



자동차부품 사출공정의 결함 검출 및 원인분석을 위한 딥러닝 알고리즘 적용에 관한 연구

고민재^a, 조용주^{a*}

Application of Deep Learning Algorithm for Defect Detection and Cause Analysis of Automotive Parts in Injection Process

Minjae Ko^a, Yongju Cho^{a*}

^aSmart Manufacturing System R&D Department, Korea Institute of Industrial Technology

ARTICLE INFO

Article history:

Received 25 October 2022
Revised 2 December 2022
Accepted 6 December 2022

Keywords:

Injection molding
Artificial intelligence
Deep learning
Defect prediction
Fault cause analysis

ABSTRACT

The injection molding industry is a key field in the manufacturing industry which is projected to expand and improve national competitiveness. However, the domestic injection molding industry is currently experiencing several difficulties, including the aging of skilled engineers and lack of investment in new technology. To address these challenges, this study aims to establish ways to reduce the time and cost required for quality inspection by quickly identifying defects in the injection molding process through defect-prediction and defect-cause analysis using an artificial intelligence algorithm. In addition, by identifying the relationship between major explanatory variables and explanatory variables for defects, we intend to establish a quantitative analysis through fundamental cause analysis of defects to implement a smart factory with an advanced response.

1. 서 론

사출 성형은 플라스틱 성형법 중의 하나로, 열가소성 수지를 가열해서 유동 상태가 되었을 때 금형의 공동부에 가압 주입하여 금형 내에서 냉각시킴으로써, 금형의 공동부에 상당하는 성형품을 만드는 방법이다. 이는 성형 사이클이 짧고 성형 능률이 높아서 자동차 부품, 생활용품 등 다양한 종류의 제품에 대한 대량 생산체제에 적합하고, 높은 정밀도의 제품을 생산하는 데 많이 활용되는 공정이다. 사출 성형은 사출기의 작업조건과 사출 금형의 온도, 원재료의 건조 상태 등 여러 요인에 의해 제품의 품질이 결정되며, 이러한 제품의 품질은 제조경영에 있어 매우 중요한 요소이다. 이러한 사출 성형 산업은 제조 산업에서 핵심 분야로써 국가 경쟁력

향상을 위해서 꾸준히 성장하고 지속되어야 하는 산업이다. 하지만 국내 사출 성형 산업은 숙련된 기술자들의 노령화 및 중소기업들로 이루어진 산업이라는 점, 그리고 신기술 개발에 대한 투자 부족 등 많은 어려움을 겪고 있다. Auto Encoder, Random Forest 등과 같은 인공지능 알고리즘을 활용하여 사출 성형 공정에서 불량요소를 빠르게 파악할 수 있고, 결함 검출을 통하여 제품의 품질을 높일 수 있을 것으로 예상되며, 이를 통하여 사출 공정에서의 업무 효율성에 기여할 수 있을 것으로 사료된다.

제조데이터 분석을 활용한 불량 예측 연구 사례로 이준혁 외는 자동차 엔진 부품을 생산하는 주조 공장을 대상으로 주조 금형 온도 데이터를 활용하여 제품의 불량을 예측하는 Random Forest 기법 기반의 데이터 분석 알고리즘을 개발하고, 개발한 알고리즘

* Corresponding author. Tel.: +82-41-589-8350

E-mail address: yjcho@kitech.re.kr (Yongju Cho).

을 현장에 적용하였다^[5]. 또한 불량 예측 결과를 가시화하고 해당 결과를 활용하여 공장 운영을 지원할 수 있는 대시보드를 개발하였다. 심현식 외는 데이터마이닝 기법 중 로지스틱 회귀 모형을 바탕으로 변수 선택법을 사용하여 PCB 제조라인의 데이터를 분석하여 어떤 공정이 영향을 끼치는지 분석하고, 공정별 분석을 시행하여 불량에 영향을 미치는 핵심 설비 요인에 관한 연구를 수행했다^[3]. 강대현 외는 인공신경망을 활용한 후판 제품에 관한 품질 예측 연구를 진행하였다^[6]. 슬래브(반제품)의 인장강도와 항복강도를 예측할 수 있는 빅데이터 및 기계학습 기반의 시스템을 제안하였고, 주요 예측부에서 도출되는 주요 영향 인자를 인공신경망을 활용하여 불량을 예측하는 연구를 수행하였다.

이와 같은 선행 연구들은 주로 데이터 분석 모델의 학습에 관한 연구이고 데이터 분석 모델의 학습 결과와 현장에서의 데이터 분석 모델 실행 및 시스템 연계는 많이 다루어지지 않았다. 본 논문에서는 사출기 및 금형의 각종 공정변수 데이터를 IoT, 센서, PLC로부터 실시간으로 취합하여, 사출 성형 공정 상태에 따른 품질 상태를 예측하고 불량에 대한 원인 파악을 통하여 제조 현장의 생산성 저하 요인인 불량률, 비가동률을 감소시키는 개선이 이루어지도록 하고자 한다. 본 연구에서 제안하는 분석 방법은 금형 사출 공정, 즉 용융시킨 원료를 금형에 주입하고 냉각시켜 금형을 열어 배출시키며 모양을 완성하는 사출성형공정으로 이루어진 유사 타 제조 현장에도 적용이 가능하며, 사출 제품에 대한 설비 및 공정데이터와 제품의 정상, 비정상 데이터 등이 필요하다.

