

딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조 개발

Development of deep learning structure for complex microbial incubator applying deep learning prediction result information

김 홍 직*, 이 원 복**, 이 승 호**★

Hong-Jik Kim*, Won-Bog Lee**, Seung-Ho Lee**★

Abstract

In this paper, we develop a deep learning structure for a complex microbial incubator that applies deep learning prediction result information. The proposed complex microbial incubator consists of pre-processing of complex microbial data, conversion of complex microbial data structure, design of deep learning network, learning of the designed deep learning network, and GUI development applied to the prototype. In the complex microbial data preprocessing, one-hot encoding is performed on the amount of molasses, nutrients, plant extract, salt, etc. required for microbial culture, and the maximum-minimum normalization method for the pH concentration measured as a result of the culture and the number of microbial cells to preprocess the data. In the complex microbial data structure conversion, the preprocessed data is converted into a graph structure by connecting the water temperature and the number of microbial cells, and then expressed as an adjacency matrix and attribute information to be used as input data for a deep learning network. In deep learning network design, complex microbial data is learned by designing a graph convolutional network specialized for graph structures. The designed deep learning network uses a cosine loss function to proceed with learning in the direction of minimizing the error that occurs during learning. GUI development applied to the prototype shows the target pH concentration (3.8 or less) and the number of cells (10^8 or more) of complex microorganisms in an order suitable for culturing according to the water temperature selected by the user. In order to evaluate the performance of the proposed microbial incubator, the results of experiments conducted by authorized testing institutes showed that the average pH was 3.7 and the number of cells of complex microorganisms was 1.7×10^8 . Therefore, the effectiveness of the deep learning structure for the complex microbial incubator applying the deep learning prediction result information proposed in this paper was proven.

요 약

본 논문에서는 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조를 개발한다. 제안하는 복합 미생물 배양기는 수집한 복합 미생물 데이터에 대해 복합 미생물 데이터 전처리, 복합 미생물 데이터 구조 변환, 딥러닝 네트워크 설계, 설계한 딥러닝 네트워크 학습, 시제품에 적용되는 GUI 개발 등으로 구성된다. 복합 미생물 데이터 전처리에서는 미생물 배양에 필요한 당밀, 영양제, 식물엑기스, 소금 등의 양에 대해 원-핫 인코딩을 실시하며, 배양된 결과로 측정된 pH 농도와 미생물의 셀 수에 대해 최대-최소 정규화 방법을 사용하여 데이터를 전처리한다. 복합 미생물 데이터 구조 변환에서는 전처리된 데이터를 물 온도와 미생물의 셀 수를 연결하여 그래프 구조로 변환 후, 인접 행렬과 속성 정보로 나타내어 딥러닝 네트워크의 입력 데이터로 사용한다. 딥러닝 네트워크 설계에서는 그래프 구조에 특화된 그래프 합성곱 네트워크를 설계하여 복합 미생물 데이터를 학습시킨다. 설계한 딥러닝 네트워크는 Cosine 손실함수를 사용하여 학습 시에 발생하는 오차를 최소화하는 방향으로 학습을 진행한다. 시제품에 적용되는 GUI 개발은 사용자가 선택하는 물 온도에 따라 목표하는 pH 농도(3.8 이하) 복합 미생물의 셀 수(10^8 이상)를 배양시키기 적합한 순으로 나타낸다. 제안된 미생물 배양기의 성능을 평가하기 위하여 공인시험기관에서 실험한 결과는, pH 농도의 경우 평균 3.7로, 복합 미생물의 셀 수는 1.7×10^8 으로 측정되었다. 따라서, 본 논문에서 제안한 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조의 효용성이 입증되었다.

Key words : Deep Learning, Graph Structure, Graph Convolution Network, Complex Microbe Incubator, Complex Microbial Data

I. 서론

비료와 분뇨와 같은 농가에서 많이 사용되는 물질들은 환경오염의 원인으로 꾸준히 제기되어 왔던 문제로서 악취로 인한 주민들의 민원 대상이 되고 있다. 또한, 무단 투기로 인해 방치되어 강물이나 하천에 흘러 들어가 수질오염을 일으키고 있다. 이러한 문제점을 해결하기 위해 최근 균의 생장을 억제하는 복합 미생물이 각광 받고 있다. 최근 연구에 의하면 유산균을 배양한 후의 pH 농도가 3.8 이하와 복합 미생물의 셀 수가 10^8 이상 배양되었을 때, 악취제거[1]에 뚜렷한 효과가 있는 것이 확인되었다. 그러나, 현재 사용되는 복합 미생물 배양기는 배양이 실패할 경우, 배양 시 사용된 여러 가지 재료들을 재사용하기가 어렵기 때문에 다시 배양해야 한다. 원하는 배양 조건을 만족할 때까지 배양을 반복해야 하는 이유로 많은 경제적 손실이 발생하게 된다. 따라서, 본 논문에서는 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조를 개발한다. 개발된 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기는 배양 실패 시 추가되는 경제적 손실을 절감시킬 수 있게 된다.

