

## 전기 차 운행 데이터를 활용한 인공지능 기반의 배터리 분석 및 평가 방법 연구

홍승모\*

### Research on artificial intelligence based battery analysis and evaluation methods using electric vehicle operation data

SeungMo Hong\*

**요약** 최근 탄소배출을 최소화하기 위해 전기자동차의 사용이 증가함에 따라 핵심 부품인 리튬이온 배터리의 상태 및 성능 분석의 중요성이 대두되고 있다. 따라서 배터리의 상태 및 성능에 영향을 줄 수 있는 배터리의 전압, 전류 및 온도뿐만 아니라 전기 자동차의 운행 데이터 및 충전 패턴 데이터를 활용한 종합적인 분석이 필요하다. 따라서 전기적 이동 수단에서 수집되는 배터리 데이터 수집 및 데이터 전처리, 단순 배터리 데이터에 추가적인 운전자 운전 습관에 대한 데이터 수집 및 전처리, 분석된 영향인자를 기반으로 인공지능 알고리즘 세부 설계 및 수정, 해당 알고리즘을 기반으로 하는 배터리 분석 및 평가 모델 설계하였다. 본 논문에서는 실시간 전기버스를 대상으로 운행 데이터와 배터리 데이터를 수집하여 Random Forest 알고리즘 활용하여 학습시킨 후, XAI 알고리즘을 통해 배터리 상태 중요 영향인자로 배터리의 상태, 운행 및 충전 패턴 데이터 등을 종합적으로 고려하여 운행 패턴에서 급가속, 급 감속, 급정지와 충 방전 패턴에서 일 주행횟수, 일일 누적 DOD와 셀 방전에서 셀 전압 차, 셀 최대온도, 셀 최소온도의 요소가 배터리 상태에 많은 영향을 미치는 인자로 확인되었으며, Random Forest 알고리즘 기반으로 배터리 분석 및 평가 모델을 설계하고 평가하였다.

**Abstract** As the use of electric vehicles has increased to minimize carbon emissions, the analyzing the state and performance of lithium-ion batteries that is instrumental in electric vehicles have been important. Comprehensive analysis using not only the voltage, current and temperature of the battery pack, which can affect the condition and performance of the battery, but also the driving data and charging pattern data of the electric vehicle is required. Therefore, a thorough analysis is imperative, utilizing electric vehicle operation data, charging pattern data, as well as battery pack voltage, current, and temperature data, which collectively influence the condition and performance of the battery. Therefore, collection and preprocessing of battery data collected from electric vehicles, collection and preprocessing of data on driver driving habits in addition to simple battery data, detailed design and modification of artificial intelligence algorithm based on the analyzed influencing factors, and A battery analysis and evaluation model was designed. In this paper, we gathered operational data and battery data from real-time electric buses. These data sets were then utilized to train a Random Forest algorithm. Furthermore, a comprehensive assessment of battery status, operation, and charging patterns was conducted using the explainable Artificial Intelligence (XAI) algorithm. The study identified crucial influencing factors on battery status, including rapid acceleration, rapid deceleration, sudden stops in driving patterns, the number of drives per day in the charging and discharging pattern, daily accumulated Depth of Discharge (DOD), cell voltage differences during discharge, maximum cell temperature, and minimum cell temperature. These factors were confirmed to significantly impact the battery condition. Based on the identified influencing factors, a battery analysis and evaluation model was designed and assessed using the Random Forest algorithm. The results contribute to the understanding of battery health and lay the foundation for effective battery management in electric vehicles.

