

빅데이터와 머신러닝 기반의 인버터 고장 분류

김민섭*, Tanvir Alam Shifat*, 허장욱*,#

*금오공과대학교 기계공학과 항공기계전자융합공학전공

Classification of Inverter Failure by Using Big Data and Machine Learning

Min-Seop Kim*, Tanvir Alam Shifat* and Jang-Wook Hur*,#

*Department of Mechanical Engineering, Department of Aeronautics, Mechanical and Electronic Convergence Engineering, Kumoh National Institute of Technology

(Received 20 July 2020; received in revised form 21 September 2020; accepted 14 December 2020)

ABSTRACT

With the advent of industry 4.0, big data and machine learning techniques are being widely adopted in the maintenance domain. Inverters are widely used in many engineering applications. However, overloading and complex operation conditions may lead to various failures in inverters. In this study, failure mode effect analysis was performed on inverters and voltages collected to investigate the over-voltage effect on capacitors. Several features were extracted from the collected sensor data, which indicated the health state of the inverter. Based on this correlation, the best features were selected for classification. Moreover, random forest classifiers were used to classify the healthy and faulty states of inverters. Different performance metrics were computed, and the classifiers' performance was evaluated in terms of various health features.

Key Words : Failure Classification(고장분류), Machine Learning(머신러닝), Inverter(인버터), Big Data(빅데이터)

1. 서 론

4차 산업혁명과 관련된 핵심 개념 중 머신러닝은 기존 공학적 지식에 인공지능과 사물인터넷 등을 융합한 내용으로 활발한 기술 개발이 이루어지고 있다^[1]. 머신러닝을 성공적으로 구현하기 위해서는 4차 산업혁명의 또 다른 핵심 개념인 빅데이터가 필요한데, 빅데이터는 디지털 환경에서 생성

되는 데이터로 그 규모가 방대하고, 생성 주기도 짧으며, 형태도 수치 데이터뿐 아니라 문자와 영상 데이터를 포함하는 대규모의 데이터를 의미하고 있다.

일상 속에서 가장 흔하게 사용하는 에너지는 전기 에너지로, 우리나라는 현재 30% 이상의 전기가 화력발전소에서 공급되고 있다. 화력발전은 환경오염을 야기하는 주요 원인을 제공하고 있기 때문에 신재생에너지를 통한 새로운 발전 방법을 이용하려는 연구가 많이 진행되고 있으며, 그중 가장 중심이 되는 에너지는 태양광 발전이라고 할

Corresponding Author : hhjw88@kumoh.ac.kr

Tel: +82-54-478-7399

수 있으나, 태양광 발전의 문제점 중 하나는 발전의 핵심 부품인 인버터의 빈번한 고장을 들 수 있다^[2].

또한, 현재 자동차의 주 연료로 사용하고 있는 석유를 대체하기 위한 것이 전기 자동차이며, 효율을 위해서는 배터리에서 나오는 직류 전원을 교류 전원으로 변환이 필요하고, 이때 사용하는 인버터는 전기 자동차에 있어 핵심 부품이라고 할 수 있다^[3].

이러한 사례들을 통해 전압을 이용한 시스템에서 핵심 부품으로 구분되는 인버터는 여러 가지 원인에 의하여 고장이 나기 때문에, 인버터의 고장분류를 통한 유지보수 및 관리의 중요성도 함께 증가하고 있다. 이에 따라 본 연구에서는 인버터의 고장유형 및 영향 분석(FMEA, Failure Mode and Effect Analysis)을 수행하여 고장형태를 파악한 후, 센서들로부터 실시간 데이터를 확보 및 추출하고, RF(Random Forest) 알고리즘을 이용하여 고장을 분류하였다.

2. 고장모드 분석 및 시험

인버터에 대한 FMEA 결과는 Table 1과 같으며, 고장원인은 과부하, 과전압, 과전류 및 과열 등 4가지로 구분할 수 있다. 표에서 알 수 있는 바와 같이, 인버터에 가장 큰 피해를 주는 고장원인은 RPN이 640인 과전압이기 때문에, 인버터의 입력 전압을 변경하면서 측정된 데이터를 이용하여 인버터의 고장 분류를 수행하였다^[4].