II. 본론

1. 복합 미생물 데이터 전처리 및 구조 변환

1.1 복합 미생물 데이터 전처리

복합 미생물 배양기로 1년동안 직접 실험을 통해 수집하여 구축한 학습 데이터를 그래프 합성곱 네트워크의 입력으로 활용할 수 있도록 데이터 전처리를 수행한다. 복합

미생물 데이터는 물 온도(temperature), 당밀(molasses), 영양제(nutrients), 식물엑기스(extract), 소금(salt) 등의 양과 pH 농도(ph), 복합 미생물의 셀 수(cell) 등이 정해진 구조에 따라 저장된다. 물 온도는 21°C부터 39°C까지 3°C 단위로 7가지, 당밀은 25kg부터 27.5kg까지 500g 단위로 6가지, 영양제는 300g부터 800g까지 250g 단위로 3가지, 식물엑기스는 200g부터 700g까지 250g 단위 3가지, 소금은 1000g부터 1500g까지 500g 단위로 2가지의 경우를 조합해 총 252개의 데이터가 포함된다. 표 1과 같이 복합 미생물 데이터에서 물 온도가 pH 농도와 복합 미생물의 셀 수에 가장 큰 영향을 끼치는 것으로 판단되며, 각 항목들에 대해 저장된 데이터의 값은 항목마다 값의 범위가 다르게 저장된다.

Table 1. Example of complex microbial data.

표 1. 복합 미생물 데이터 예시

| Temperature (°C) | Molasses (g) | Nutrients (g) | Extract (g) | Salt (g) | pH (pH) | Cell (cells) |
|------------------|--------------|---------------|-------------|----------|---------|--------------|
| 39 | 27500 | 550 | 450 | 1500 | 3.21 | 3.30E+08 |
| 39 | 25500 | 800 | 700 | 1500 | 3.71 | 2.80E+08 |
| 33 | 26000 | 550 | 450 | 1000 | 3.78 | 1.30E+08 |
| 33 | 26000 | 300 | 200 | 1000 | 3.98 | 1.00E+08 |
| 27 | 25000 | 300 | 200 | 1000 | 4.08 | 4.00E+07 |
| 27 | 25500 | 550 | 450 | 1000 | 4.34 | 4.00E+07 |
| 21 | 25000 | 300 | 200 | 1500 | 4.32 | 3.00E+07 |
| 21 | 26500 | 550 | 450 | 1500 | 4.15 | 1.70E+07 |

항목마다 값의 범위가 일정하지 않은 데이터는 딥러닝 학습에 있어 불안정한 요소로서, 값이 큰 데이터는 가중치에 큰 영향을 받고 값이 작은 데이터는 가중치에 적은 영향을 받는 것이 그 예시이다. 따라서, 당밀, 영양제, 식

* Dept. Electronic Engineering, Hanbat National University

** VISIONTECH

★ Corresponding author

E-mail : shlee@cad.hanbat.ac.kr, Tel : +82-42-821-1137

※ Acknowledgment

This research was funded by Technology Development Program (S3013451), Ministry of SMEs and Industry-university cooperation base platform (R&D), Republic of Korea.

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ICAN (ICT Challenge and Advanced Network of HRD) program (IITP-2023-RS-2022-00156212) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation)

This results was supported by "Regional Innovation Strategy (RIS)" through the National Research Foundation of Korea(NRF) funded by the Ministry of Education(MOE)(2021RIS-004)

Manuscript received Jan. 27, 2023; revised Feb. 18, 2023; accepted Feb. 26, 2023.

