

제조 분야를 위한 데이터 분석 모델

이기성¹, 이종찬^{2*}

¹호원대학교 컴퓨터게임학부, ²군산대학교 컴퓨터정보공학과

Data Analytics in Manufacturing

Gi-Sung Lee¹, Jong-Chan Lee^{2*}

¹Division of Computer & Game, Howon University

²Department of Computer Information Engineering, Kunsan National University

요약 하이테크 제조 분야에서는 인더스트리 4.0(Industry 4.0)과 스마트 팩토리(Smart Factory)를 통하여 끊임없이 변화하는 시장 요구사항에 대응함으로써 혁신적인 제품을 신속하게 제공하기 위하여 노력하고 있다. 정부의 스마트 팩토리 보급 사업을 통하여 중소중견기업들의 제조 데이터의 수집 및 활용도가 높아지고 있지만, 데이터의 수집과 활용에 여전히 괴리가 있는 것이 현장의 시각이다. 데이터의 수집은 물리적으로 한계가 없으나, 활용의 측면에서 수집한 데이터로부터 의미 있는 정보를 추출하여 제조 현장에 활용하기 위해서는 고수준의 분석작업이 필요하므로 아직은 데이터를 활용하고 있는 기업의 수가 적은 것이 현실이다. 본 연구에서는 주요 공정 데이터를 자동으로 수집하고, 이 데이터를 기반으로 설비 모니터링과 AI 기술을 활용한 설비의 최적 공정조건을 제시한다. 이를 통하여 기존 설비의 설정값과 더불어 주요 외부 환경을 고려하여 설비의 생산 최적화 설정 정보를 제공함으로써 휴먼 오류를 최소화하고 불량률을 개선할 수 있다.

Abstract The high-tech manufacturing sector is striving to deliver innovative products faster by responding to ever-changing market requirements through the search for new methods, such as Industry 4.0 and Smart Factory. Although the collection and utilization of manufacturing data by small and medium-sized companies are increasing through the government's smart factory supply project, the collection and utilization of data are still different from the perspective of the field. Data collection is not physically limited, but extracting meaningful data from collected data in terms of utilization and using it at the manufacturing site requires high-quality analysis. Thus, the number of companies actually using the data is small. In this study, key process data was automatically collected. The optimal process conditions for equipment using AI technology and equipment monitoring are suggested based on these data. Through this, the defect rate can be improved by minimizing human error by providing production optimization setting information of a facility considering the main external environment and the setting value of the existing facility.

Keywords : Big-data, AI, Manufacturing Data, Data Analysis, Process Conditions

본 논문은 2022년도 호원대학교 지원에 의하여 출판되었음.

*Corresponding Author : Jong-Chan Lee(Kunsan National Univ.)

email: chan2000@kunsan.ac.kr

Received August 26, 2022

Revised October 4, 2022

Accepted October 7, 2022

Published October 31, 2022

1. 서론

산업 설비의 자동화와 데이터의 광범위한 적용에 기반을 둔 4차 산업혁명 시대에 제조업 분야의 변화가 크게 주목을 받고 있다. 인더스트리 4.0을 표방하며 4차 산업혁명을 선도하고 있는 독일은 사물인터넷, 빅데이터, 인공지능, 클라우드 등 첨단기술을 제조 분야에 도입하고 스마트 제조(Smart manufacturing) 환경 구축에 기술을 집중하고 있다. 제조업 중심의 4차 산업혁명은 미국, 일본, 중국 등 주요 국가에서도 기술개발에 자본 역량을 집중하고 있다. 스마트 제조는 각종 센서를 통해 수집된 데이터를 분석하여 제조 공정상의 불량률을 최소화하거나, 설비 이상을 사전 감지해 고장을 예방하고, 자동화된 통합 제조 공정의 구축에 활용된다. 이를 구현하기 위한 핵심기반 기술이 바로 인공지능이다. 즉, 현실 공간에서 수집된 데이터를 가상 공간에서 분석하고, 도출된 솔루션 및 의사결정 사항을 다시 현실 공간에 전달해 자동으로 기기를 제어하고 공정을 최적화하는 전 과정을 인공지능이 수행한다[1-3].