**Key Words** : Battery condition, Electric vehicle, Lithium-ion batteries, Random forest, XAI

\*Department of Telecommunication Engineering, Soongsil University

\*Corresponding Author : Department of Telecommunication Engineering, Soongsil University (omnu@ssu.ac.kr)

Received November 07, 2023

Revised November 14, 2023

Accepted December 03, 2023

## 1. 서론

국내외에서 탄소배출을 최소화하기 위한 방법으로 전기자동차에 대한 관심이 증가하면서 핵심 부품인 리튬이온 배터리의 상태 및 성능을 분석하는 방법에 대한 연구가 활발하게 진행 되고 있다[1]. 일반적으로 배터리 상태 및 성능에 대한 분석은 BMS(Battery Management System)에서 추출되는 배터리의 전압, 전류, 온도 등의 배터리 상태 데이터만을 통해 이루어지지만, 배터리의 구조 및 원리를 고려하였을 때 배터리의 충전방식 및 충전 환경, 배터리가 방전되는 시간 동안 발생하는 방전 패턴도 많은 영향을 미친다.[2-4] 또한 전기자동차의 운행 패턴에 따라서도 영향을 받을 수 있다. 현재 인공지능 기반 배터리 분석 및 평가방법 연구가 활발히 이루어지고 있으며, 배터리 분석 분야는 전기 자동차, 재생 에너지 저장장치, 휴대용 전자 장치 등 다양한 응용분야에 매우 중요하다.[5-7]

본 논문에서는 전기 차 배터리 상태 모니터링을 위한 각종 데이터 수집 및 수집 전처리 프로세스 설계 구현 또한 추가적으로 운전자의 운전 습관에 대한 데이터 수집 및 전처리 과정을 소개한다. Machine learning 학습 중 Random forest XG Boost, Light GBM 3개의 알고리즘을 테스트 하여 학습된 알고리즘을 통해 나오는 예측 값과 실제 값들 간의 MSE (Mean Squared Errors)값들을 비교를 하였다. 예측을 위한 머신러닝 알고리즘을 선택 후 모델 학습 후 XAI(Explainable Artificial Intelligence) 알고리즘을 통해 배터리 성능을 개선시킬 주요 영향 인자의 중요도를 확인하였다.

## 2. 본론

### 2.1 전기버스 데이터 수집

전기 차 배터리의 양질의 데이터를 추출하기 위하여 데이터 수집 대상을 전기버스로 운행 데이터를 기준으로 추출하였다. 전기버스는 일반 전기 승용차보다 배터리 용량 및 배터리 셀 수가 3개 이상 많아 더 많은 양의 배터리 데이터를 수집할 수 있다. 또한 시중에

운행 중인 전기 버스의 노선버스의 경우 같은 노선을 일정한 속도로 운행하기 때문에 일반적인 전기 자동차의 비 규칙적인 운행보다 많은 종류의 변수를 통제할 수 있어 양질의 데이터를 수집할 수 있었다.

표 1. 전기버스의 운행 정보  
Table 1. The driving data of electric buses

Category	Mileage(km)			Number of round trips per day
	Accum.	6 Months	Daily Average	
Reg.A-1	102,273	34,386	247	6
Reg.A-2	38,593	28,801	202~253	4 ~ 5
Reg.B-1	108,085	38,562	220~235	4 ~ 4.5
Reg.B-2	101,563	35,876	207	3.5

데이터 표본을 전기버스로 설정한 후, LTE 통신이 가능한 OBD-2 모듈을 활용/장착하여 전기버스의 BMS (Battery Management System) 데이터와 차량의 CAN 데이터 접근을 통해 차량 기본 정보 및 차량 운행 시 데이터를 종합적으로 수집하였다. 표 1은 2개의 지역 각 2개의 노선씩 총 4개 버스에 대해서 6개월간 종합적으로 수집하였다. 실제 운영 중인 부산(Reg.A)과 서울(Reg.B) 지역의 전기 버스의 노선 6개월간 주행거리와 일평균 주행거리 1일 왕복 운행 수를 보여주고 있다. Reg.B-2는 다른 전기 버스에 비해 6개월 주행거리가 짧음을 보여주고 있다. Reg.B-2 지역버스의 일평균 충전빈도는 낮고 야간은 재충전 없이 SOC 14%까지 운행 한 데이터이다.