This is an Open-Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License(<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

물엿기스, 소금 등을 원-핫 인코딩[2] 기법을 사용하여 0 과 1로 나타내기 위해 항목마다 항목의 가지 수 만큼의 크기를 갖는 벡터로 변환한다. 한편, pH 농도와 복합 미생물의 셀 수의 경우 0과 1로 나타내는 데이터의 수 만큼의 크기를 갖는 벡터가 필요하므로 메모리의 공간 낭비가 심각하다. 따라서, 식 1과 같이 최대-최소 정규화 방법을 사용하여 데이터를 변환한다.

$$x_{normalization} = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (식 1)$$

그림 1은 전처리된 복합 미생물 데이터를 나타낸다.

| Index | Temperature | One-Hot encoding | | | | | | | | | | Normalization | | | | | |
|-------|-------------|------------------|---|---|---|---|-------------------|---|---|-----------------|---|---------------|----|------|---|-------|-------|
| | | Molasses [0 ~ 5] | | | | | Nutrients [0 ~ 2] | | | Extract [0 ~ 2] | | Salt [0 ~ 1] | pH | Cell | | | |
| 0 | 39 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.179 | 0.821 |
| 1 | 39 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0.460 | 0.692 |
| 2 | 33 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0.5 | 0.308 |
| 3 | 33 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.612 | 0.231 |
| 4 | 27 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.668 | 0.077 |
| 5 | 27 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0.814 | 0.077 |
| 6 | 21 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0.803 | 0.051 |
| 7 | 21 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0.707 | 0.018 |

Fig. 1. Example of pre-processed complex microbial data.

그림 1. 전처리된 복합 미생물 데이터 예시

1.2 복합 미생물 데이터 구조 변환

| Index | Temperature | Molasses [0 ~ 5] | | | | | Nutrients [0 ~ 2] | | | Extract [0 ~ 2] | | Salt [0 ~ 1] | pH | Cell |
|-------|-------------|------------------|---|---|---|---|-------------------|---|---|-----------------|---|--------------|-------|-------|
| 0 | 39 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0.179 | 0.821 |
| 1 | 39 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.460 | 0.692 |
| 2 | 33 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.5 | 0.308 |
| 3 | 33 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.612 | 0.231 |
| 4 | 27 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0.668 | 0.077 |
| 5 | 27 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.814 | 0.077 |
| 6 | 21 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0.803 | 0.051 |
| 7 | 21 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0.707 | 0.018 |

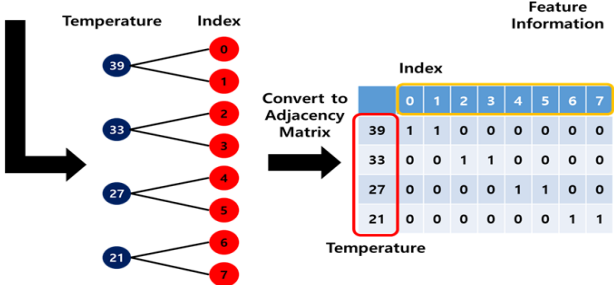


Fig. 2. Example of feature information and adjacency matrix.

그림 2. 속성 정보와 인접 행렬 예시

전처리된 복합 미생물 데이터는 그림 2와 같이 변환할 수 있다. 복합 미생물 데이터의 인덱스와 물 온도를 연결하고 연결된 두 데이터 사이의 연결 여부를 1과 0으로 나타냄으로써 그래프 구조를 형성한다. 물 온도마다 여

러 인덱스와 연결될 수 있으며, 트리 형태의 그래프 구조로 나타낼 수 있다. 인덱스는 전체 복합 미생물 데이터에서 해당 인덱스의 데이터 정보를 읽어오는 주소값을 나타내며, 주소값에 해당하는 당밀, 영양제, 식물엿기스, 소금 등의 양과 pH농도, 복합 미생물의 셀 수 등의 속성 정보를 가리킨다. 그래프 구조는 인접 행렬로 변환할 수 있으며, 학습 과정에서 속성 정보와 같이 입력 데이터로 활용된다.

2. 딥러닝 네트워크 구조 설계

본 논문에서 제안하는 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조는 그림 3과 같은 과정을 거쳐 학습을 진행한다. 진행 과정은 전처리된 데이터의 인접 행렬과 속성 정보는 그래프 합성곱 네트워크의 입력으로 활용되며, 네트워크의 출력인 확률 정보를 바탕으로 최적의 배양 조건을 찾기 위해 학습을 진행한다. 학습 과정에서 소규모 데이터 학습에서 유용한 Cosine 손실함수[5]를 사용하여 학습을 진행하며, Loss가 임계값 보다 작을 때까지 반복한다.