Table 1 FMEA results of inverter

Part name	Function	Failure mode	Cause of failure	Severity	Occurrence	Detection	RPN
MOSFET	Switching control force generation	Short	Over load	10	4	3	120
Capacitor	Noise rejection	Faulty accumulator	Over voltage	10	8	8	640
Diode	Rectification action	Short circuit	Over current	10	8	4	320
Coil	Current change inhibition	Short	Over heating	10	5	6	300

인버터로부터 데이터를 수집하기 위해 Fig. 1에 나타낸 바와 같이 Power Supply, 직류 전압 센서, 인버터, 교류 전압 센서와 Arduino 등으로 시험장치를 구성하였으며, 센서에서 측정된 데이터를 Excel과 PLX-DAQ를 이용하여 저장하고, 데이터를 도식화하여 분석하였다.

인버터의 사양은 Table 2와 같으며, 인버터에 입력되는 전압을 정격전압인 12V보다 높은 15V로 설정하였을 때 약 28일 이후 작동이 중지하였으며, Fig. 2와 같이 고장이 발생하였다. 고장을 나타낸 Fig. 2(b)는 Capacitor 파손으로 인해 Capacitor 내부에 있던 물질이 노출 및 분산되어 있는 모습을 보인다.

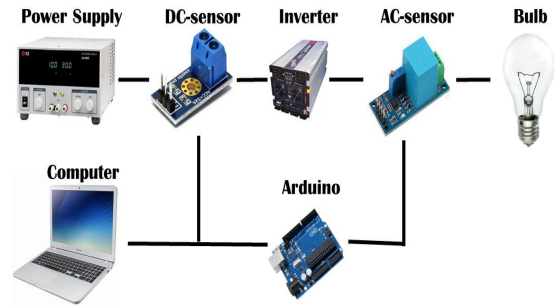


Fig. 1 Configuration of inverter measurement equipment

Table 2 Specifications of inverter

Model	Rated voltage	Output voltage	Maximum power
PI-100	12V	220v	150W

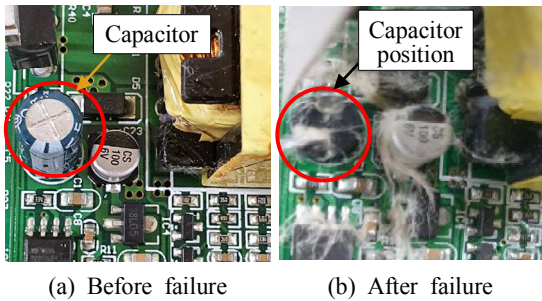


Fig. 2 Photos of before and after capacitor failure in inverter caused by high voltages

인버터에 220V를 사용하는 LED 전구를 연결하여 정격전압(12V)과 정격전압보다 높은 전압(14V, 15V)을 부가하였을 때 출력전압의 변화를 Fig. 3과 같이 측정하였다. 입력전압이 12V[Fig. 3(a)]와 13V[Fig. 3(b)]에서는 출력전압이 205~235V로 정상적인 작동 패턴을 보였으나, 입력전압이 14V[Fig. 3(c)]에서는 높은 입력전압으로 인해 출력전압이 13일째부터 정상작동 범위(205~235V)에서 벗어난 불안정한 상태를 보였으며, 입력전압이 15V[Fig. 3(d)]에서도 출력전압이 10일째부터 정상작동 범위를 벗어난 불안정한 상태가 되어 고장상태로 판단하였다. 이러한 결과를 바탕으로 입력전압이 12V와 13V일 때 데이터를 정상데이터, 14V와 15V일 때 데이터를 고장데이터로 분류하였다.

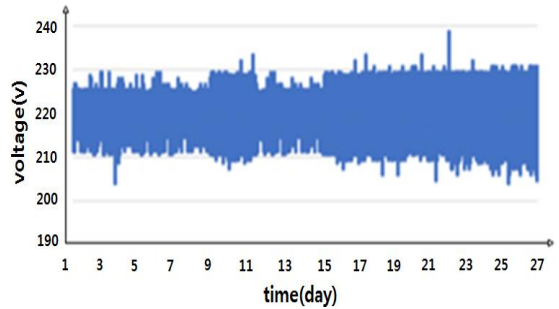
3. 머신러닝에 의한 인버터 고장분류

3.1 데이터 차원 축소

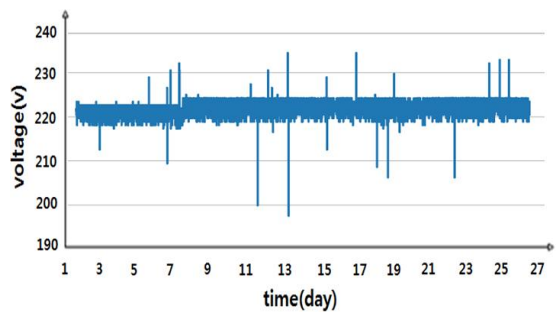
인버터의 입력전압에 대한 출력전압의 변화를 이용하여 고장을 분류하기 위해 입력전압 12V~15V에 대하여 10초당 1개의 전압 데이터를 샘플링하였고, 시험은 4주 동안 실행하여 1개의 시험당 약 350,000개의 데이터를 수집하였다.