현재 국내에서 4차산업 디지털 전환의 시도는 금융, 통신 등 일부 분야에서 빅데이터, AI 기술을 적용하여 활용되고 있으나, 제조 분야에 대한 디지털 전환 시도는 미흡한 상황이다. 스마트 제조의 지능형 어플리케이션의 경우 미국 대비 국내 기술 수준은 현재 76.8%로 격차는 2년으로 분석된다. 스마트 제조 분야의 디지털 전환과 관련 기술의 적용이 미흡한 원인은 작은 기업 규모, 전문인력 부족, 데이터 부재, 도입 효과 불확실 등이다. 특히 대다수의 중소기업은 요소 기술 부재, 자금 역량 부족으로 인하여 불확실한 디지털 전환 기술의 도입을 피하는 경향이 크다[4-6].

데이터는 대규모의 작업 과정의 자동화, 작업 시간의 단축 등 제조업에 큰 이점을 가져오지만, 저장 및 처리되는 데이터의 양이 기하급수적으로 증가하므로 제조 회사는 데이터의 새로운 활용법을 찾아야만 한다[7-9]. 데이터 분석은 다양한 위험 예측과 판단이 필요한 제조 공정에서 문제 상황을 사전에 예측하고 예방을 도와 기업의 큰 손실을 막아주고, 소비자의 피드백을 데이터 처리하여 더 나은 제품을 개발 및 생산에 효율성을 부여할 수 있으며, 양품 생산을 위한 생산 품종, 라인별 최적 공정조건을 도출할 수 있게 한다[10-12].

본 연구에서는 최적 공정조건 분석을 수행한다. 최적 공정조건 분석은 과거의 최적 물성 조건에 맞는 양품 생

산 시기를 기반으로 미래의 양품 생산을 위한 라인별 최적 공정조건을 도출하고 적용하는 것이다. 대부분의 생산라인에서는 제품을 생산하면서 공정조건변경은 현업의 경험과 감에 의존하므로, 해당 분석을 통해 데이터 기반의 생산 공정의 기술축적 및 기술 표준화로 생산성 및 품질 향상을 위한 초석을 마련할 수 있다.

2장에서는 데이터 분석을 위한 세부 방법을 기술하고 3장에서 분석 결과를 보인다. 4장에서 결론을 맺는다.

2. 제안 시스템 구조

2.1 시스템 구성

데이터의 수집 → 표준 최적 공정 관리조건 도출 → 현장 적용 → 모니터링 단계로 진행되며, 최적 조건을 도출하고, 실제 현장의 공정 운영 조건에 설정하여 운영 최적화까지 수행한다. 빅데이터의 분석을 위해서는 대량의 데이터를 생산하는 구조가 필요한데, 제조 환경에서는 다수의 센서로 이루어진 사물인터넷으로부터 방대한 데이터를 생성한다. 센서로부터 생성되는 실시간 데이터는 지속적인 생성, 비정형성 데이터, 방대한 데이터의 양 등의 특징을 갖는다.

이런 사물인터넷 환경이 목적에 맞게 동작하기 위해서는 빅데이터의 가공과 분석이 필요하다. 즉 수집된 데이터 자체로는 가치를 확보할 수 없고, 목적에 맞는 데이터 분석과 처리가 필요하다. 따라서 아래와 같은 데이터의 수집 및 전처리, 그리고 정밀한 분석 과정을 통하여 데이터의 정확성 및 신뢰성을 확보해야 한다. Fig. 1에 AI 기반 데이터 분석 과정을 보인다.

① 데이터 신뢰도 향상

- 설비의 주요 공정 데이터와 공장 환경 정보의 분석 및 처리를 통하여 회귀분석 모델 개발을 위한 양질의 데이터를 제작한다.

② 데이터의 수집 및 전처리 시스템

- 설비 데이터와 외부 온·습도 데이터를 기반으로 빅데이터 수집 장치를 구축한다.

● 설비로부터 공정 데이터(온도, 압력, 습도, 모터 속도 등)를 수집한다.

● 공장 내 온·습도 및 원료건조 온도, 온수기 I/O 온도, 금형 표면 온도를 수집한다.