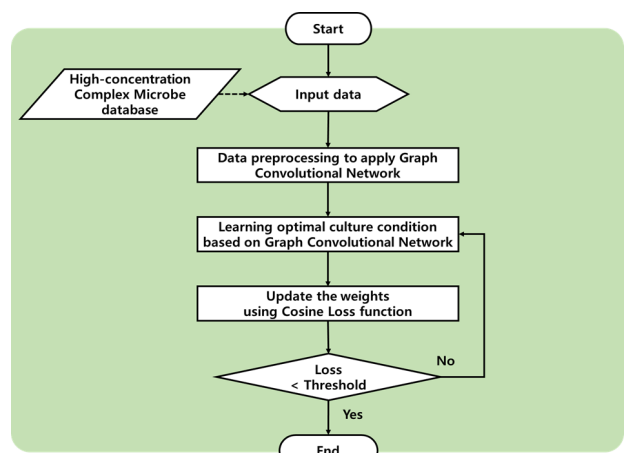


Fig. 3. Learning process of Deep Learning network.

그림 3. 딥러닝 네트워크의 학습 과정

설계한 그래프 합성곱 네트워크는 그림 4와 같으며, 각 층의 입력과 출력은 괄호 안의 숫자와 같다. 3개의 그래프 합성곱[3] 층은 입력 데이터로부터 인접 행렬과 속성 정보 데이터를 전달받아 가중치와 곱하며, 배치 정규화[4]와 ReLU 활성화 함수를 통해 다음 층으로 해당 층의 출력을 전달한다. 마지막에 위치한 완전 연결 층은 최종 그래프 합성곱 층의 출력을 1차원 벡터 형태로 변형하여 임베딩 벡터를 출력 후, 이진 크로스 엔트로피 함수를 사용하여 최종 출력인 배양 가능 확률을 도출한다.

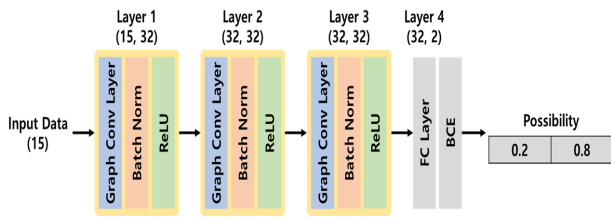


Fig. 4. Structure of designed graph convolutional network.
그림 4. 설계한 그래프 합성곱 네트워크의 구조도

그래프 합성곱 층은 그림 5와 같은 과정을 거쳐 학습이 진행된다. 각 층의 역할은 속성 정보 데이터에 가중치 곱셈, 활성화 함수, 배치 정규화 등의 딥러닝 기법을 수행하여 데이터를 업데이트하고 업데이트된 결과에 인접 행렬을 곱함으로써 행렬의 성분이 1인 부분만 값을 추출하는 결과를 얻을 수 있다. 이후, 배치 정규화 방법을 사용하여 학습 데이터에 대한 가중치 값을 고르게 분산시켜 학습의 효율을 높인다.

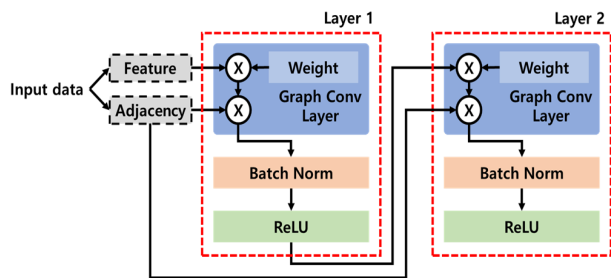


Fig. 5. Operation process of graph convolution layer.
그림 5. 그래프 합성곱 층의 연산 과정

그림 6과 같이 입력받은 그래프는 그래프 합성곱을 통해 임의의 노드의 속성 정보와 인접 행렬을 이용하여 이웃 노드에 대한 가중치를 업데이트하며, 가중치가 업데이트된 그래프 구조를 벡터화하여 완전 연결 층을 통해 최종적으로 배양 가능 확률을 출력한다.