### 2.2 수집 데이터 전처리 프로세스 설계

인공지능 기반 배터리 분석 및 평가하기 위해서는 데이터 수집을 한 후 데이터 전처리를 해야 한다. 누락된 값, 이상 값 및 기타 불규칙성을 처리하기 위해 데이터를 정리하고 전처리를 해야 한다. 이 단계는 훈련 및 분석에 사용되는 데이터의 품질을 보장하는 필수적이다. 수집한 데이터를 EDA (Exploratory Data Analysis) 통해 군집화, 데이터 시각화, 차원 축소를 통해 분석 값을 머신러닝 학습 데이터로 가공하기 위한 데이터 전처리 프로세스를 그림 1과 같이 설계하였다. 설계 및 구축된 프로세스는 데이터 특성에 따라 실

시간으로 들어오는 데이터를 집계함수 batch 파일을 통해 분석 방법 주기에 맞추어 적재된 데이터를 생산하였으며, 생산된 데이터를 기반으로 SQL(Structured Query Language)를 활용해 Data Mart를 진행하였다. 이 프로세스를 통해 최종적으로 표 2와 같이 정제되지 않은 로우 데이터에서 분석을 위한 데이터 70개를 추출하였다.

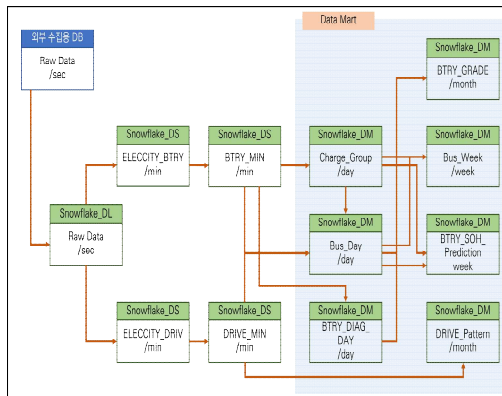


그림 1. 데이터 전처리 프로세스 흐름도  
Fig. 1. Data Preprocessing Process flow chart

표 2. 실험 로우 데이터 세트  
Table 2. The raw data set for analyze

Data Categories		Sum.
Driving Data	Vehicle Information ID Date Time et c.	24
Battery Data	Vehicle Information ID Cell voltage Pack current Battery Temperature etc.	46 types

그림 1의 데이터 전처리 프로세스를 통해 차량 및 배터리 기본정보, 배터리 초 단위 데이터, 배터리 1분 집계, 운전 데이터 1분 집계, 배터리 충전 정보, 배터리 및 운전 데이터 1일 집계, 배터리 등급 산정, 배터리 수명 예측을 위한 주간 집계, 주간 SOH (State of Health) 감소량 등의 데이터로 전 처리하였다.

### 2.3 수집 데이터 전처리 프로세스 분석 결과

통신 모듈을 통해 수집한 각 종 데이터 (Raw Data) 그리고 이를 전 처리하여 생성한 Data Mart를 EDA 분석하여, 각 데이터간의 상관관계 및 배터리 성능 평가에 주요 영향을 끼치는 인자를 도출하였다. 표 3은 EDA 분석 간 도출한 배터리 상태 주요 영향 인자를 나타내었다.

표 3. EDA 분석 간 도출한 배터리 상태 주요 영향인자  
Table 3. Main influencing factor of battery condition derived from EDA analysis

Spec.	Data list	Description
Route Info.	DISTANCE	Total traveled distance (Weekly)
Charging/Discharging Pattern	DOD	Daily average Depth of Discharge
	C_RATE	Daily average discharge C-RATE
	DCH_DOD	Daily Accumulated Discharge Amount
	DRIVE_POWER_SUM	Cumulative Drive System Power Consumption
	DRIVE_POWER_P	Average Drive System Power Consumption
Driving Pattern	SOC	Average State of Charge
	SPEED	Daily Average Speed (above 0)
	DIFF_SPEED	Daily Average Speed Variation (second by second)
	ACC_POS	Daily average accelerator position
	ACC	Rapid Acceleration Count
	DEC	Rapid Deceleration Count
	DEC_STOP	Hard Stop Count
Cell Status	EVCU	Heater on Count (second by second)
	CELL_DIFF	Daily Average Cell Voltage Difference (Second-by-Second)
	CELL_MXTMP	Cell Max average temperature
Target	CELL_MNTMP	Cell Min average temperature
	Target	DIFF_SOH_CH
Target	Target	DIFF_SOH_DCH