● 데이터 기반의 주요 공정 데이터 분류 SW를 구현

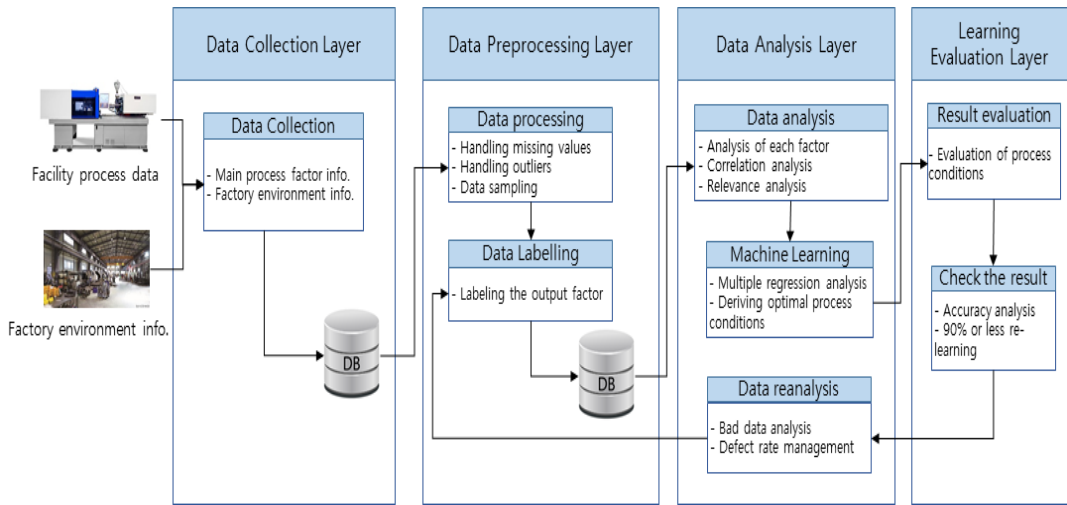


Fig. 1. AI-based data analysis procedure

한다.

- 수집된 데이터가 저장 가능한 데이터베이스 H/W 및 S/W 시스템을 구축한다.
- 수집한 데이터로부터 결측치, 이상치, 데이터 샘플링을 수행한다.
- 샘플링된 데이터의 라벨링을 수행한다.

③ 데이터 분석 시스템 구현

- 각 공정 데이터의 제약 사항을 확인하고, 여러 가지 조건이 최종 제품에 미치는 영향을 파악한다.
- 상관분석을 통한 인자 간의 관련 유·무를 파악한다.
- 다중 회귀분석과 딥러닝 기반 회귀분석의 예측을 모두 시도하여 우수한 모델을 선택한다.
- 불량분석을 통하여 외부 상황에서의 생산 상태, 불량률을 개선한다.

2.2 데이터의 수집 장치 및 AI 모델 인터페이스

Fig. 2와 같이 각 설비의 공정상의 인자, 공장 내의 온·습도 정보를 수집하고, 공정에 필요한 조건을 인공지능 S/W와 연계한다. 입력 데이터에 대한 딥러닝 회귀분석을 이용하여 학습하고 회귀계수(가중치)를 산출한 후, 사출 장비 및 환경 조건을 추론하여 공정 및 환경에 반영한다. 수집된 데이터의 분석을 토대로 예측 모델을 개발하는데, 일반적인 회귀모델과 딥러닝 기반의 회귀모델을 제안하고 더 나은 예측 성능을 보이는 모델을 선택한다.

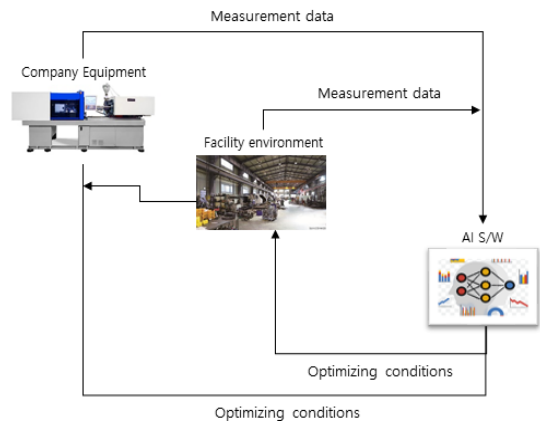


Fig. 2. AI Model Interface

이를 위하여 우선, 각 설비와 제어기를 연결하여 생산 수량을 자동으로 수집하고 설비의 성형 가동 값을 수집하여 원료조건 온도, 온수기 입출력 온도, 금형 표면 온도, 공장 내 온습도와 동기화가 필요하다. 데이터의 수집은 TCP기반의 MODBUS 프로토콜의 인터페이스를 통하여 각 설비의 현재 생산되는 설정값을 읽어서 DB에 저장하도록 구성한다. 머신러닝 학습에 필요한 입력 데이터를 입력 및 DB 저장, 생산설비의 데이터 및 입력 데이터 사상 작업을 통하여 학습 코어에 학습데이터를 전달하고 예측 데이터확인 가능한 UI(User Interface)를 구축한다.