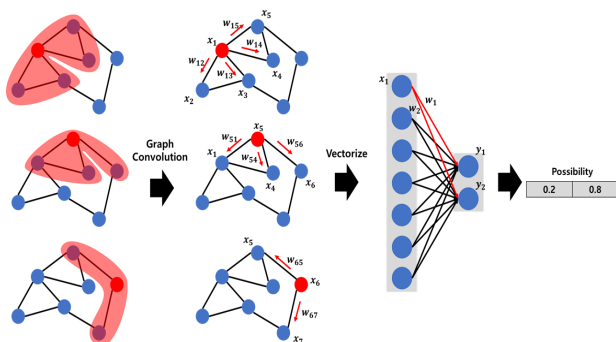


Fig. 6. Example of final output calculation process.
그림 6. 최종 출력 연산 과정 예시

3. 설계한 딥러닝 네트워크 학습

본 논문에서 제안하는 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조는 손실 함수 (Loss function)로 Cosine 손실함수를 사용하여 학습을 진행한다. 학습 시에는, 전체 데이터 10을 기준으로 8 : 1 : 1(학습 : 평가 : 테스트)로 분배한다. 학습 횟수(Epoch)는 500번으로, 학습률은 0.0005로 설정하였으며, Adam 옵티마이저를 사용하여 딥러닝 네트워크를 학습한다. 그림 7은 설계한 딥러닝 네트워크의 학습 결과 그래프를 나타낸다.

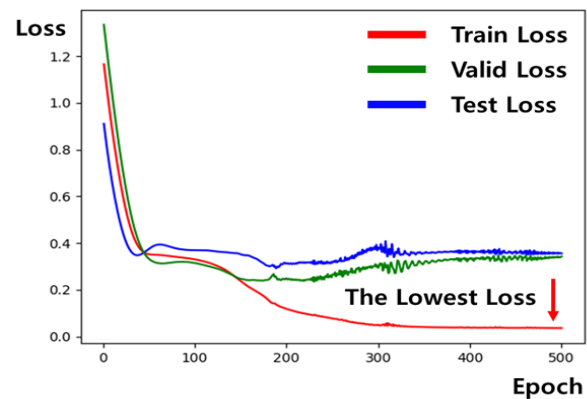


Fig. 7. Learning result graph of designed deep learning network.

그림 7. 설계한 딥러닝 네트워크의 학습 결과 그래프

4. 시제품에 적용되는 GUI 개발

시제품에 적용되는 GUI는 사용자가 선택하는 물 온도와 그에 따른 당밀, 영양제, 식물엑기스, 소금 등의 양과 재입력, 다음, PLC 전송 버튼 등의 화면구성으로 개발한다. GUI 화면에는 사용자가 선택하는 물 온도에 따라 목

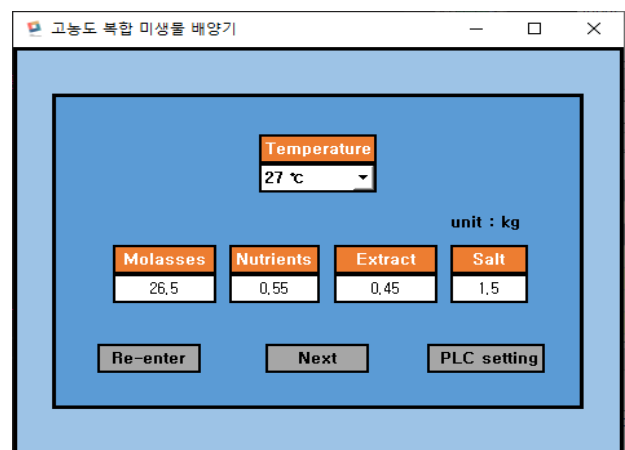


Fig. 9. Example of GUI Configuration Screen.

그림 9. GUI 구성 화면 예시

표하는 pH 농도(3.8 이하) 복합 미생물의 셀 수(10^8 이상)를 배양시키기 적합한 순으로 나타낸다. 다음에 사용자가 직접 값을 확인하고 복합 미생물 배양기의 PLC에 전송될 수 있도록 LAN 선을 이용한 TCP/IP 통신을 바탕으로 설계한다. 그림 10은 PCT 터치패널에 적용된 GUI를 나타낸다.

5. 성능 실험

5.1. 실험 방법

본 논문에서 제안한 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기의 성능을 평가하기 위하여, 공인 인증기관에서 pH 농도, 복합 미생물의 셀 수 등의 테스트를 수행하였다. 본 논문에서 제안한 딥러닝 예측 결과 정보인 다음 5개의 조건에 대해 배양하여 샘플을 채취 후에, pH 농도가 3.8 이하인지, 복합 미생물의 셀 수가 10^8 이상인지 측정하였다.