2. 배경 방법론

본 연구에서 활용하는 주요 분석기법인 Auto Encoder, Random Forest, SHAP(SHapely Additive exPlanations)에 대해서 Table 1에 정리하였다. Auto Encoder는 비지도 방식으로 훈련된 인공신경망이고, SHAP은 기계학습 모델의 결과를 예측하고 설명하기 위한 방법이다. Random Forest는 분류, 회귀 분석 등에 사용되는 양상별 학습 방법이다.

Table 1 Characteristics of 3 analytical techniques
(a) Auto encoder

Description	Algorithm to learn the model approximating the output value to the input value
Strengths	Data compression (saving memory), reducing data dimensions, and extracting attributes
Weakness	Learning loss occurs when normal and abnormal data are similar
Key utilization	Anomaly detection and denoising

(b) SHAP

Description	Algorithm that can explain the predicted value of observation x by calculating the contribution of each feature to the prediction
Strengths	Providing the importance of each variable for individual predictions
Weakness	Calculation slows down when there are many input vectors to predict
Key utilization	Description of the prediction of the model

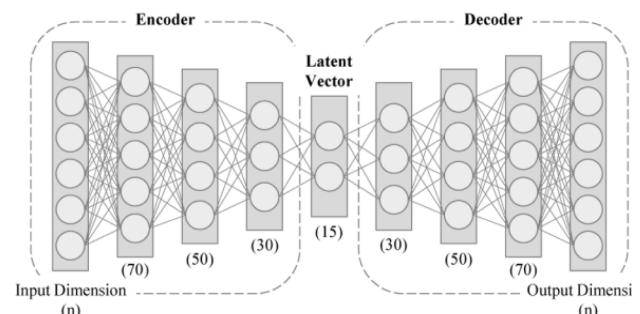
(c) Random forest

Description	A tree-based model that constructs a number of decision trees for classification and uses ensembled results
Strengths	High accuracy and good generalization performance
Weakness	Vulnerable to overfitting and abnormal data
Key utilization	Classification and regression

2.1 Auto Encoder

Auto Encoder 모델은 불량 탐지 및 예측 분야에서 주로 활용되는 딥러닝 알고리즘 중 하나로, 입력 데이터와 같은 출력데이터를 얻기 위하여 입력 데이터의 특성을 요약하여 저차원의 잠재 벡터로 축소하는 인코더와 요약된 잠재 벡터를 다시 복원하는 디코더의 형태로 구성되며 Auto Encoder 모델의 구조는 Fig 1과 같다^[1]. Auto Encoder 모델은 재구축된 데이터와 원본 데이터와의 비교를 통하여 두 데이터 사이의 차이가 작아지도록 학습한다. 이러한 차이는 재구축 오차(reconstruction error)로 정의되며 벡터의 크기를 구하기 위한 유clidean norm(euclidean norm)으로 구성되어 있다^[8].

Auto Encoder는 데이터를 축소하고 복원하는 과정에서 데이터의 패턴을 학습한다. 불량 예측 알고리즘으로 활용하기 위하여 학습 단계에서는 정상 데이터만을 사용하기 때문에 학습된 패턴을 기반으로 정상 데이터를 효과적으로 재구축하지만, 이전에 입력되지 않았던 불량데이터에 대해서는 정확히 복원하지 못하므로 원본 데이터와 재구축 데이터 사이에 큰 재구축 오차가 산출된다. 따라

**Fig. 1 Structure of the auto encoder model**

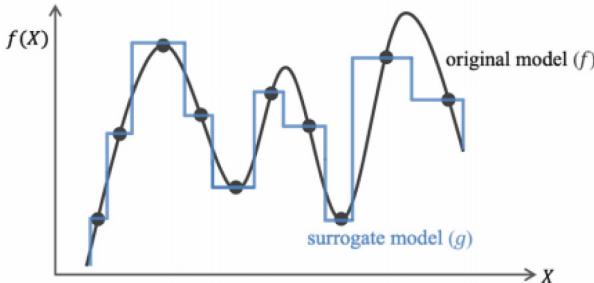


Fig. 2 Structure of the SHAP algorithm

서 Auto Encoder 모델에서는 재구축 오차가 높은 관측치를 이상치로 간주한다^[2].

