

# A Study on Defect Prediction through Real-time Monitoring of Die-Casting Process Equipment

Chulsoon Park · Heungseob Kim<sup>†</sup>

Department of Industrial & Systems Engineering, Changwon National University

## 주조공정 설비에 대한 실시간 모니터링을 통한 불량예측에 대한 연구

박철순 · 김흥섭<sup>†</sup>

창원대학교 산업시스템공학과

In the case of a die-casting process, defects that are difficult to confirm by visual inspection, such as shrinkage bubbles, may occur due to an error in maintaining a vacuum state. Since these casting defects are discovered during post-processing operations such as heat treatment or finishing work, they cannot be taken in advance at the casting time, which can cause a large number of defects. In this study, we propose an approach that can predict the occurrence of casting defects by defect type using machine learning technology based on casting parameter data collected from equipment in the die casting process in real time. Die-casting parameter data can basically be collected through the casting equipment controller. In order to perform classification analysis for predicting defects by defect type, labeling of casting parameters must be performed. In this study, first, the defective data set is separated by performing the primary clustering based on the total defect rate obtained during the post-processing. Second, the secondary cluster analysis is performed using the defect rate by type for the separated defect data set, and the labeling task is performed by defect type using the cluster analysis result. Finally, a classification learning model is created by collecting the entire labeled data set, and a real-time monitoring system for defect prediction using LabView and Python was implemented. When a defect is predicted, notification is performed so that the operator can cope with it, such as displaying on the monitoring screen and alarm notification.

**Keywords :** Condition based Monitoring, Machine Learning, Predictive Maintenance, Sensor Data, Real-time Data Acquisition and Analysis

### 1. 서 론

최근 IOT 센서 기술과 인공지능 기술을 바탕으로 제조 시스템의 지능화를 위한 많은 연구가 이루어지고 있다[8, 9, 10, 17, 18]. 하지만, 국내 제조산업의 근간이라 불리는

대표적인 뿌리산업 중 하나인 다이캐스팅 주조업체들은 영세한 기업이 많고, 열악한 작업환경으로 인해 신규인력 유입이 적고 외국인 노동자로 대체되고 있다. 따라서 생산 및 품질 기술 지식 전수의 단절, 설비문제에 대한 미흡한 대응으로 산업 전반의 품질 및 수익성이 저하되는 악순환이 반복되고 있는 실정이다. 현재 대부분의 다이캐스팅 업체에서는 불량 발생시 체계적인 접근보다는 현장 전문가가 시행착오를 반복하며 문제를 해결하기 때문에 불량 대

Received 25 November 2022; Finally Revised 12 December 2022;  
Accepted 13 December 2022

<sup>†</sup> Corresponding Author : heungseob79@changwon.ac.kr

처가 체계적이지 못하고 경험과 지식을 갖춘 경력자의 은퇴로 인해 대응에 어려움을 겪고 있다. 다이캐스팅 공정에서는 제품을 성형할 때 투입되는 소재량의 불균형, 금형 온도 변화, 용탕 청정도 확인 미흡, 주조 압력 및 냉각 조건 임의 변경 등 조건 설정 오류 등에 의해 불량 발생 수 있다. 또한 제어장치 고장, 센서 오류 등에 의해 주조 파라미터의 설정치와 실적치에 차이가 발생하는 등의 원인으로 불량이 발생할 수 있다. 다이캐스팅 주조 현장에서는 <Table 1>과 같은 불량 대응방안 예시에서 보여주는 것처럼 불량 발생시 작업자의 경험과 지식에 의존하여 대응하고 있다. 즉, 주조 업계에서는 이와 같은 주조조건 파라미터 관리의 어려움으로 인한 주조결함 발생, 현장 작업자들의 표준화된 작업조건의 임의 변경으로 인한 공정 불량률 증가 및 공장 가동률 저하 등의 문제 발생에도 불구하고 이를 위한 체계적인 방법론 및 시스템의 부재로 어려움을 겪고 있다[15]. 따라서, 다이캐스팅 공정에서 품질예측을 위해 주조 파라미터 모니터링 등을 통한 제품 불량률 예측하기 위한 다양한 연구들이 수행되어왔다. 자동차 엔진 부품을 생산하는 주조 공장을 대상으로 주조 금형 온도 데이터 모니터링을 통해 제품의 불량률 예측하기 위해 앙상블 기반 랜덤포레스트(Random Forest) 알고리즘을 개발하여 현장에 적용한 사례가 있다[19]. 다이캐스팅 제품의 불량 검출을 위하여 비지도 학습 기반 불량 검출 모델을 제안한 사례도 있다[20].