수집되고 전처리된 데이터에서 배터리 상태 및 수명에 영향을 주는 영향 인자는 다음과 같다.

총 방전패턴으로 일평균 DOD (Depth of Disc-charge, Max-Min), 일평균 방전 C-RATE, 일 누적 방전량, 평균 구동계 소요전력, 누적 구동계 소요전력과 평균 SOC(State of Charge)인자를 추출하였으며, 운전 패턴으로 일평균 속도, 일평균 초 단위 속도변화, 일평균 Accelerator position, 급가속 횟수, 급 감속 회수, 급정지 횟수, Heater on Count(초 단위) 요소를 추출하였다. 셀 상태로 Cell Max 평균 온도, Cell Min 평균 온도, 일평균 초단위 셀 전압 차 인자를 추출하였다. Target으로 충전시점 계산 SOH 차이, 방전 시점 계산 SOH 차이 인자를 추출하였다. 배터리 총 방전 패턴에 따른 배터리 수명 영향 인자와 운전자의 운전 패턴에 인자를 개선하여 배터리 수명을 개선시킬 인자로 확인하였다.

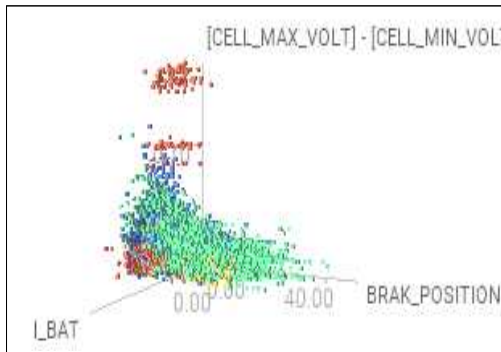


그림 2. 셀 전압차와 브레이크 간 상관관계  
Fig. 2. Correlation between cell voltage difference and break

또한 배터리 수명에 영향을 미치는 인자를 식별하기 위해 전기 버스 운행 중 기록 관리되고 있는 인자 별로 차량 간, 운행 기간 내 유의미한 차이 변화가 있었는지를 데이터를 탐색하여 분석 하였다. 그림 2는 셀 전압 차와 브레이크 간 상관관계를 나타낸 것이다. 운전 중 브레이크는 배터리의 총 방전이 급격하게 교차되어 셀 전압차를 크게 하였다. 따라서 셀 전압차가 수명에 미치는 영향 인자로 추출하였다.

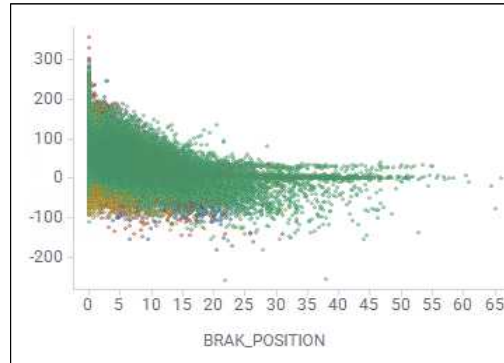


그림 3. 전류 분포와 브레이크 간 상관관계  
Fig. 3. Correlation between current distribution and break

그림 3은 방전 및 충전 전류의 분포와 브레이크 간 상관관계를 나타낸 그림이다. 운전 중 브레이크 방전 전류와 충전 전류가 급하게 교차되며, 브레이크를 약하게 밟는 경우 충전 전류가 더 발생한 것을 확인하였다.