2.2 SHAP

설명 가능한 인공지능 알고리즘(cXplainable AI, XAI)은 분석에 활용한 딥러닝 모델의 출력 결과를 잘 이해하고 설명하기 위한 방법이다^[7]. 이 알고리즘을 활용하면 분석한 모델의 결과를 정량적으로 산출하여 예측 결과에 대한 직관적인 설명을 제공할 수 있다. SHAP 알고리즘은 전체 관측치를 활용하여 대리 모델을 구성하고 그 결과를 해석하며, 각 변수들이 각각의 예측값을 도출할 때 기여하는 정도를 평가한다^[4]. Fig 2와 같이 설명 대상 모델과 가장 유사한 예측값을 도출할 수 있도록 여러 선형 함수가 결합된 형태의 대리 모델을 구축하고, 이를 해석하여 설명 결과를 제공한다. SHAP 알고리즘에서는 특정 변수를 제거하고 모델을 재구축했을 때 예측 성능이 크게 변화한다면 해당 변수의 중요도가 높음을 가설로 두고 있다. 이때 예측 성능의 변화 정도는 Shapley Value로 정의된다. 특정 변수에 대한 Shapley Value가 클수록 예측값을 도출하는데 많은 기여를 하는 것으로 해석되며, 이는 설명 대상이 되는 관측치에 중요한 변수임을 의미한다.

2.3 Random Forest

Random Forest는 과적합(over fitting)을 방지하기 위해 최적의 기준 변수를 랜덤 선택하는 지도학습 알고리즘으로 결정트리 알고리즘을 앙상블 시킨 모델이다. 검출, 분류, 회귀 등 다양한 문제에서 활용되고 결측치를 다루기도 용이하다. 또한, 대용량 데이터 처리에 효과적이며 모델의 노이즈를 심화시키는 과적합 문제를 회피하여, 모델 정확도를 향상시킨다. Classification 모델에서는 상대적으로 중요한 변수를 선정하고 우선순위를 정할 수 있다. Random Forest는 제일 먼저 Bagging이라는 과정을 거치는데 이는 트리를 만들 때 Training set의 부분집합을 활용하여 트리를 형성하는 것을 말한다. 즉 모든 트리는 각기 다른 데이터를 바탕으로 형성되지만 모두 Training set의 부분집합이 된다^[3]. Random Forest는 높은 정확성과 간편하고 신속한 학습 및 테스트, 다중

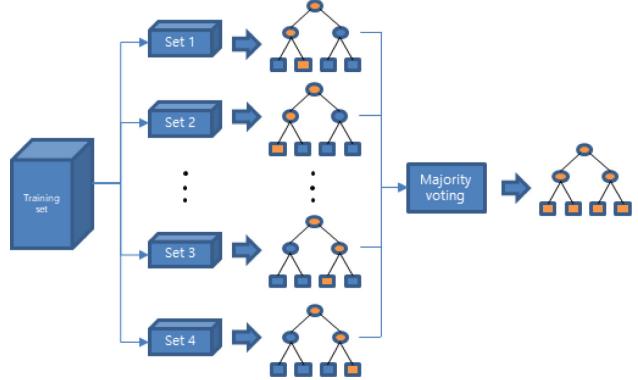


Fig. 3 Structure of the random forest

클래스 알고리즘의 특성이 있으며 기존의 결정트리의 단점을 배깅 및 랜덤 노드 최적화와 같은 방법으로 극복한다.

3. 사출 데이터분석 및 결과

3.1 사출 데이터 분석 개요 및 데이터 수집

사출 데이터 분석 알고리즘의 전체 개요는 Fig. 4와 같다. 사출 기 I/F를 통해 데이터를 수집하고 이를 전처리하여 AI 분석 데이터 셋을 설정한다. 그리고 분석 목표에 따른 AI 학습 모델을 선정하여 상관분석 기반의 다양한 지표들을 개발 및 계산한다. 계산된 지표들을 테스트해서 예측 모델이 실제값과 비교하여 90% 이상 유사 경향을 보일 경우, 해당 AI 모델을 선정하여 현장 불량원인분석 및 품질 예측 알고리즘을 적용하여 실제 작업자들이 필요한 데이터 혹은 지표별 우선순위에 따라 가시화하여 보여준다. 본 연구에 쓰인 데이터는 자동차 부품 Guard Air Upper를 생산하는 사출업체의 생산데이터를 PLC 인터페이스를 통하여 수집하였다. PC 부하로 인한 누락 발생의 위험성을 줄이고자 데이터 수집 주기를 1분 주기로 확정하였으며, 불량데이터는 설비에서 나오는 데이터를 기반으로 수작업과 검증을 통해 불량이 발생한 시점과 불량유형 등을 명기하여 수집하였다.

사출 공정에는 Table 2와 같이 사출 시간, 압력, 온도, 속도, 압력 등의 분류 하에 성형 횟수, 사출 시간, 행정 시간, 사출 압력, 사출 압력, 사출 속도, 보압, 보압 시간, 보압 속도, 계량 거리, Suck-back, 보압 전환, 냉각 시간, 배압, 회전수, 바렐 온도, 오일 온도, Cushion, 사출량, 형개시간, 형폐시간, 충전 시간, 계량 시간, 전환압력, 전환속도, 전환 위치 등 중요데이터 42종의 많은 변수가 존재한다.