다이캐스팅을 포함한 제조업 분야에 머신러닝 알고리즘을 적용하기 위한 대량 데이터 수집 및 라벨링 작업을

위해서는 많은 비용이 수반되기 때문에 대안을 찾기 위한 연구들이 수행되어왔다. 불량 여부 뿐만 아니라 불량 타입을 예측하기 위해서는 라벨링 작업 후에 분류모형을 개발해야 하는데 다이캐스팅 공정의 경우 주조 후에 열처리, 사상 등의 후공정 작업이 수행되고 이러한 후공정 작업중에 주조 불량 여부가 판단되기 때문에 주조단계에서 수집된 파라미터 데이터와 불량품 간의 매칭을 통한 라벨링을 하는 것이 어려운 실정이다. 따라서 다이캐스팅 공정 시뮬레이션을 통해 학습 데이터를 생성하고 생성한 데이터를 이용하여 LSTM 기반 시계열 모델을 적용하여 모델 학습을 수행한 연구도 있다[21]. Kittur et al.[16]은 다이캐스팅 공정의 품질을 예측하는 머신러닝 모형을 만들기 위해 다이캐스팅 파라미터들에 대한 회귀식을 이용해서 학습 데이터셋을 생성한 후에 이 데이터를 이용하여 역전파신경망 모형을 만든 후에 고압 다이캐스팅 공정에 활용하는 방안을 제시하기도 하였다. Lee et al.[22]은 금형 온도 등의 공정변수를 이용하여 인공신경망(Artificial Neural Network, ANN) 모델을 개발하여 주조결함을 예측하는 연구를 수행하였다.

따라서, 다이캐스팅 주조 제품은 주조를 마친후 열처리, 사상가공 등의 후공정이 끝나야 결함 여부를 정확히 알 수 있으므로 주조 작업시 미리 조치하지 못해 대량 불량 발생의 원인이 될 수 있다. 즉, 주조 시점에서 육안검사 등 비전검사를 수행하기 어려운 다이캐스팅 주조공정에서 실시간으로 수집 가능한 주조 파라미터를 이용해서 불량 타입별로 불량 여부를 판단할 수 있는 방안이 요구되고 있다.

<Table 1> Examples of Die-casting Defects

Defect type	Cause of defect	Prognostic factor	Countermeasure
Plunger tip/Sleeve damage	- Casting velocity unevenness - Back flash	- Plunger tip/Sleeve life/Plunger oil shortage	- Vision inspection - Plunger tip noise/Sleeve abrasioncheck(1ea/Shift)
Low casting pressure	- Bubble, Short shot	- N2 gas Leakage	- Real time monitoring of casting pressure
High casting pressure	- Fusion, deform, mold damage	- High bath temperature - Spray nozzle position setting error - Spray nozzle blockage	- Opening check of Spray nozzle position/blockage - Daily report logging (warm-up shot)
Low mold temperature	- Bubble, Short shot	- Mold preheating lack	- Daily report logging (warm-up shot)
Low vacuum	-Blister	- Vacuum block blockage	- Vacuum block replacement before warm-up(daily)
Deformation	- Work size error	- Spray nozzle position setting error - Spray nozzle blockage	- Opening check of SPRAY NOZZLE position/blockage - Daily report logging (warm-up shot)
Biscuit thickness defect	- Short shot	- Mold moisture inconsistency	- Biscuit thickness check (27± 7mm)
Eject pinprotrusion	- Assembly size error	- Eject pin length shortage	- Initial off-set check - Patrol check(+0.2/-0.5)
Appearance defect	- Customer complain	- Casting work standard deviation	- Daily report logging - Total appearance inspection
Low density	- Soldering defect	- Density control inconsistency	- Blister test 2EA/Shift - Casting work standard