2.3 데이터 처리

Fig. 3에 데이터 처리 과정을 보인다. 수집된 데이터를 센서 데이터, 공정 데이터, 품질 데이터로 분리한다. 사출 공정상의 데이터는 완전 무작위 결측(Missing Completely At Random, MCAR)이라고 가정하고 결측값(Missing Value)을 처리한다. 결측값 처리 방법으로 사전 조사 결과 결측치가 10% 미만이면 결측치 데이터를 삭제하고, 10% 이상이면 공정 인자 값의 평균이 아닌 각 인자의 특성을 고려하여 값을 삽입하는 회귀 삽입(Regression Imputation)을 수행한다.

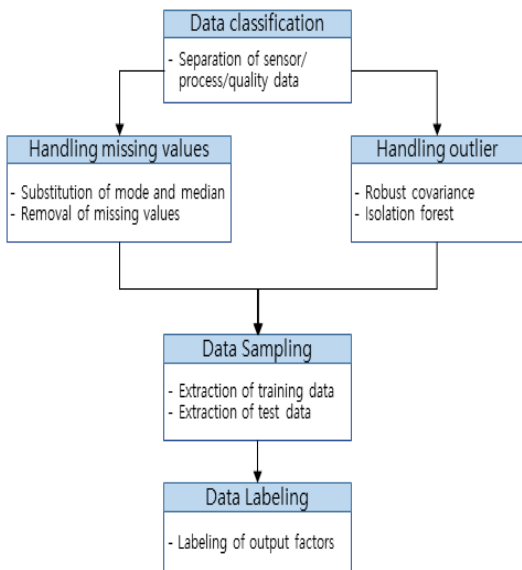


Fig. 3. Data preprocessing process

이상치(outlier)는 정상적인 데이터의 분포 범위 밖에 위치하는 값으로서 분산을 과도하게 증가시켜 데이터 분석 모델의 예측력을 악화시키는 주요 원인이다. 따라서 제조 공정 데이터 분석의 경우, 제품의 불량 원인을 찾아내기 위하여 특정 공정의 센서 값 변화를 분석하여 데이터 이상치를 추출한다. 이상치의 판별을 위하여, 제조 공정의 불량률에 연관된 센서의 평균으로부터 일정기준으로 상한선과 하한선을 설정하고, 이를 벗어났을 때와 제품 불량과의 관계를 분석한다. 탐지 방법으로서 두 방법을 적용하는데, 정규분포에서 97.5% 이상 또는 2.5%의 이하에 포함되는 값을 이상치로 판별하는 Robust covariance와 의사결정 트리 기반 이상탐지 기법으로 랜덤으로 데이터를 분리하여 이상치를 판별하는 Isolation forest를 적용한다. 처리 방법으로는 우선, 하한값과 상

한값을 결정한 후 하한값보다 적으면 하한값으로 대체, 상한값보다 크면 상한값으로 대체한다. 두 번째로 평균의 표준편차를 이용한다(하한값 = 평균 - 데이터 개수*표준편차, 상한값 = 평균 + 데이터 개수*표준편차). 97.5% 이상 혹은 이하값을 이상치로 제거하거나 대체한다.

위와 같이 정제된 데이터를 다중선형 회귀분석을 위하여 학습에 사용될 데이터와 테스트를 위하여 사용될 데이터로 구분하여 샘플링을 수행한다. 이 샘플링된 데이터에 라벨링 작업을 수행한다.

2.4 데이터 분석

Fig. 4에 데이터 분석 과정을 보인다. 우선, 각 인자 특성 분석을 수행한다. 우선 각 공정 인자(온도, 냉각시간, 성형시간 등)의 특성을 분석하고, 각 인자의 공정 내 역할과 사출 공정의 불량률 감소 조건을 분석한다.