이 중 상관관계 분석 등을 통하여 지수가 높은 변수들을 우선순위로 하여 AI 데이터셋 주요 변수로 설정하였으며 각 변수가 정상값을 벗어나면 어떠한 불량이 발생하고 불량을 관리할 때 영향이 있는지 살펴보았다.

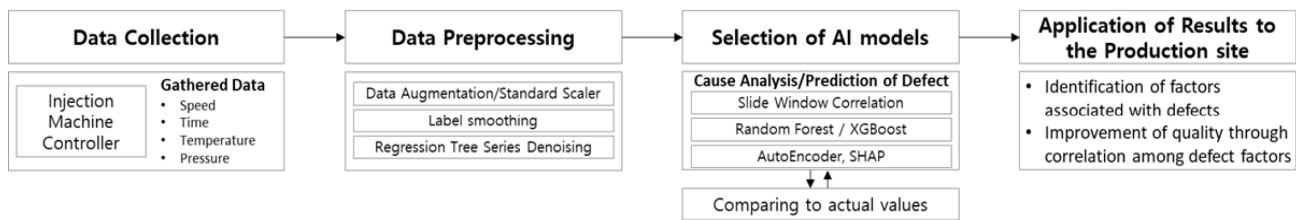


Fig. 4 Procedure for developing and applying AI algorithms

Table 2 Injection process variables

Temperature	Cylinder temperature, Mold temperature, Barrel temperature, Hopper temperature, Oil temperature, Drying temperature etc.
Pressure	Injection pressure, Packing pressure, Max/Average back pressure, Switch over pressure, etc.
Time	Injection time, Packing pressure time, Filling time, Plasticizing time, Cycle time, Clamp close time, Drying time etc.
Position	Supply, Cushion, Suck-back, Clamp open position etc.
Speed	Max injection speed, Max/Average screw RPM, etc.

사출 성형 주요 불량으로는 금형 내에 원료의 충전량이 부족한 경우 발생하는 쇼트숏(short shot), 성형품 주입부 이외의 부분으로 용융 플라스틱이 흘러나와 고화되는 플래시(flash), 수축단계에서 냉각이 지연되어 부분적으로 함몰이 일어나는 싱크마크(sink mark), 성형품 표면에 광택이나 거칠어짐이 발생하는 플로우마크(flow mark), 금형 내 두 개 이상의 유동 선단이 만날 때 선이 생기는 웨ルド라인(weld line), 성형품 표면에 원료의 유동 방향으로 은백색 줄이 생기는 은조흔(silver streak) 외에 흑점, 기포, 콜드슬러그 등이 있다.

작업자는 불량현상의 원인을 정확하게 파악하기 힘들고 불량현상을 해결하기 위하여 많은 시간이 소요된다. 이와 같이 사출 공정에는 많은 변수와 불량현상이 있기 때문에 인공지능을 활용한 딥러닝 알고리즘을 통하여 분석을 수행하면 높은 효율을 보일 것으로 예상된다. 그러나 인공지능 모델의 경우 많은 데이터를 기반으로 학습하여 추후 생산된 성형품의 불량 상태를 확인할 수 있지만 실제 현장에서는 불량품의 수는 매우 적기 때문에 좋은 인공지능 모델을 구현하기가 어려운 경우가 발생한다. 본 논문에서는 이러한 문제를 해결하기 위하여 데이터 전처리 과정에서 데이터 증강(data augmentation)과 표준화(standard scaler)를 적용해 데이터 불균형을 해소하고자 하였다.

3.2 데이터 전처리

설비에서 수집되는 데이터 중 노이즈로 인해 데이터의 품질이 낮을 수 있으므로 이를 해결하기 위해 전처리 기법을 적용한다.

먼저 상관관계 분석을 통해 데이터 간의 상관관계를 수치로 확인할 수 있다. 각 데이터별 상관관계 및 현장 전문가의 지식을 통해 데이터 관계 및 상태 등 데이터에 대한 정보를 확인하고 Fig 5와 같이 전체 데이터 특징들의 상관관계 분석을 통해 데이터 간의 관계를 그래프화하여 데이터의 흐름이 깨지거나 이상 징후가 나타나는 시점 혹은 관련 정보를 탐지할 수 있다.

두 번째로 데이터 표준화를 수행한다. 분석에 사용할 제조데이터는 MPa, mm, sec 등 특성마다 다른 단위로 기술되었기 때문에 그대로 사용하면 학습 속도가 느리며 효율이 떨어지게 된다. 이러한 문제를 해결하기 위해 사용되는 것이 데이터 표준화와 정규화 작업이다. 이는 각기 다른 특성들의 값의 범위를 비슷하게 만들어주는 작업으로 특성 스케일링 또는 데이터 스케일링으로 불린다. 본 연구에서는 사이킷런 함수인 Standard Scaler를 이용하여 표준화를 수행하였다.