제조분야에서 불량 발생 등 정상 패턴을 따르지 않는 상황을 구분하기 위해서는 대표적인 비지도학습 머신러닝 기법의 하나인 이상치 탐지 (Anomaly detection) 알고리즘을 일반적으로 사용할 수 있다[25]. 데이터 이상치를 찾는 문제에는 다양한 연구가 이루어져 왔으며 최근에는 머신러닝을 이용한 이상치 탐지 분야에서 많은 연구가 이루어지고 있다. 이상치 탐지 기법에는 몇 가지 어려움이 있는데, 첫째는 데이터 불균형 발생 문제이다. 일반적으로 이상치 데이터가 정상 데이터에 비해 훨씬 적기 때문에 학습 편향이 발생할 수 있다. 이러한 이슈를 다루기 위해 데이터 마이닝과 머신러닝 분야에서 많은 연구가 이루어져 왔다[6, 14, 26, 32]. 둘째, 이상치 탐지 (Outlier Detection) 모형의 학습 및 검증에 대해서는 정확하게 라벨링된 데이터가 필수적인데 이러한 데이터를 얻기 위한 라벨링 방법에 대해 많은 연구에서 주요한 이슈로 다루어 왔다[1, 5, 28, 29]. 또한, 이상치 탐지를 위한 Isolation Forest, LOF(Local Outlier Factor), One Class SVM, Autoencoder 등의 머신러닝 알고리즘들에 대해 많은 연구가 이루어져 왔다[4, 5, 31].

본 논문에서는 확보 가능한 불량률을 하이퍼파라미터로 사용하여 라벨링 작업을 수행할 수 있는 Isolation Forest 모형을 이용하여 이상치인 불량 데이터셋을 분리해내는 접근방법을 제시한다. 불량 여부 라벨링이 완료된 데이터셋에 대해 불량 타입별로 불량예측을 수행하기 위해서는 불량 데이터셋에 대해 불량 타입별로 군집을 나누고 불량 타입별 예측 진단을 위해 불량 타입 라벨링 작업을 다시 수행할 필요가 있다.

불량 타입 특성과 같이 유사성을 가지는 데이터들을 그룹 또는 클래스로 묶기 위한 방법을 제공하는 군집분석 (Clustering Analysis)은 데이터 마이닝과 머신러닝 분석의 핵심 기술중 하나로 간주되고 있으며 K-Means, DBSCAN, GMM(Gaussian Mixture Models) 등의 알고리즘을 포함하고 있다[2, 13, 24]. K-Means는 클러스터 센터 개수로 하이퍼파라미터  $k$  값을 지정하고 각 데이터와 클러스터 센터 사이의 거리를 계산해서 가장 가까운 클러스터에 데이터를 할당하는 방식으로 군집을 형성토록 한다. K-Means 알고리즘은 초기 클러스터 개수인  $k$  값을 지정해야 한다는 단점이 있고, DBSCAN은 클러스터 개수를 미리 지정할 필요는 없지만 클러스터의 밀도가 다양할 경우 성능이 나쁘다고 알려져 있으며, GMM은 K-Means 보다는 유연하고 클러스터 개수 지정이 가능하지만 데이터가 Gaussian 분포를 따름을 가정해야 하는 단점이 있다[3, 11, 12, 23, 30].

제조 분야에 군집분석을 적용한 사례로 Seo [27]에서는 설비가 운전중에 정상상태로부터 이탈여부를 탐지하기 위해 군집분석 기반 설비이상 진단 절차를 제안하였고 현장

에서 수집된 데이터를 통해 설비의 비정상 상태를 탐지할 수 있음을 보였다. Choi et al. [7]에서는 제조 공정 데이터의 불량 위치의 검출을 위해 K-means와 DBSCAN 군집분석 알고리즘을 적용해서 PCB 관련 공정 데이터를 기반으로 어떤 클러스터링 알고리즘이 PCB의 불량 위치를 더 효과적으로 분할하여 불량 밀집 구역을 더 잘 표현할 수 있는지 실험하였고, 최적의 클러스터를 형성하기 위한 파라미터 값을 결정하기 위한 성능평가 실험을 수행하였다.

본 연구에서는 다이캐스팅 주조공정의 설비에서 실시간으로 수집되는 주조조건 파라미터 데이터에 대해 전체 불량률 및 불량타입별 비율을 기반으로 라벨링 작업을 수행한 후 머신러닝 기술을 이용하여 불량 타입별 주조 불량 발생을 예측할 수 있는 방안을 제시하고자 한다. 적용 대상기업의 주조공정 파라미터 데이터는 기본적으로 주조설비 컨트롤러를 통해 접근 가능하며 MES의 데이터베이스에 실시간으로 저장되고 있다.