## 2.4 예측을 위한 머신러닝 알고리즘 선택

위와 같은 EDA 분석을 통해 추출한 배터리 상태 주요 영향인자를 바탕으로 학습 데이터를 생성하고, 이를 활용하여 배터리 잔존가치 예측을 위한 머신러닝 알고리즘을 분석하였다. 일반적인 알고리즘에는 벡터 머신, 의사결정 트리, 랜덤 포레스트, 신경망, RNN이 있다. 본 논문에서는 Tree 구조의 ML알고리즘 중 RF(Random Forest) 사용하기 위하여 RF(Random Forest), XGB (eXtreme Gradient Boost), LGBM(Light GBM) 3개의 알고리즘을 적용하여 학습을 시켜 수행한 결과를 비교하였다. Random Forest는 여러 모델을 독립적으로 학습 시키고, 이들의 예측을 결합하여 최종 예측을 하는 앙상블 학습 기법이다. 무작위화는 이 학습 기법의 핵심이다. XGBoost, LightGBM는 앙상블 학습 제품군에 속하며 특히 대규모 데이터 세트 및 분산 컴퓨팅 환경에 적합한 기계 학습 알고리즘이다.

그림 4는 Tree 구조의 ML 알고리즘인 RF(Random Forest), XGB(XG Boost), LGBM(Light

GBM) 3개의 알고리즘을 적용하여 학습을 수행한 결과 값을 비교한 그래프이다. 학습된 알고리즘을 통해 출력되는 예측 값 및 실제 값을 MSE(Mean Squared Errors)알고리즘을 적용하여 비교 및 분석한 결과, 최종 예측 모델을 위한 알고리즘으로 가장 작은 데이터 간 차를 갖는 RF를 선정하였다.

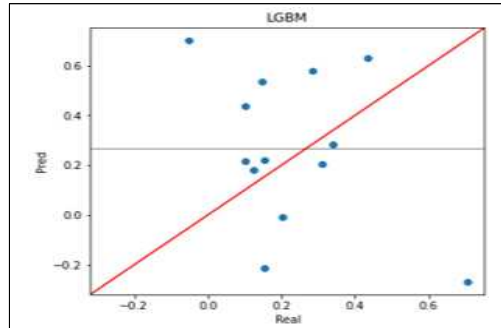
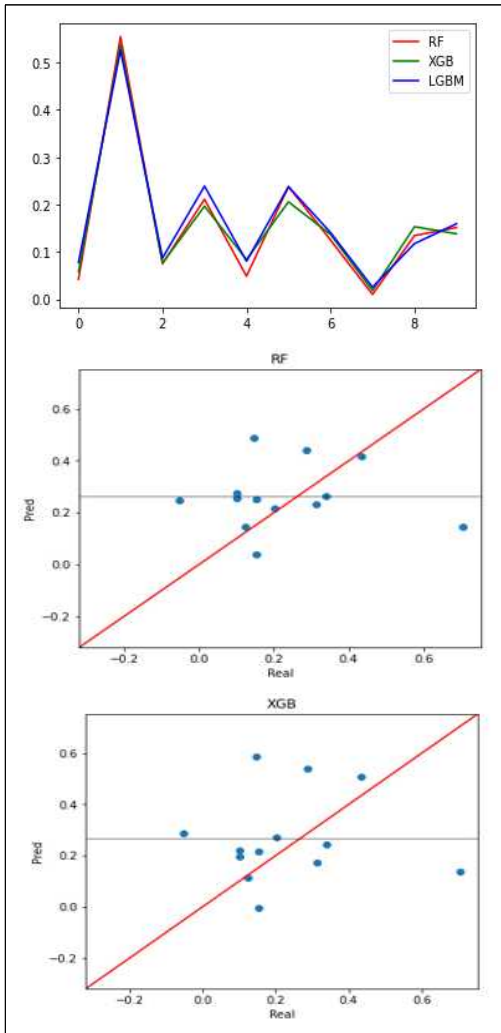