세 번째로, 수집한 사출 데이터에서 불량데이터의 비율은 전체 데이터의 약 2% 정도이다. 데이터의 심한 불균형은 모델 학습 성능 저하의 원인이 된다. 이를 해결하기 위해 데이터 증강(data augmentation) 기법을 사용하였다. 이는 데이터의 양을 늘리기 위해 원본에 각종 변환을 적용하여 개수를 증강하는 기법이다. 라벨링이 되지 않은 데이터에 라벨을 붙이고 데이터 특성들의 평균과 표준편차를 구한 후 이를 통해 정규분포를 만들어서 불량데이터 생성 모델을 구현하여 데이터 불균형을 완화해 성능을 향상 시켰다.

3.3 평가지표

본 연구에서는 결합 검출 및 원인분석 알고리즘의 성능을 평가하기 위하여 오차 행렬(confusion matrix)을 사용하였다. 오차 행렬은 Training을 통한 예측 성능을 측정하기 위해 예측값과 실제 값을 비교하기 위한 표이다. Fig. 7에서 보이는 바와 같이 T는 True, F는 False, P는 Positive, 그리고 N은 Negative를 의미한다. 즉, TP와 TN은 실제값을 맞게 예측한 것이며, FP와 FN은 실제값과 다르게 예측한 부분을 의미한다.

전체 중 모델이 바르게 분류한 비율을 정확도(accuracy)라고 하며 직관적으로 모델 예측 성능을 나타내는 지표이다. 그러나 이진 분류인 데이터, 특히 불균형한 데이터 분포의 경우 정확도 수치

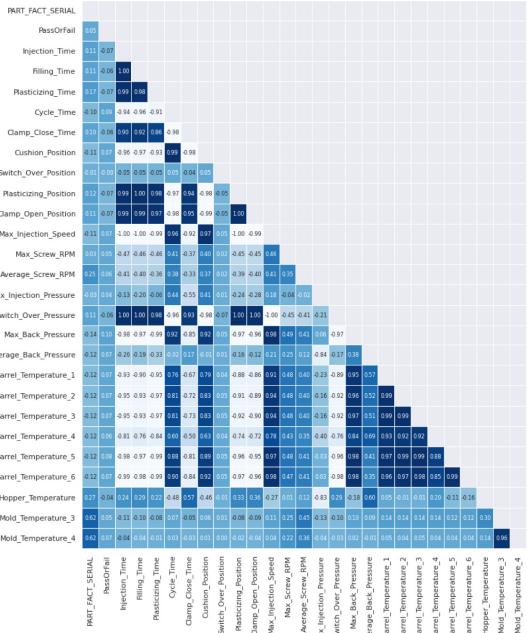


Fig. 5 Correlation analysis of variables

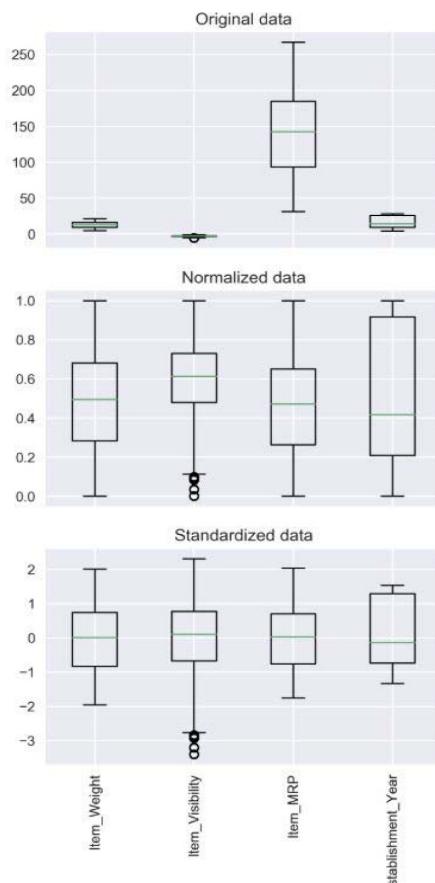


Fig. 6 Data standardization

하나로 성능을 평가하지 않는다. 모델이 Positive라 분류한 것 중 실제값이 Positive인 비율을 정밀도(precision)라고 한다. 예측값

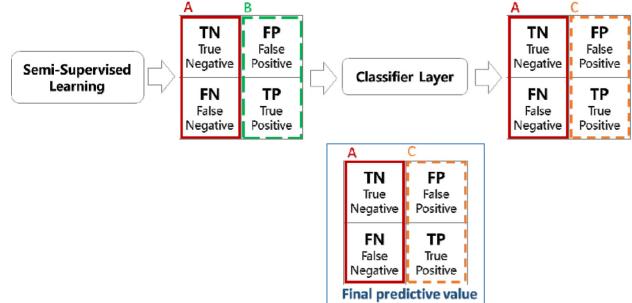


Fig. 7 False prediction process

이 얼마나 정확한지를 나타낸다. 실제값이 Positive인 것 중 모델이 Positive라고 맞게 분류한 비율을 재현율(recall)이라고 하며, 실제 정답을 얼마나 맞추었는지를 평가한다. 마지막으로 정밀도와 재현율의 조화 평균값을 나타내는 F1 Score는 주로 분류 클래스 간의 데이터 불균형이 심각할 때 정확도 대신 사용하며 수치가 높을수록 좋은 모델이라고 판단한다. 각 범주를 예측하는 기준이 되는 임계값의 위치에 따라 정밀도나 재현율이 달라지기 때문에 문제의 상황에 따라 적절한 임계값을 선택할 필요가 있다.