불량 예측을 위한 분류분석(Classification)을 수행하기 위해서는 주조공정 파라미터에 대한 라벨링 작업을 수행해야 한다. 불량 타입별 불량 예측을 수행하기 위해서는 수집한 데이터에 대한 라벨링 작업이 선행되어야 한다. 첫째, 양불 라벨링 작업을 위해서 후가공 업체에서 식별되는 전체 불량률 정보를 Isolation Forest 이상치 탐지 모형의 하이퍼파라미터로 사용하여 양불 라벨링 작업을 수행한다. 둘째, 분리된 불량 데이터셋을 대상으로 불량 타입 개수를 K-Means 군집분석 모형의 하이퍼파라미터로 이용하여 불량 타입별로 군집을 나눈후에 불량 타입 라벨링 작업을 수행한다. 마지막으로 라벨링을 마친 전체 데이터셋을 취합하여 분류학습 모델을 만든다. 불량발생 및 불량타입 예측을 위한 최적모형을 현장에 배치하여 실시간으로 수집되는 주조공정 파라미터를 이용해서 모니터링 작업을 수행한다. 불량이 예측될 시 모니터링 화면에 표시 및 알람통지 등 작업자가 대처할 수 있도록 정보를 제공한다. 본 연구에서는 LabView와 Python 환경에서 불량예측을 위한 실시간 불량 예측 모니터링 시스템을 구현한다.

## 2. 데이터 전처리

본 연구에서 수집 가능한 다이캐스팅 파라미터 데이터는 결측치와 고정값 컬럼들을 제외하면 <Table 2>와 같이 15개의 파라미터로 구성되어 있다.

본 연구에서는 머신러닝 모형 개발을 위해 약 1개월간의 주조 파라미터에 대한 실측치 데이터를 수집하였다. 공정변수별 Scatter plot을 통해 시계열 분석을 수행해 보면 <Figure 1>과 <Table 3>과 같다. <Figure 1>에서 가로축의 Production Time은 다이캐스팅 샷 (Die-Casting shot)을 의

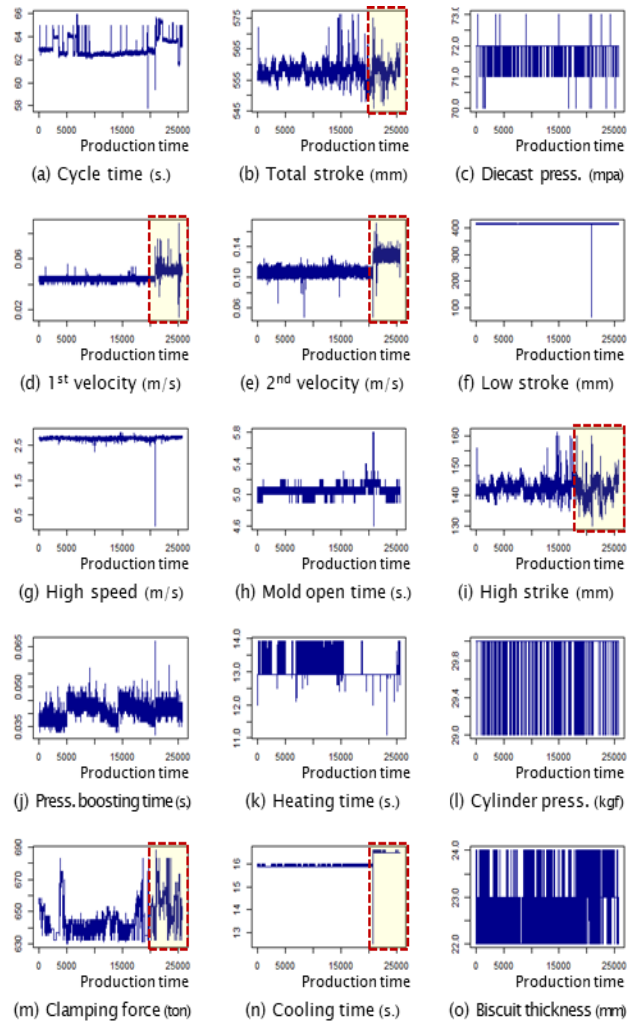
<Table 2> Die-Casting Parameters

No	Column	Unit	Sampling	Sze	Source	ref
1	Cycle Time	Sec	1ea/shot	Double byte	PLC	MES
2	Total Stroke	mm	1ea/shot	Double byte	PLC	MES
3	Casting Pressure	MPA	1ea/shot	Double byte	PLC	MES
4	1st Velocity	m/s	1ea/shot	Double byte	PLC	MES
5	2nd Velocity	m/s	1ea/shot	Double byte	PLC	MES
6	Low stroke	mm	1ea/shot	Double byte	PLC	MES
7	High speed	m/s	1ea/shot	Double byte	PLC	MES
8	Mold open time	Sec	1ea/shot	Double byte	PLC	MES
9	High stroke	mm	1ea/shot	Double byte	PLC	MES
10	Pressure boost time	Sec	1ea/shot	Double byte	PLC	MES
11	Heating time	Sec	1ea/shot	Double byte	PLC	MES
12	Cylinder pressure	Kgf	1ea/shot	Double byte	PLC	MES
13	Clamping force	Ton	1ea/shot	Double byte	PLC	MES
14	Cooling time	Sec	1ea/shot	Double byte	PLC	MES
15	Biscuit thickness	mm	1ea/shot	Double byte	PLC	MES

