인할 수 있었다. 주간과 야간 평균은 6월부터 8월까지 일정하게 유지하다 이후 9월부터 11월까지 하락하는 것을 확인할 수 있었다.

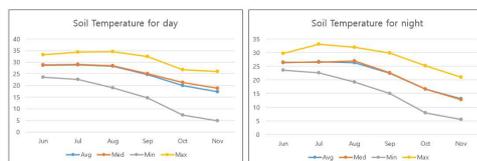


그림 10. 주야간 토양온도 데이터 차트
Fig. 10. Data chart for day and night soil temperature

3.2.5 토양수분

토양수분 데이터 분석 결과, 9월을 제외한 평균, 중

양값, 최솟값, 최댓값 모두 주간과 야간의 차이가 거의 없는 것을 확인할 수 있었으며, 데이터의 변동성이 적었다. 주간과 야간 평균 25%~28%를 유지하는 것을 확인할 수 있었다.

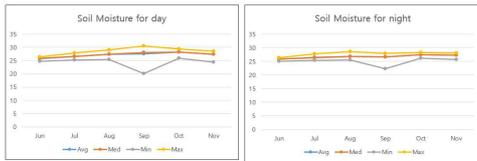


그림 11. 주야간 토양수분 데이터 차트
Fig. 11. Data chart for day and night soil moisture

3.2.6 일사

일사 데이터 분석 결과, 야간은 주간보다 평균이 매우 낮은 $0w/m^2$ 에 수렴하는 것을 확인할 수 있었다. 주간은 6월부터 9월까지 상승한 후 하락하는 모습을 확인할 수 있었다. 최솟값의 경우 주간과 야간 모든 월이 $0w/m^2$ 이었으며, 최댓값은 주간은 8월에 $753w/m^2$, 야간은 $60w/m^2$ 인 것을 확인할 수 있었다.

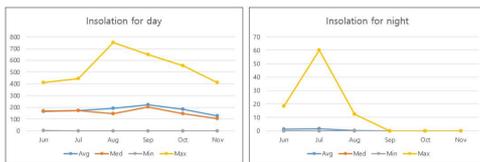


그림 12. 주야간 일사 데이터 차트
Fig. 12. Data chart for day and night insolation

3.2.7 토양EC

토양EC 데이터 분석 결과, 7월과 9월의 최댓값을 제외하면 주간과 야간의 평균, 중앙값, 최솟값, 최댓값 모두 큰 차이가 없는 것을 확인할 수 있었다. 주간과 야간 평균은 6월부터 8월까지 하락 후, 9월부터 10월까지 상승하였으며, 이후 11월에 다시 하락하는 모습을 확인할 수 있었다.

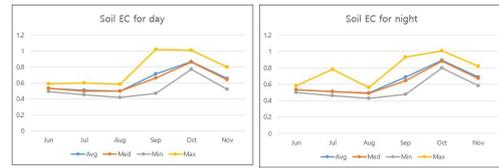


그림 13. 주야간 토양EC 데이터 차트
Fig. 13. Data chart for day and night soil EC

3.2.8 EC

EC 데이터 분석 결과, 주간과 야간의 차트가 토양EC와 비슷한 모습을 보이는 것을 확인할 수 있었다. 또한 7월의 최댓값을 제외하면 주간과 야간의 평균, 중앙값, 최솟값, 최댓값 모두 큰 차이가 없는 것을 확인할 수 있었다. 주간과 야간 평균은 6월부터 8월까지 하락 후, 9월부터 10월까지 상승하였으며, 이후 11월에 다시 하락하는 모습을 확인할 수 있었다.