각 독립변수 간의 상관분석(Correlation Analysis)을 수행한다. 각 인자의 상관관계가 클수록 다중 공선성(Multicollinearity) 문제가 발생한다. 각 인자 간의 상관계수(Correlation Coefficient, r)가 0에 근접하는 인자들을 입력 인자로 선정한다. 데이터 분석 시에 부정확한 종속변수 값을 도출하므로, 독립변수 간의 상관계수를 구하고, 분산팽창요인(VIF, Variance Inflation Factor)을 구하여 이 값이 10을 넘는다면 보통 다중 공선성의 문제가 있다고 판단한다. 다중 공선성 문제를 해결하기 위하여 상관관계가 높은 독립변수 중 하나 혹은 일부를 제거, 변수 변형, 새로운 관측치를 이용하거나, 주성분 분석(PCA, Principle Component Analysis)를 이용한 개각행렬(Diagonal Matrix)의 형태로 공선성을 제거한다.

입력 인자와 출력 인자 간의 상관관계를 분석한다. 사출 설정값에 영향을 미치는 인자들을 선정해야 한다. 출력 인자와의 상관계수(Correlation Coefficient)가 1이나 -1에 근접하는 입력 인자들을 선정하는데, 결정계수(Coefficient of Determination)는 r^2 로 계산한다. 단순 상관분석과 다중 상관분석(Multiple Correlation analysis)을 이용하여 분석을 수행하고, 입력 인자별로 출력 인자와의 상관계수를 구하여 선형성이 높은 입력 인자를 출력 인자를 구하기 위한 입력값으로 선정한다.

불량 분석(Defect Analysis)을 수행한다. 주요 인자의 변동/변화에 따른 품질 분석/예측을 수행하는 것으로서, 불량 공정 여부 확인 후 또는 최종 제품에서의 불량 여부 확인 후에 불량 라벨링된 데이터를 확보한다. 불량

데이터 분석을 통하여 각 공정의 데이터가 얼마나 불량에 기여하는지를 도출하고 각 인자의 표준편차 또는 이상치를 확인하고, 그 인자가 제품 불량에 주는 영향을 판단한다.

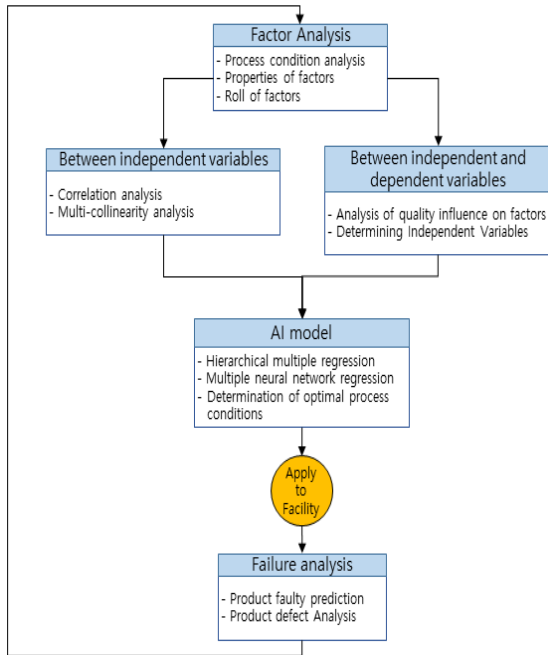


Fig. 4. Data Analysis Process

2.5 AI 모델

Fig. 5에 AI 분석 모델을 보인다. 다중신경망은 입력과 출력이 직접 연결이 불가하므로 비선형성을 포함하며, 구성된 층만큼 더 다양한 패턴을 학습(회귀)한다. 은닉층과 노드의 개수에 비례하여 2개 이상의 회귀분석이 가능하므로 다수개의 은닉계층으로 입력 데이터의 변환을 단계별로 추정하고 계층을 증가시켜 더 많은 패턴을 찾는다.

순전파(Feed-forward) 및 손실함수(Loss Function)로 n 개의 입력 데이터(설비의 온도, 압력, 속도, 공장 온도 등)가 공정조건에 미치는 영향을 파악하고, 최종적으로 공정조건 설정값이 출력한다. 입력 인자, 출력 인자를 고려하여 입·출력 계층을 확정한다. 어느 수준의 신경망을 구축해야 하는지 예측이 불가하므로 은닉계층 수를 결정하기 위하여 반복적인 테스트를 수행한다. 과적합(Overfitting)을 방지하고, 불량률이 허용 오차를 벗어나는 경우, 사용 데이터와 불량 데이터를 기준으로 모델 추가 학습을 수행한다.