먼저, 수집된 데이터들의 특징 추출을 수행하기에 앞서 각 특징은 최대 1,000개의 데이터 포인트에 걸쳐서 계산되기 때문에 `numpy.reshape()` 함수를 사용하여 1,000개당 1개의 데이터, 총 350개의 데이터(1000, 350)의 형태로 변형시켰다.

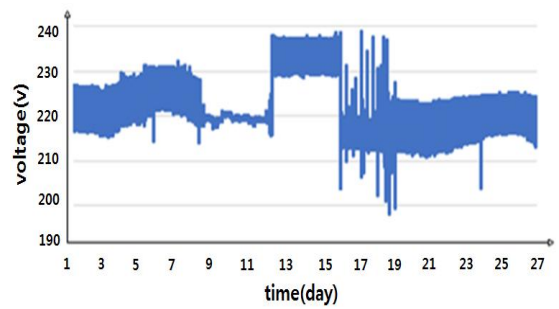
그 후 각 행 데이터를 `numpy` 패키지의 공식을



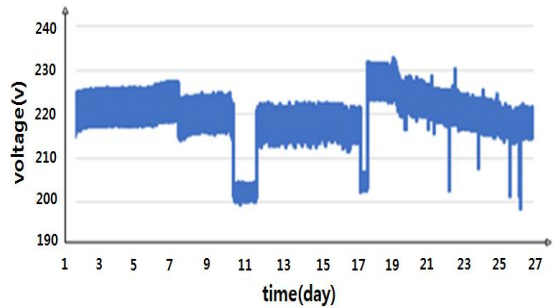
(a) 12V input



(b) 13V input



(c) 14V input



(d) 15V input

Fig. 3 Output voltage of the inverter

이용하여 AVG(Average), RMS(Root Mean Square), VAR(Variation), STD(Standard Deviation), MAX(Maximum) 및 MIN(Minimum)의 6가지 특징으로 계산하여 데이터를 형성하였고, 고차원 성분들을 저차원 데이터로 차원 축소를 위해 PCA(Principal Component Analysis), TSNE (t-Stochastic Neighbor Embedding), LLE(Locally Linear Embedding) 및 LDA(Linear Discriminant Analysis)의 4가지 기법을 사용하였다. 여기서, PCA는 고차원 데이터의 분산이 최대가 되는 축을 찾아 선형 연관성이 적은 저차원으로 변환하며, TSNE는 고차원 공간에서 데이터의 벡터가 2차원 공간에서도 유사하도록 데이터간의 특성을 보존하면서 차원을 축소하고, LLE는 고차원 공간에서 인접해 있는 데이터들 사이의 선형적 구조를 보존하면서 저차원으로 변환한다. 또한, LDA는 PCA와 유사하나, 고차원 데이터를 2차원 공간에서 데이터 범주를 명확하게 구분할 수 있는 특정한 축을 찾아 저차원으로 축소한다^[5-6].

이러한 4가지 기법을 이용하여 Fig. 4와 같이 차원을 축소하였다. 그림으로부터 PCA와 LDA는 12V의 데이터 경계를 분명하게 찾아볼 수 있지만, 13V, 14V, 15V의 데이터 경계는 식별할 수 없기 때문에 좋은 결과를 가져왔다고 볼 수 없으며, LLE는 모든 전압 데이터에서 경계를 찾아볼 수 없는 결과가 도출되었다. 그러나 TSNE는 전압 데이터의 경계가 분명하게 나타나는 결과를 가져왔으므로 인버터 고장분류를 위한 데이터 차원축소는 TSNE 기법이 최적으로 판단된다.

인버터의 고장분류는 Fig. 5와 같이 정상상태로 판단한 12V, 13V와 비정상상태로 판단한 14V, 15V 입력전압을 대상으로 아두이노 UNO를 통하여 출력단의 전압 데이터를 수집하였다. 수집된 센서데이터의 분류를 위해 TSNE 기법을 사용하여 데이터 차원을 축소하였고, RF(Random Forest) 알고리즘을 적용하여 데이터 분류를 수행하였으며, 최종적으로 훈련된 모델의 정확도 측정과 성능평가를 진행하였다.