그림 4. RF, XGB, LGBM의 MSE 값 비교 그래프  
 Fig. 4. MSE value comparison graph of RF, XGB, and LGBM

표 4. RF,XGB 알고리즘 MSE 값 테이블  
 Table 4. RF,XGB algorithm MSE value table

Random Forest		XG Boost	
MSE		MSE	
X	Y	X	Y
0.443494	0.333995	0.243092	0.56875
0.418198	0.440087	0.38813	0.5
0.482846	0.502991	0.38824	0.475
0.590198	0.516162	0.436376	0.5
0.616084	0.547574	0.484375	0.55625
0.698222	0.685587	0.437308	0.2875
0.444352	0.534	0.590077	0.45625
0.397551	0.552456	0.607319	0.525
0.251187	0.526543	0.430482	0.84375
0.569597	0.716033	0.566639	0.8
0.436734	0.758953	0.95869	0.4125

표 4는 RF(Random Forest), XGB(XG Boost)의 MSE 값의 X,Y 값들의 테이블이다. 모델이 추출된 특징을 기반으로 영향을 미치는 인자를 예측하는 회기 작업이 포함된다. 따라서 RF(Random Forest) MSE 값이 다른 알고리즘 보다 직선상에 가깝게 분포되어 있어 오차 값이 가장 적은 분포를 나타내고 있다는 것을 확인하였다.

표 5. RF 알고리즘을 학습한 인자의 예측 영향도 값  
Table 5. Predicted influence value of the factor learned by the RF algorithm

feature_name	feature_importance
CELL_DIFF	0.257969
SOH_DCH_LAG_1	0.206015
EVCU	0.159623
CELL_MNTP	0.107164
COUNT	0.058995
BRK_POS	0.053218
SOC	0.041701
C_RATE	0.022198
ACC	0.016295
SYSTEM_POWER_P	0.013489

표 5는 RF(Random Forest) 알고리즘으로 학습한 인자의 예측 영향도 값을 나타낸 표이다. EDA 분석을 통해 도출된 학습 데이터 세팅 및 선정된 RF 알고리즘을 활용하여 모델 학습을 시킨 후, 모델의 결정 계수를 확인하였다. 모델의 결정 계수 값을 통해 배터리 성능 평가에 영향을 미치는 인자를 추출하였다.

2.5 모델 학습 후 주요 영향인자 중요도 분석

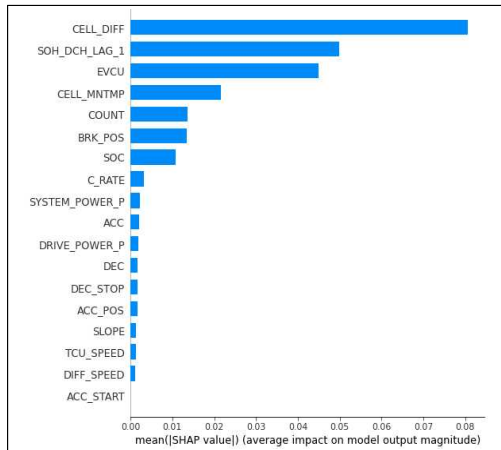


그림 5. XAI 값을 통해 확인한 영향 인자 중요도  
Fig. 5. Influence Weight determined through XAI values

그림 5는 XAI 알고리즘을 통해 확인한 영향 인자 중요도를 나타낸 그래프를 나타내었다. 제일 중요한 인자는 일평균 초 단위 셀 전압차이며, 두 번째로는 방전 시점 계산 SOH 차이이다. 셋 번째 주요 인자는 Cell Min 평균온도이며, 급정지 횟수, 배터리 충전 상태 (SOH), 일평균 방전 C-RATE, 시스템 전력, 급가속 회수, 누적 구동 계 소요전력으로 나타났다. 따라서 주요 인자별로 미치는 절댓값 및 음양의 부호 값을 추출하여, 추출한 인자별 중요 영향도를 바탕으로 추후 모델 재학습 및 고도화에 반영하여 모델의 예측을 강화할 수 있다.