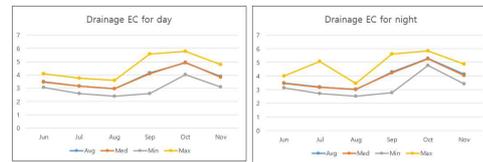


그림 14. 주야간 EC 데이터 차트
Fig. 14. Data chart for day and night EC

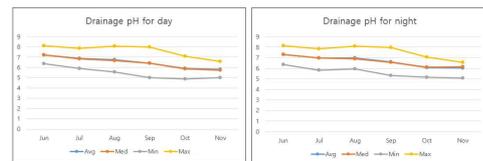


그림 15. 주야간 pH 데이터 차트
Fig. 15. Data chart for day and night pH

3.2.9 pH

pH 데이터 분석 결과, 8월 최솟값을 제외하면 앞선 토양EC, EC와 같이 주간과 야간의 평균, 중앙값, 최솟값, 최댓값 모두 큰 차이가 없는 것을 확인할 수 있었으며, 6월부터 11월까지 천천히 평균이 하락하는 것을 확인할 수 있었다.

3.3 상관관계 히트맵 분석

데이터 차트 분석 결과, 온도와 토양온도, 토양EC와 EC 등 월별 데이터 변화가 비슷한 모습을 보이는 생육 환경들 데이터가 있다는 것을 확인할 수 있었다. 이러한 관계를 분석하기 위해 데이터에 상관관계를 계산하고, 상관관계를 한눈에 파악할 수 있도록 주간과 야간을 나눠 데이터 히트맵을 그렸다.

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (1)$$

다음과 같이 정의된 계산식 (1)은 피어슨 상관계수 계산식이다. n은 데이터의 크기를 의미하며, x와 y는 각각 데이터 x와 데이터 y를 의미하며, xi와 yi는 i번째의 x와 y값을 의미한다 [5].

표 1. 상관계수 값
Table 1. Correlation coefficient value

값(절대값)	의미
0.9~1.0	매우 강한 양(음)의 상관관계
0.7~0.9	강한 양(음)의 상관관계
0.5~0.7	중간 정도의 양(음)의 상관관계
0.3~0.5	약한 양(음)의 상관관계
0.1~0.3	매우 약한 양(음)의 상관관계
0.0~0.1	거의 없는 양(음)의 상관관계

표 1은 상관계수 값 의미를 나타내는 표이다. 상관계수 값은 -1과 1 사이의 값을 가지며, 절댓값이 1에 가까울수록 강한 상관관계를 나타내고, 0에 가까울수록 두 데이터 간의 상관관계가 없다는 것을 의미한다 [6].

3.3.1 주간 상관계수 히트맵

주간 상관계수 분석 결과, 온도와 토양온도, 토양EC와 EC 간에는 매우 강한 양의 상관관계, 온도와 일사, 습도와 CO2, CO2와 pH 간에는 중간 정도의 양의 상관관계를 보였다. 습도와 일사, CO2와 토양수분, 토양온도와 토양EC, 토양온도와 EC, 토양EC와 pH, EC와 pH에는 중간 정도의 음의 상관관계가 보였으며, 그 외 데이터는 약하거나 매우 약하거나,

거의 없는 양(음)의 상관관계를 보였다.

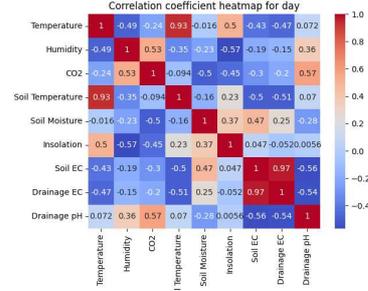


그림 16. 주간 상관계수 히트맵
Fig. 16. Correlation coefficient heatmap for day

3.3.2 야간 상관계수 히트맵

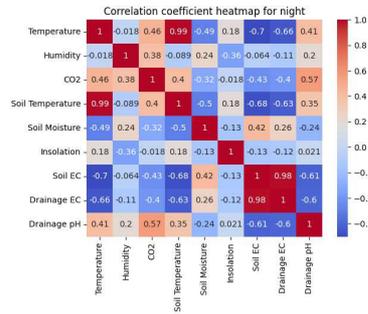


그림 17. 야간 상관계수 히트맵
Fig. 17. Correlation coefficient heatmap for night

야간 상관계수 분석 결과, 주간과 같게 온도와 토양온도, 토양EC와 EC 간에는 매우 강한 양의 상관관계를 보였고, 온도와 토양EC 간에는 강한 음의 관계를 보였다. CO2와 pH 간에는 중간 정도의 양의 상관관계, 온도와 EC, 토양온도와 토양EC, 토양온도와 EC, 토양EC와 pH, EC와 pH 간에는 중간 정도의 음의 상관관계를 보였으며, 그 외 데이터는 약하거나 매우 약하거나, 거의 없는 양(음)의 상관관계를 보였다.