3.2 RF 알고리즘

TSNE 차원으로 축소된 데이터를 분류하기 위해

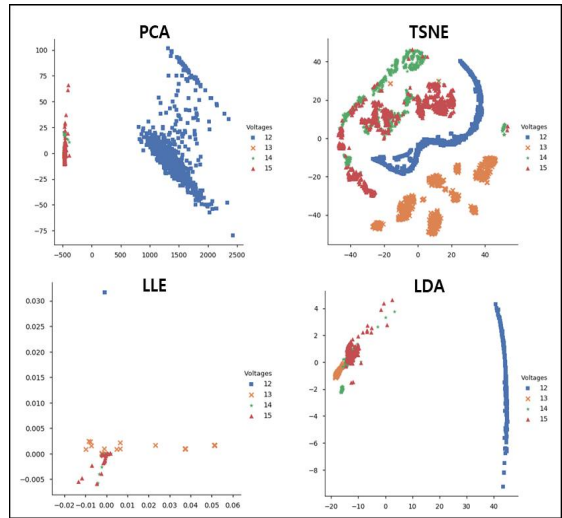


Fig. 4 Technique performed for dimension reduction

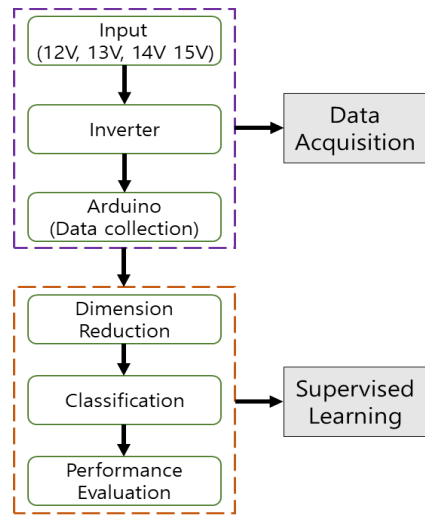


Fig. 5 Flowchart of proposed fault classification

대규모 데이터 분류에 적합하고 정확도가 높은 RF(Random Forest) 알고리즘을 이용하여 분류 모델을 학습하고 시험을 진행하였다. RF 알고리즘은 기준을 설정하여 구분된 결과에 대해 다수결의 원칙을 적용하는 방법으로, Fig. 6과 같은 결정 Tree를 기반으로 만들어졌다. 결정 Tree는 결정을 내리기 위해 사용하는 Tree로서, 하나의 Tree는 계층

구조로 이루어진 Node들과 Edge들의 집합으로 이루어져 있다.

만약, 하나의 결정 Tree에 많은 Feature가 선택된다면 오버피팅의 발생을 야기하는데, 이를 방지하기 위해 나온 것이 RF이다. 여러 가지 Tree들에 적절한 수의 Feature를 랜덤으로 분배하여 Tree들이 내린 예측값 중 가장 많이 나온 값을 최종 예측값으로 정하는 알고리즘이다.

RF에서는 다양한 입력 파라미터가 있는데, 결정 Tree의 최대 깊이(Max_depth)와 형성할 결정 Tree의 개수(N_estimator)를 설정해 주어야 한다. Max_depth는 데이터를 예측하는 결정 Tree의 최대 깊이로서 완벽하게 Class 값이 결정될 때까지 Tree의 가지를 분할한다. N_estimator는 동일한 데이터를 이용하여 반복 구분할 횟수를 의미하며, N_estimator=3이라면 Tree 3가지 형성되고, 각각의 Tree 결과를 조합하여 다수결의 원칙에 따라 데이터의 상태가 결정된다. 따라서 결정 Tree의 깊이와 개수가 증가할수록 정확도는 향상된다^[7-8].