- ① 물 온도 39°C / 당밀 25kg / 영양제 300g / 식물엑기스 200g / 소금 1kg
- ② 물 온도 39°C / 당밀 27.5kg / 영양제 550g / 식물엑기스 450g / 소금 1.5kg
- ③ 물 온도 36°C / 당밀 25kg / 영양제 300g / 식물엑기스 200g / 소금 1kg
- ④ 물 온도 36°C / 당밀 26kg / 영양제 300g / 식물엑기스 200g / 소금 1kg
- ⑤ 물 온도 33°C / 당밀 26.5kg / 영양제 300g / 식물엑기스 200g / 소금 1.5kg

5.2 실험 결과

실험 결과는 표 2와 같이 pH 농도의 경우 평균 3.7로, 복합 미생물의 셀 수는 1.7×10^8 로 측정되었다. 따라서,

Table 2. Experimental results for pH concentration and high concentration complex microbial cell number.
표 2. pH 농도 및 복합 미생물 셀 수에 대한 실험 결과

| | Temperature (°C) | Molasses (kg) | Nutrients (g) | Extract (g) | Salt (kg) | pH | Cell count |
|------|------------------|---------------|---------------|-------------|-----------|-----|-------------------|
| ① | 39 | 25 | 300 | 200 | 1 | 3.7 | 1.7×10^8 |
| ② | 39 | 27.5 | 550 | 450 | 1.5 | 3.7 | 2.0×10^8 |
| ③ | 36 | 25 | 300 | 200 | 1 | 3.7 | 1.9×10^8 |
| ④ | 36 | 26 | 300 | 200 | 1 | 3.7 | 1.1×10^7 |
| ⑤ | 33 | 26.5 | 300 | 200 | 1.5 | 3.7 | 2.0×10^8 |
| Mean | | | | | | 3.7 | 1.7×10^8 |

본 논문에서 제안한 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조의 효용성이 입증되었다.

III. 결론

본 논문에서는 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조를 개발하였다. 복합 미생물 배양기로 1년동안 수집하여 구축한 학습 데이터를 바탕으로 딥러닝 네트워크를 사용하여 해당 데이터에 대해 복합 미생물 데이터 전처리, 복합 미생물 데이터 구조 변환, 딥러닝 네트워크 설계, 설계한 딥러닝 네트워크 학습, 시제품에 적용되는 GUI 개발 등으로 구성되었다. 실험 결과, pH 농도의 경우 평균 3.7로, 복합 미생물의 셀 수는 1.7×10^8 로 측정되었다. 따라서, 본 논문에서 제안한 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기를 위한 딥러닝 구조의 효용성이 입증되었다. 개발된 딥러닝 예측 결과 정보를 적용하는 복합 미생물 배양기는, 배양 실패시 추가되는 경제적 손실을 절감시킬 수 있게 되어 향후 사업성에서 밝은 전망을 나타낼 수 있을 것이다. 향후 연구과제로는 더욱 높은 정확도를 갖는 딥러닝 네트워크를 설계하여 배양기의 성능을 높이는 방법에 대한 연구가 필요하다고 사료된다.

References

- [1] Kim, C. M., et al, "The effect of feed additives that enhance the activity of intestinal microorganisms on the reduction of ammonia generated from pig manure slurry," *Proceedings of the Korean Environmental Sciences Society Conference. The Korean Environmental Sciences Society, 2020.*
- [2] Rodríguez, Pau, et al, "Beyond one-hot encoding: Lower dimensional target embedding," *Image and Vision Computing 75*, pp.21-31, 2018.
DOI: 10.48550/arXiv.1806.10805
- [3] Kipf, Thomas N., and Max Welling, "Semi-supervised classification with graph convolutional networks," *arXiv preprint*, 2016.
DOI: 10.48550/arXiv.1609.02907
- [4] Ioffe, Sergey, and Christian Szegedy, "Batch

normalization: Accelerating deep network training by reducing internal covariate shift,” *International conference on machine learning. PMLR*, 2015.

DOI: 10.48550/arXiv.1502.03167

[5] Barz, Bjorn, and Joachim Denzler, “Deep learning on small datasets without pre-training using cosine loss,” *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020. DOI: 10.48550/arXiv.1901.09054