3. 결론

본 논문에서는 EDA 분석을 기반으로 추출된 배터리 상태 주요 영향 인자를 기반으로 3가지 ML 모델 중 RF(Random Forest)이 가장 뛰어난 성능을 보이는 것을 확인하였다. 또한 이를 XAI 알고리즘을 활용하여 인출한 인자의 영향 중요도를 확인하는 방법을 제시 하였다. 타 모델과 차별 점으로, 실제 운영 중 전기 버스의 배터리 데이터를 실시간으로 수집, 운행 간 발생 할 수 있는 배터리 영향 인자 데이터를 수집하여 모델 설계하였으며 배터리 분석을 위한 데이터 중 분석 환경에 맞추어 배치 파일 통해 전처리 실시하였다. 배터리의 상태, 운행 및 충전 패턴 데이터 등을 종합적으로 고려하여 배터리 상태 영향 인자로 운행 패턴에서 급가속, 급 감속, 급정지와 충 방전 패턴에서일 주행횟수, 일일 누적 DOD와 셀 방전에서 셀 전압 차, 셀 최대온도, 셀 최소온도의 요소가 배터리 상태를 많은 영향을 미치는 인자로 확인되었다. Random Forest AI 알고리즘을 기반 데이터 수집/전처리를 통해 분석 예측과정의 모델 개발을 하였다. 이는 배터리 상태 예측 및 잔존가치 평가 분석에 적합함을 보였다. 따라서 추후 전기적 이동 수단 배터리의 기본 데이터 및 운행 중 데이터를 종합적으로 활용하여 배터리 상태를 실시간으로 분석하여 이를 통해 전기 차 배터리의 가장 큰 이슈인 화재 예방할 뿐만 아니라 배터리 재사용 및 재활용에도 유용하게 사용될 수 있을 거라고 판단되어진다.

## REFERENCES

- [1] S. Kim, "A review on electrochemical model for predicting the performance of lithium secondary battery," J. KES, vol. 22, no. 1, pp. 43-52, 2019.
- [2] J. Park, "Due to difference in uniformity of electrical characteristics between cells in a battery pack SOC estimation performance comparative analysis," TKPE, vol. 24, no. 1, pp. 16-24, 2019.
- [3] Zhang, Lide, et al. "Accurate online power estimation and automatic battery behavior based power model generation for smart phones." Hardware/Software design and System Synthesis(CODES+ISSE), 2010 IEEE/ACM/IFIP International Conference on. IEEE, 2010
- [4] S.J.Park, G.S.Song, and S.M. Park, "A Study on the parameters Estimation for SOC and SOH of the battery," J. Korean Soc. of Industry Convergence, vol. 23, no. 5, pp.853-863, Oct. 2020
- [5] S. H. Kang, D. H. Kim, J. H. Bae, T. W. Noh, and B. K. Lee, "Machine Learning based SOH Estimation Algorithm Using a Linear Regression Analysis," J. Power Electronics Conference, pp. 88-90, Nov. 2020
- [6] K. M. Lee, "Artificial intelligence : From Turing test to deep learning," Saeng Reung Publisher, pp. 847, Apr. 2019.
- [7] S. W. Kim, P. Y. Lee, D. H. Han, and J. H. Kim, "Multiple linear regression model based voltage imbalance estimation for high-power series battery pack," J. of IKEEE, vol. 23, no. 1, pp. 1-8, Mar. 2019
- [8] Alex Sherstinsky, "Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network," Physica D: Nonlinear Phenomena, Volume 404, 132306, ISSN 0167-2789, 2020.

## 저자약력

홍 승 모 (Seungmo Hong)

[정회원]



- 2018~ 현재 숭실대학교 전자정보공학부 IT융합학과 교수

〈관심분야〉 마이크로파, 안테나, 디지털 신호처리, 인공지능