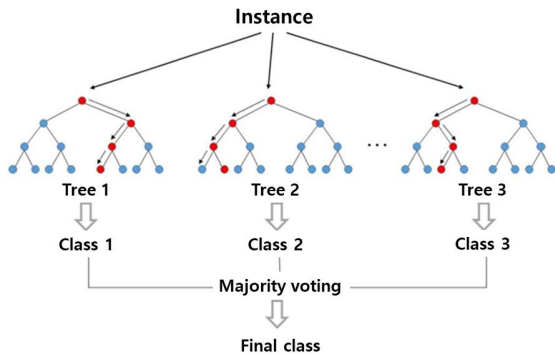


Fig. 6 Random forest classification process

		True Label	
		True	False
Predicted Label	True	True Positive	False Positive
	False	False Negative	True Negative

Fig. 7 Example of performance evaluation indicator

그리고 분류 모델에 대한 Accuracy는 식 (1)을 사용하였으며, Fig. 7에는 식 (1)에 대하여 True와 False (2×2 Matrix)로 표시하였다.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FN} + \text{TN} + \text{FP}} \quad (1)$$

여기서, TP는 True Positive로 실제 True인 정답을 분류 모델이 True라고 예측한 것이고, TN은 True Negative로 실제 False인 정답을 False라고 예측한 것이며, FP는 False Positive로 실제 False인 정답을 True라고 예측한 것이다. 그리고 FN은 False Negative로 실제 True인 정답을 분류 모델이 False라고 예측한 것을 의미한다^[7].

4. 고장분류 결과 및 고찰

RF 알고리즘을 이용한 머신러닝에서는 Max_depth=200, N_estimator=1,000으로 수행하였고 입력된 파라미터는 RF classifier에 다른 결과와 비교하여 가장 최상의 결과를 제공하였다.

RF 분류 모델은 전압마다 1,000개의 데이터 중 700개의 데이터를 사용하여 학습하였고, RF 분류 모델 학습의 결과를 Fig. 8에 Confusion matrix로 나타내었다. 입력전압이 12V, 13V, 14V, 15V에 대해, 각각 700개의 데이터 중 700개(100%)를 분류하도록 학습되었다.

또한, 학습된 RF 분류 모델은 전압마다 1,000개의 데이터 중 300개의 데이터를 사용하여 시험에 활용하였으며, 결과에 따른 Confusion matrix를 Fig. 9에 나타내었다. 12V와 13V의 경우, 각각 300개의 데이터 중 300개(100%)를 12V와 13V로 분류하였다. 그러나 14V는 300개의 데이터 중 266개(89%)의 데이터를 14V로 분류하였으며, 15V는 300개의 데이터 중 263개(88%)의 데이터를 15V로 분류하였다. 14V와 15V에 있어서 분류 성능의 일부 오차 발생원인은 Fig. 4의 TSNE 결과로부터 14V와 15V 데이터 중에 일부가 중복된 것을 확인할 수 있었던 바와 같이 14V와 15V의 고장특성이 유사한 상태로 RF 모델의 학습이 수행되었기 때문으로 사료된다.

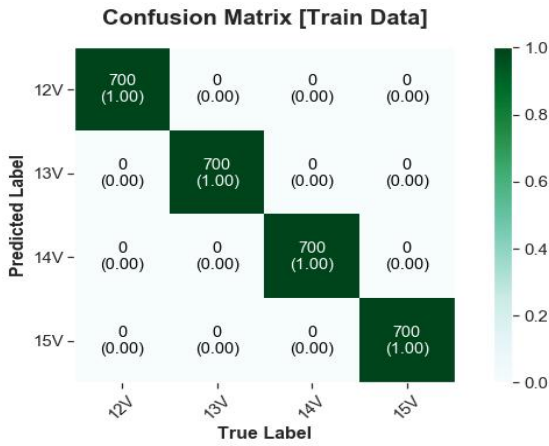


Fig. 8 Confusion matrix of RF(Train Data)

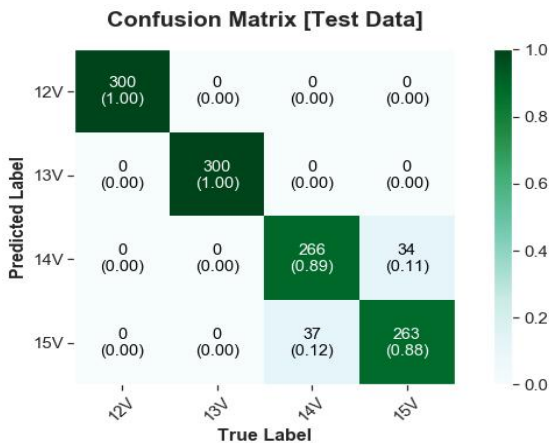


Fig. 9 Confusion matrix of RF(Test Data)

Table 3 Classification results of RF

Voltage	Accuracy	Precision	Recall	F1-score
12V	0.94	1.00	1.00	1.00
13V		1.00	1.00	1.00
14V		0.88	0.89	0.88
15V		0.89	0.88	0.88

RF 알고리즘을 이용한 모델의 분류성능 평가를 수행한 결과를 Table 3에 나타내었다. Precision은

모델이 True라고 분류한 것 중에서 실제 True인 것의 비율인데, 예를 들어 모델이 12V라고 예측했는데 예측한 데이터가 실제로 12V인지를 알아볼 수 있는 지표이며, 식 (2)와 같이 표현할 수 있다. Recall은 실제 True인 것 중에서 모델이 True라고 예측한 것의 비율이고, 식 (3)과 같이 나타낼 수 있으며, F1-score는 Precision과 Recall의 조화평균으로 식 (4)와 같이 나타낼 수 있다^[9].