3.4 모델 적용 및 결과

데이터 전처리를 통해 정제된 데이터셋을 결합 검출 모델 및 원인분석 모델에 적용하고 Table 3, 4에 정리하였다.

결합 검출 모델 개발을 위하여 제품의 불량 여부를 판단하는 분류 모델을 생성한다. 분류 모델을 통한 예측값을 활용하여 품질 검사 횟수를 줄이거나 없앨 수 있으며, 품질검사 전 미리 불량을 예측할 수 있다. 먼저 라벨데이터와 비라벨데이터를 사용하여 Auto Encoder를 통하여 비지도 학습을 수행한다. 이렇게 사전 학습된 Auto Encoder를 바탕으로 자기지도학습을 수행한다. Auto Encoder의 마지막 부분에 분류할 수 있는 레이어(classifier)를 추가한 후 라벨데이터만 활용하여 추가학습을 진행하였다.

Fig. 8과 같이 False Negative 가 0건이기 때문에 모델이 실제 불량을 미검출하지는 않는다. 또한 기존 False Positive의 경우를 최대한 줄여서 모델이 판정 결과를 토대로 불량을 검출해야 하는 대상 건수를 줄였다. 본 모델을 이용해서 제품 품질검사의 시간 및 비용을 줄일 수 있다. 불량 비율은 희박하게 발생하지만, 소수라도 불량품 미검출은 이후 더 큰 문제를 야기할 수 있으므로 불량을 검출하기 위한 전수검사가 현장에서 진행되는 경우가 많다. 생산 데이터를 이용하여 불량을 예측해 검수량을 줄일 수 있다면 기업의 수율 향상에 기여할 수 있을 것으로 판단된다.

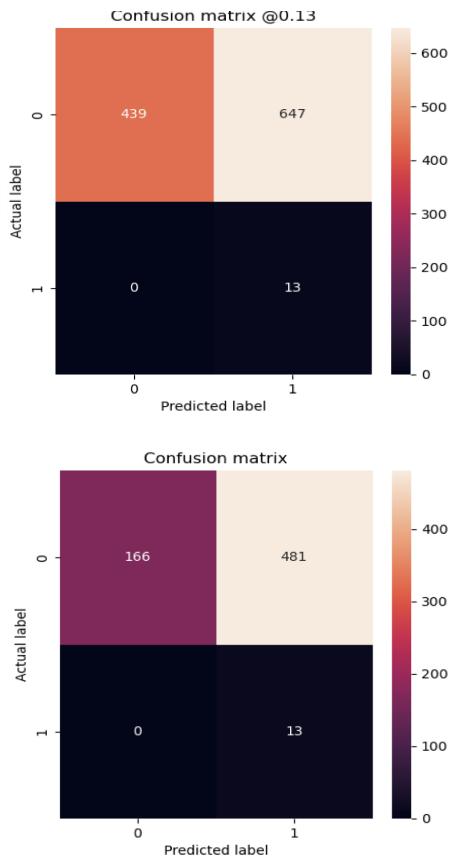
불량에 대한 주요 원인분석은 예측 결과에 영향을 준 변수를 산출하여 불량에 대한 원인분석을 수행할 수 있게 한다. 신경망 기반의 알고리즘은 입력값과 출력값 간의 관계를 알 수 없는 Black box 모델이므로 이를 극복하려는 방안이 필요하다. SHAP 기법을

Table 3 Application of defect detection analysis model

Analysis technique	- Unsupervised learning and self-supervised learning using Auto Encoder
Effect	- Reducing the time and cost of product quality inspection

Table 4 Application of cause analysis model

Analysis technique	- Extracting the main causes of defects using SHAP - Extraction of major and related variables that affect defect using random forest
Effect	- Prevention of defects through management of important factors for defects - Improvement of defects through the relationship between major factors and factors affecting the defect

**Fig. 8 Data confusion matrix**

활용하여 각 변수의 유무에 따라 기존 모델의 예측이 얼마나 많은 영향을 받는지 계산한다. 복잡한 Black-box 모델 f 를 해석이 쉬운 형태의 대리 모델 (surrogate model) g 로 근사하고 각 변수를 추가 및 제거하는 방법으로 각 변수의 중요도를 해석하는 방식을 취한다. 대리 모델 g 를 만들 때는 블랙박스 모델 $f(x)$ 와 $g(x')$ 가 같아지도록 원래의 입력값인 x 를 단순화된 x' 로 바꾸어준다. 그리고 x' 들을 활용하여 만든 대리모델 g 의 가중치(coefficient)를 가지고

Table 5 Major cause factors for defective type ‘gas’

Factor	SHAP value
Average screw RPM	0.127284
Clamp open position	0.069465
Max back pressure	0.031705
Injection time	0.010504

Table 6 Major cause factors for defective type ‘short shot’

Factor	SHAP value
Average screw RPM	0.120784
Clamp open position	0.073884
Max back pressure	0.032684
Injection time	0.008275

중요도를 판단한다. 주요원인분석 결과 발생하는 주요 원인은 파악할 수 있으나 불량유형을 구분할 수는 없다. Table 5, 6은 주요 불량 Gas와 Short shot의 SHAP Value 값에 의한 주요 원인을 나타낸다.