미하며 세로축 공정변수 명칭은 각 그림에서 그림제목으로 표기하였다. 그림에서 확인할 수 있는 것처럼 전체스트로크, 1속도, 2속도, 형체력, 냉각시간 등의 일부 공정 파라미터에서 20일차부터 운영조건의 변화가 발생했음을 확인할 수 있다. 즉, 생산일자 기준으로 1차 블록 (1~19일)과 2차 블록 (20~24일)으로 분리하여 공정변화에 대한 유의성을 확인할 필요가 있다. <Figure 2>는 공정조건 변화에 대한 유의성을 판단하기 위해 수행한 Box-Plot 그래프 분석 결과를 보여주고 있다. 여기서 1속도(m/s), 2속도(m/s), 형체력(ton), 냉각시간(sec)에서 수집한 조건 파라미터의 시계열에서 유의미한 공정변화가 발생했음을 확인할 수 있다. 즉, 예비적 군집분석을 수행한 결과 20~24일차 데이터 블록에서 다수가 불량품으로 오분류되는 결과를 초래하고 있음을 발견하였다. 이로 인해 20~24일차(5일) 데이터 블록의 경우 공정 운영조건 변화로 인해 군집분석 및 분류모형 학습에 혼선을 초래할 가능성이 예상되므로 학습모형을 개발할 때는 공정조건에 따른 정상/불량품 및 불량원인 판별 모형 개발에는 1~19일차 데이터 블록만을 활용하도록 한다.

<Figure 3>에서는 1~19일차 데이터 블록의 시계열을 보여주고 있으며 유의미한 공정변화가 없음을 확인할 수 있다.

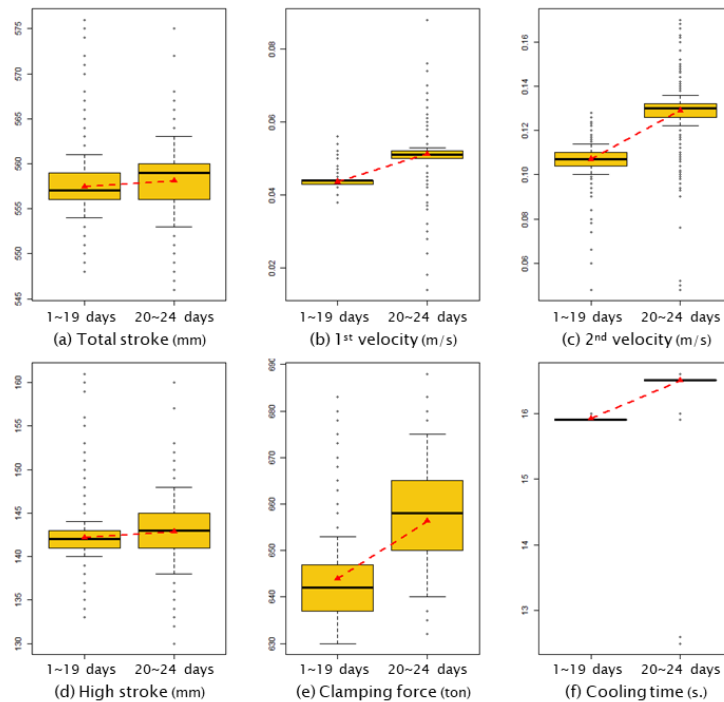
전처리 과정에서 마지막으로 공정변수들에 대한 차원축소를 고려하기 위해서 상관분석을 수행하였다. <Figure 4>에서 볼 수 있듯이 상관관계수 절대값이 0.8 이상인 경우가 전체스트로크(V2)와 고속스트로크(V9), 주조압력(V3)과 실린더압력(V12)의 경우에서 관측되었다. 따라서 본 연구에서는 고속스트로크(V9)와 실린더 압력(V12)을 제거하고 전체스트로크(V2)와 주조압력(V3)을 선택하여 13개 주요 파라미터를 이용하여 머신러닝 모형을 개발하도록 한다.