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1-score = 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} \quad (4)$$

$$= 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

RF 모델의 분류성능 평가에 따르면, 12V~15V 전압에 대해 Accuracy는 94%가 산출되었으며, 12V와 13V에서는 Precision, Recall 및 F1-score에서 모두 100%의 결과를 나타낸다. 그러나 14V에서는 Precision 88%, Recall 89% 및 F1-score 88%를 표시하고 있으며, 15V에서는 Precision 89%, Recall 88% 및 F1 score 88%를 확인할 수 있다.

5. 결론

엔지니어링 어플리케이션에서 널리 사용되는 인버터를 대상으로 빅 데이터와 머신러닝을 수행하여 고장을 분류하였다. 얻어진 결과를 요약하면 다음과 같다.

- 1) 인버터의 입력전압(12V~15V)에 따른 출력전압 데이터의 차원축소는 TSNE가 PCA, LEE 및 LDA 등의 기법보다 우수하다.
- 2) RF 알고리즘을 이용한 분류 성능은 12V~15V의 인버터 입력전압에 대해 Accuracy가 94%로 산출되었다.

- 3) 인버터의 입력전압이 12V와 13V에서는 Precision, Recall 및 F1-score에서 모두 100%의 결과를 표시하였다.
- 4) 인버터의 입력전압이 14V에서는 Precision 88%, Recall 89% 및 F1-score 88%를 표시하였으며, 15V에서는 Precision 89%, Recall 88%, F1-score 88%를 나타내었다.

후 기

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 Grand ICT연구센터지원사업의 연구결과로 수행되었음(IITP-2020-2020-0- 01612).

REFERENCE

1. Choi, H., Kim, T. K., Heo, G. R., Hur, J. W. and Choi, S. D., "A Study on the Prediction of Failure of Fuel Pump Based on Big Data and Machine Learning", Journal of the Korean Society of Manufacturing Process Engineers, Vol. 2019, No. 4, pp. 231-231, 2019.
2. Yoon, Y. H., "Smart Monitoring System for Improving Solar Power System Efficiency", The Journal of the Institute of Internet, Broadcasting and Communication, Vol. 19, No. 1, pp. 219-224, 2019.
3. Lee, S. T., Cho, J. H. and Kim, D. K., "Development Trends in Electric Vehicle Driving Systems", Journal of the Korean Institute of Power Electronics, Vol. 16, No. 2, pp. 32-37, 2013.
4. Min, J. K., Ra, B. H. and Huang, U. S., "Failure Mode Design of PV Inverter According to Grid Conditions", Journal of the The Korean Institute of Power Electronics, Vol. 2016, No. 7, pp. 24-25, 2016.
5. Sun, Z., Xing, W. and Guo, W., "A Survey on Dimension Reduction Algorithms in Big Data Visualization", Cloud Computing, Smart Grid and Innovative Frontiers in Telecommunications, pp. 375-395, 2019.
6. Song, W., Wang, L., Liu, P., "Improved T-SNE Based Manifold Dimensional Reduction for Remote Sensing Data Processing", Multimedia Tools and Applications, Vol. 78, No. 4, pp. 4311-4326, 2019
7. Shifat, T. A. and Hur, J. W., "An Improved Stator Winding Short-circuit Fault Diagnosis using AdaBoost Algorithm", In 2020 International Conference on Artificial Intelligence in Information and Communication(ICAIC), pp. 382-387, 2020.
8. Shin, K. S and Baek, J. G., "Machine System Failure Prediction using Multivariate State Estimation Technique(MSET) and Real Time Contrast(RTC)", Journal of the Korean Institute of Industrial Engineers, Vol. 2017, No. 11, pp. 2203-2208, 2017.
9. Shifat T. A. and Hur, J. W., "EEMD assisted supervised learning for the fault diagnosis of BLDM motor using vibration signal", Renewable Energy, pp. 3981-3990, 2020.