SHAP value는 불량 유형별 관측치를 취합하여 각 변수의 영향도에 절대값을 취해 평균을 낸 것으로 각각의 관측치에 대해 SHAP 분석을 하면 보다 자세한 주요 원인 인자가 산출된다. 불량에 대한 중요 인자로 Average screw RPM이 도출되었는데 사출을 위한 스크류 회전 시 평균 속도에 이상이 있으면 제품 표면에 스크래치나 자국을 발생시킬 가능성이 크다고 고려된다. 이러한 유형의 불량을 방지하기 위해 스크류 속도의 제어를 중점 관리사항으로 볼 수 있다. 또한 Average screw RPM이 측정되기 이전 인자 중 선형관계가 높은 인자들을 탐색하여 스크류 속도에 직접적으로 영향을 미치는 인자를 찾아 조정할 수 있다.

주요 원인분석은 주요인자와 불량과의 관계를 도출한 것이고, 근본 원인분석은 불량에 영향을 미치는 주요 변수와 그에 영향을 미치는 여러 인자를 파악하는 것을 의미한다. 이러한 인자 간의 관계를 통해 불량에 대한 근본 원인이 되는 인자 또는 관련 인자를 제어하여 품질 개선에 도움을 준다. 따라서 데이터 분석 관점에서 필요한 기능은 불량에 대한 주요 인자의 산출과 각 인자 간의 관계를 파악하는 것이다. 그러나 인자 간의 관계를 정확히 파악하기 위해서는 막대한 양의 데이터와 실험이 필요하기에 본 연구에서는 Random Forest를 활용한 상관관계 분석을 수행하였다. 기초적인 방법이지만 데이터양이 적고 데이터 품질이 낮은 중소기업 현실에서 충분히 활용할 만한 가치가 있다고 판단하였다.

주요 원인분석을 통해 강한 상관성을 보이는 인자는 Average screw RPM, Clamp open position, Max back pressure, Injection time 등이 있음을 확인하였고 이 인자들이 양/불 여부에 영향을 끼치는 근본 원인과 어떤 상관관계를 가지는지

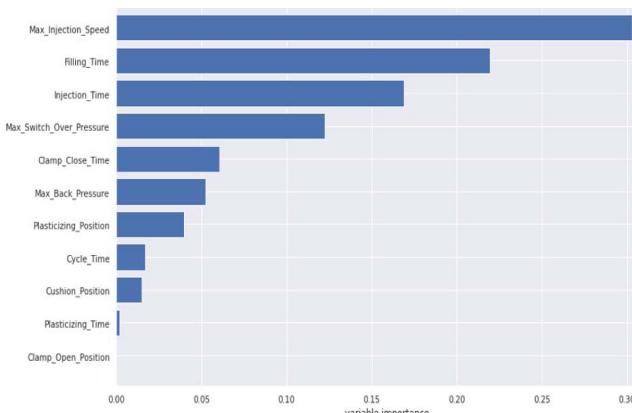


Fig. 9 Data features by random forest

Table 7 Root cause analysis for defective type ‘gas’

Factor	SHAP value	Related factor	Related factor
Average screw RPM	0.127284		
Clamp open position	0.069465		
Max back pressure	0.031705	Cushion position(0.997)	Injection time(-0.982) Filling time(-0.98) Max Injection speed(0.989)
Injection time	0.010504		

Table 8 Root cause analysis for defective type ‘short shot’

Factor	SHAP value	Related factor	Related factor
Average screw RPM	0.120784		
Clamp open position	0.073884		
Max back pressure	0.032684	Cushion position(0.996) Plasticizing position(-0.891)	Injection time(-0.974) Filling time(-0.971) Max Injection speed(0.983)
Injection time	0.008275		

Random Forest 모델을 통해 도출하였다. 결과에 영향을 끼친 주요 특징들을 추출하면 Fig. 9와 같다. 상관관계가 있는 변수를 선별하고, 사전에 파악한 공정 지식을 통해 변수 간의 선후 관계를 파악하여 Table 7, 8에 정리하였다. Gas 불량에서의 주요 원인 Max back pressure와 가장 관련 있는 인자는 Cushion position이며, Cushion position에 관련 인자는 Injection time, Filling time, Max Injection speed 등이 있다. 이 결과를 통해 실제 제어할 수 있는 인자를 살펴보고 관계를 조절하여 불량 개선을 수행할 수 있다.