4. 결론

본 논문에서는 스마트팜 테스트베드 설치 후 RS-48

5 통신을 통해 9개의 스마트팜 생육환경 데이터를 획득하고 데이터베이스에 저장하였다. 이후 데이터베이스에서 엑셀 시트 형식으로 내려받아 데이터 분석을 진행하였다. 히스토그램 분석을 통해 전체, 주간, 야간을 구분하여 데이터 분포를 확인하였고, 데이터 차트 분석을 통해 평균, 중앙값, 최소값, 최대값을 파악하여 월별 데이터 변화 추이를 확인하였다. 마지막으로 주간과 야간을 구분하여 상관관계 히트맵 분석을 통해 데이터의 상관관계를 파악하였다. 결과 주간에서는 온도와 토양온도, 토양EC와 EC 간에 매우 강한 양의 상관관계를 보였으며, 야간에는 온도와 토양온도, 토양EC와 EC 간에 매우 강한 양의 상관관계를 보였으며, 온도와 토양EC 간에 강한 음의 관계를 보였다. 본 연구에서는 단기간 데이터를 분석하여 얻은 결과를 바탕으로 진행하였기에 생육환경 데이터의 상관관계를 파악하였으나, 더 정확한 결과를 얻기 위해서는 장기간 데이터 획득 및 분석이 필요하다. 따라서 추후 연구에서는 장기간 데이터 획득을 통해 더 자세하고 정확한 분석을 진행할 예정이다.

REFERENCES

- [1] M. S. Choi, "A Study on the Efficient Implementation Method of Cloud-based Smart Farm Control System", Journal of Digital Convergence, vol. 18, no. 3, pp. 171-177, December, 2020
- [2] T. Y. Ju, S. C. Kim, E. K. Kim, "Development of a Low Cost Smart Farm System for Cultivating High Value-added Specialized Crops", The Journal of the Korea Institute of Electronic Communication Sciences, vol. 16, no. 4, pp. 743-748, August, 2021
- [3] K. B. Min, H. J. Park, "A Study on the Patent Trend of 'Smart Farm' in Domestic through Network Analysis", Journal of Korea Institute of Information, Electronics, and Communication Technology, vol. 15, no. 5, pp. 413-422, October, 2022
- [4] H. K. Choi, H. H. Ahn, Y. N. Jeong, B. K. Lee, "A Smart Farm Environment Optimization and Yield Prediction Platform based on IoT and Deep Learning", Journal of Korea Institute of Information, Electronics, and Communication Technology, vol. 12, no. 6, pp. 672-680, December, 2019
- [5] Mukaka, Mavuto M, "A guide to appropriate use of correlation coefficient in medical research.", Malawi medical journal, vol. 24, no. 3, pp. 69-71, September, 2012
- [6] Ratner, Bruce, "The correlation coefficient: Its values range between +1/-1, or do they?", Journal of targeting, measurement and analysis for marketing, vol. 17, no. 2, pp. 139-142, May, 2009

저자약력

한 석 호(Seok-Ho Han)

[정회원]



- 2023년 2월 : 원광대학교 디지털 콘텐츠공학과(학사)
- 2022년 9월~현재 : 한국전자기술연구원 연구원

<관심분야> 신호 및 영상처리, 디지털트윈, 기계 및 심층 학습

장 훈 석(Hoon-Seok Jang)

[정회원]



- 2014년 8월 : 광주과학기술원 기전공학과(공학석사)
- 2019년 2월 : 광주과학기술원 기전공학과(공학박사)
- 2020년 2월~현재 : 한국전자기술연구원 선임연구원

<관심분야> 신호 및 영상처리, 증강 및 혼합 현실, 기계 및 심층 학습