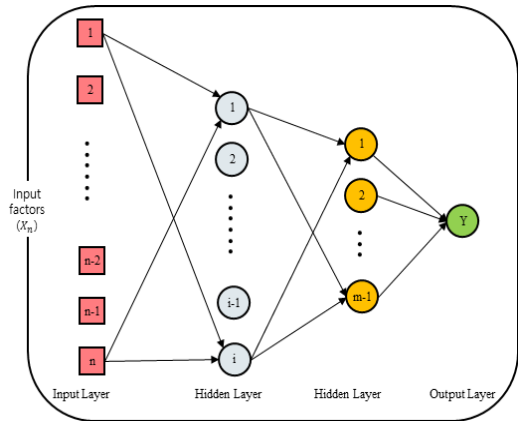


Fig. 5. Multiple neural network regression analysis

3. 분석 평가

Table 1에 제조 데이터 분석을 위한 주요 공정 데이터를 보인다. 주요 인자를 중심으로 기술하였으며, 다중 회귀 분석을 통하여 분석 결과를 제시한다.

Table 1. Process factors for analysis

Classification	Category	
Data type	Text, Number	
Data Properties	Approximately 1400 numerical data of injection molding equipment	
Data collection method	1. PLC, Temperature and Humidity sensor 2. Receive cycle: By Lot, Every 20 seconds	
Main parameters	Mold surface temperature	°C
	Input temperature to the water heater	°C
	Output temperature to the water heater	°C
	the drying temperature of the raw material	°C
	Input pressure to the injection device	kgf/cm ²
	Output pressure to the injection device	kgf/cm ²
	Temperature in the factory	°C
	Humidity in the factory	g/m ²

VIF를 이용하여 다중공선성을 검토한 결과, 모든 독립변수가 1.5 미만 분석되었다. 이를 통하여 다중공선성이 발생했을 가능성이 낮다고 판단했다. 또한 모형의 결정계수 R^2 은 0.851이고, 자유도 조정 후의 결정계수가

0.823로 산출되었다. 따라서 독립변수로 인하여 종속변수의 분산 중 85.1%가 설명된다고 할 수 있다. F 검정은 0.1% 수준으로 유의하며, 95% 신뢰구간으로 독립변수의 편회귀계수는 5% 수준에서 유의하고 모집단의 영향이 존재하는 것으로 분석되었다.

4. 결론

스마트 제조의 근간은 빅데이터 활용에 있다. 센서 기술, 데이터 산출 등을 기반으로 한 공정 디지털화의 중요성이 점점 대두되고 있다. 반면에 제조업은 세계적으로 생산 비용 절감이 거세게 요구되고 있으며 동시에 제품의 제조 원가와 공정 과정에 대한 정보는 갈수록 투명하게 공개되어 제품의 가격, 성능 경쟁력에 대한 기대치는 날이 갈수록 높아지고 있으므로, 이미 어느 정도 효율성을 갖추고 있는 생산 체계를 한 단계 더 효율화해야 하는 과제에 당면해 있다. 본 연구에서는 제조업체에서 수년 동안 수집한 많은 양의 데이터를 분석하고 활용할 수 있는 방안을 모색하여 작업에 실질적인 도움, 생산성 증대, 그리고 품질의 일관성을 유지하기 위해 AI 모델을 개발하고 적용하고자 하였다. 제조업체의 생산설비, 제품, 생산환경, 작업자로부터 얻어지는 빅데이터의 분석을 기반으로 설비의 최적 공정조건을 제시하는 방법을 제시하였다. 추후 연구로서 품질의 불량 원인을 사후 분석이 아닌 제조 공정 중 실시간 분석으로 분석하는 방법을 제시하고자 한다.