4. 결 론

본 연구에서는 금형 산업의 사출성형공정 고도화를 위하여 제품의 결함 검출과 그 원인분석 연구를 수행하였다. Auto Encoder를 통하여 불량이 극히 적은 불균형 데이터로 예측 모델을 생성하고, 이를 통해 제한된 환경에서 학습데이터에 포함되지 않은 장비와 제품에서 생성된 데이터도 예측할 수 있도록 하였다. 이러한 결함 검출 모델 분석을 통해 품질검사 시간과 비용을 줄일 수 있을 것으로 보인다. 또한 Random Forest 및 SHAP 분석 방법을 통하여 불량에 대한 주요 설명변수 및 설명변수 간의 관계를 파악하여 불량에 대한 근본 원인분석을 통해 정량적인 분석 기반을 마련하여 사전 대응이 가능하게 함으로써 제조공정을 최적화하여, 궁극적으로 제조기업의 수율 개선 활동에 도움을 주고자 하였다.

결함 검출과 같은 품질 예측은 제조공정에서 발생하는 데이터를 활용하여 산출물에 대한 계측값을 예측하는 방법으로 불량 확률이 높다고 예측된 제품에 대해서 선택적으로 품질검사를 수행하게 되면, 실제로 전수검사를 수행하는 수준의 품질관리 결과를 기대할 수 있다. 또한 품질에 영향을 미친 변수를 파악하여 원인 탐색이 수월하다. 기존에는 노련한 작업자의 감에 의존하거나 공정을 멈춰 장비를 확인하였으나 데이터와 예측 모델을 통해 원인을 빠르게 파악할 수 있다.

근본 원인분석은 문제의 원인을 중간 단계에서가 아닌 직접적인 지점 파악하여 이에 대한 조치를 통해 가장 효율적인 대책과 개선책을 마련할 수 있는 분석방법이다. 데이터 분석을 통해 불량과 상관관계가 높은 인자 뿐만 아니라 인과관계를 파악하면 정량적으로 불량에 가장 큰 요인을 확인할 수 있고 개선점을 바로 파악할 수 있다는 장점이 있다.

향후 데이터 양이 증가하고 보다 고도화된 데이터 분석기법을 적용할 경우 좀 더 정교한 예측 및 원인분석이 가능해질 것이다.

후 기

이 연구는 중소벤처기업부의 스마트제조혁신기술개발사업(R&D) “자동차 부품산업을 위한 지능형 가치사슬 플랫폼 기술개발(RS-2022-00140875)” 과제의 지원을 받아 작성하였습니다.

References

- [1] Rumelhart, D. E., Hinton, G. E., Williams, R. J., 1985, Learning Internal Representations by Error Propagation, Institute for Cognitive Science (ICS Report 8506), CA, U.S.A..
- [2] An, J., Cho, S., 2015, viewed 29 Nov. 2022, Variational

- Autoencoder Based Anomaly Detection using Reconstruction Probability, Special Lecture on IE, SNU Data Mining Center, <<http://dm.snu.ac.kr/static/docs/TR/SNUDM-TR-2015-03.pdf>>.
- [3] Shim, H. S., Kim, C. W., 2015, Fault-Causing Process and Equipment Analysis of PCB Manufacturing Lines Using Data Mining Techniques, KIPS Transactions on Software and Data Engineering, 4:2 65-70, <https://doi.org/10.3745/KTSDE.2015.4.2.65>.
- [4] Lundberg, S. M., Lee, S.-I., 2017, A Unified Approach to Interpreting Model Predictions, Proceedings of the 31st International Conference on Neural Information Processing Systems, 4768-4777.
- [5] Lee, J., Noh, S. D., Kim, H. J., Kang, Y. S., 2018, Implementation of Cyber-Physical Production Systems for Quality Prediction and Operation Control in Metal Casting, Sensors, 18:5 1428, <https://doi.org/10.3390/s18051428>.
- [6] Kang, D. H., Oh, S. G., Park, J. H., Seo, M. S., 2019, Prediction of Quality of Backplate Products Using Artificial Neural Network, Proceedings of the Korean Society for Quality Management Autumn Conference, 146.
- [7] Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F., 2020, Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI, Inf. Fusion., 58, 82-115, <https://doi.org/10.1016/j.inffus.2019.12.012>.
- [8] Kim, S. Y., Lee, J. Y., Mok, C. H., Kim, S. H., Moon, S. H., Kyeong, Y. Y., Jin, Y. G., Lee, W. J., Choi, J. M., Kim, S. B., 2022, Defect Detection and Cause Analysis on Display Electrostatic Chuck FAB Process Using Deep Learning Algorithms, J. Soc. Korea Ind. Syst. Eng., 48:2 151-162, <https://doi.org/10.7232/JKIIE.2022.48.2.151>.



Minjae Ko

Senior researcher of Smart Manufacturing System R&D Department in Korea Institute of Industrial Technology. Her research interests are the analysis and systematization of manufacturing data.

E-mail: minjko@kitech.re.kr



Yongju Cho

Principal researcher of Smart Manufacturing System R&D Department in Korea Institute of Industrial Technology. His research interest is the development of AI use case using manufacturing data.

E-mail: yjcho@kitech.re.kr