References

- [1] Tao, F., Qi, Q., Liu, A., Kusiak, A., "Data-driven Smart Manufacturing", The Journal of Manufacture System, vol. 48, pp. 157-169, 2018.
DOI: <https://doi.org/10.1016/j.jmsy.2018.01.006>
- [2] Baicun Wang, "Smart Manufacturing and Intelligent Manufacturing: A Comparative Review" Elsevier BV, vol. 7, no. 6, pp. 738-757, 2021.
- [3] Koren, Y., Gu, X., Guo, W., "Reconfigurable Manufacturing Systems: Principles, Design, and Future Trends", Frontiers of Mechanical Engineering, vol. 13, pp. 121-136, 2018.
DOI: <https://doi.org/10.1007/s11465-018-0483-0>
- [4] D. B. Ko, J. M. Park, "A Study on the Visualization of Facility Data Using Manufacturing Data Collection Standard", The Journal of The Institute of Internet, Broadcasting and Communication, vol. 18, no. 3, pp. 159-166, 2018.
DOI: <https://doi.org/10.7236/IIBC.2018.18.3.159>
- [5] S. Y. Kim, "A Case Study of the Introduction of Smart Factory Operation Management (FOM) in the Fourth Industrial Revolution Era", Journal of Korean Association of Computer And Accounting, vol. 16, no. 1, pp. 43-62, 2018.
DOI: <https://doi.org/10.32956/kaoca.2018.16.1.43>
- [6] K. S. Noh, S. Park, "An Exploratory Study on Application Plan of Big Data to Manufacturing Execution System", The Society of Digital Policy and Management, vol. 12, no. 1, pp. 305-311, 2014.
DOI: <https://doi.org/10.14400/JDPM.2014.12.1.305>
- [7] Liang, S., Rajora, M., Liu, X., Yue, C., Zou, P., Wang, L., "Intelligent Manufacturing Systems: A Review", International Journal of Mechanical Engineering and Robotics Research, vol. 7, no. 3, pp. 324-330, 2018.
DOI: <https://doi.org/10.18178/ijmerr.7.3.324-330>
- [8] Wang, B., Tao, F., Fang, X., Liu, C., Liu, Y., Freiheit, T., "Smart Manufacturing and Intelligent Manufacturing: A Comparative Review", Engineering, vol. 7, no. 6, pp. 738-757, 2020.
DOI: <https://doi.org/10.1016/j.eng.2020.07.017>
- [9] Mittal, S., Khan, M. A., Romero, D., Wuest, T., "A Critical Review of Smart Manufacturing & Industry 4.0 Maturity Models: Implications for Small and Medium-sized Enterprises (SMEs)", Journal of Manufacturing Systems, vol. 49, pp. 194-214, 2018.
DOI: <https://doi.org/10.1016/j.jmsy.2018.10.005>
- [10] Mourtzis, D., Vlachou, E., Milas, N., "Industrial Big Data as a Result of IoT Adoption in Manufacturing", Procedia CIRP, vol. 55, pp. 290-295, 2016.
DOI: <https://doi.org/10.1016/j.procir.2016.07.038>
- [11] H. Y. Cho, "On-line Process Data-driven Diagnostics Using Statistical Techniques", Journal of the Korea Academia-Industrial cooperation Society, vol. 19, no. 3 pp. 40-45, 2018.
DOI: <https://doi.org/10.5762/KAIS.2018.19.3.40>
- [12] D. Y. Lee, J. W. Park, S. H. Cho, J. S. Lee, "Maintainability Prediction of Guided Missile based on Machine Learning using Field Data", Journal of the Korea Academia-Industrial cooperation Society, vol. 23, no. 2 pp. 518-526, 2022.
DOI: <https://doi.org/10.5762/KAIS.2022.23.2.518>

이 기 성(Gi-Sung Lee)

[중신회원]



- 1996년 2월 : 송실대학교 컴퓨터 과학과 (공학석사)
- 2001년 2월 : 송실대학교 컴퓨터 과학과 (공학박사)
- 2001년 3월 ~ 현재 : 호원대학교 컴퓨터게임학과 교수

<관심분야>

모바일 멀티미디어, 네트워크 보안, 머신러닝

이 중 찬(Jong-Chan Lee)

[정회원]



- 1996년 8월 : 송실대학교 컴퓨터 과학과 (공학석사)
- 2000년 8월 : 송실대학교 컴퓨터 과학과 (공학박사)
- 2000년 10월 ~ 2005년 2월 : 한국전자통신연구원 선임연구원
- 2005년 3월 ~ 현재 : 군산대학교 컴퓨터정보공학과 교수

<관심분야>

머신러닝, 빅데이터, 블록체인