<Figure 1> Scatter Plot Analysis

<Table 3> Process Condition Change Analysis

Parameter	V1		V2		V3		V4		V5	
	Cycle time		Total Stroke		Diecast Pressure		1st Velocity		2nd Velocity	
Date Block	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24
Mean	62.7761	63.8735	557.4258	558.0898	71.9919	71.9939	0.0436	0.0512	0.1070	0.1290
Stddev	0.5574	0.8396	2.0906	3.1454	0.0964	0.0941	0.0012	0.0028	0.0040	0.0064
CV	0.0089	0.0131	0.0038	0.0056	0.0013	0.0013	0.0267	0.0545	0.0375	0.0496
Parameter	V6		V7		V8		V9		V10	
	Low Stroke		High Speed		Mold open time		High Stroke		Pressure Boosting time	
Date Block	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24
Mean	415.1333	414.7038	2.7049	2.7276	5.0346	5.0723	142.1602	142.8346	0.0410	0.0417
Stddev	0.3399	12.1846	0.0244	0.0838	0.0639	0.1279	2.0947	3.1440	0.0030	0.0021
CV	0.0008	0.0294	0.0090	0.0307	0.0127	0.0252	0.0147	0.0220	0.0724	0.0502
Parameter	V11		V12		V13		V14		V15	
	Heating time		Cylinder Pressure		Clamping Force		Cooling Time		Biscuit Thickness	
Date Block	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24	1~19	20~24
Mean	13.0104	12.9092	29.9939	29.9956	643.9362	656.3242	15.9232	16.4971	22.6294	22.9453
Stddev	0.2679	0.1053	0.0777	0.0665	8.5964	11.0575	0.0422	0.1611	0.6323	0.5295
CV	0.0206	0.0082	0.0026	0.0022	0.0133	0.0168	0.0027	0.0098	0.0279	0.0231



<Figure 2> Result of Significance Analysis of Die-Cast Parameters

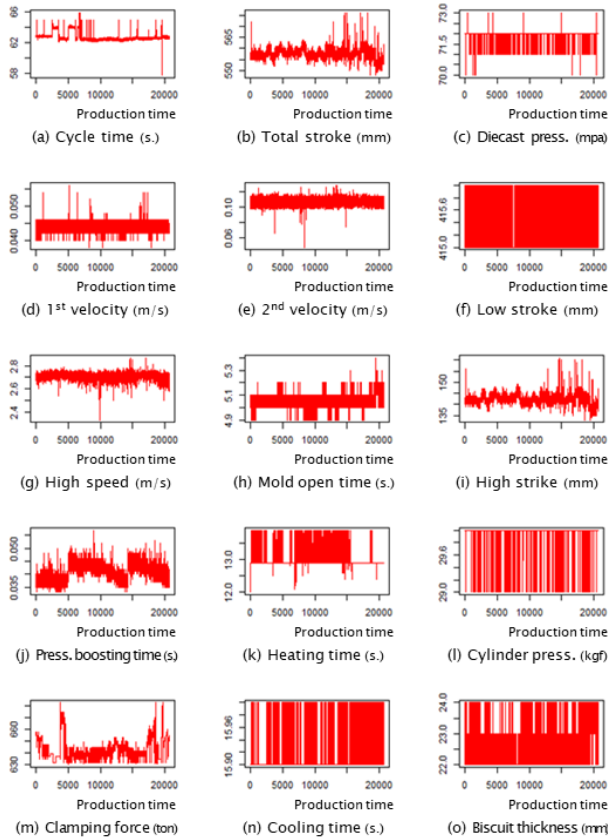
### 3. 머신러닝을 이용한 예측모델 개발

#### 3.1 군집분석을 이용한 라벨링 작업

불량 예측/판단 및 불량원인 분석 알고리즘을 개발하기 위해서는 분류학습(Classification) 모형을 개발해야 하므

로 수집한 공정 데이터에 대한 라벨링 작업이 필요하다. 수집한 주조 파라미터 데이터와 1차 후가공 업체의 열처리 작업에서 식별된 주조 불량률 정보를 바탕으로 주조단계에서는 식별되지 못하는 양품·불량품에 대한 공정조건을 추정하기 위해서 이상치 탐지 기술을 적용하였다. 대표적인 이상치 탐지 알고리즘들에는 Isolation Forest, LOF,

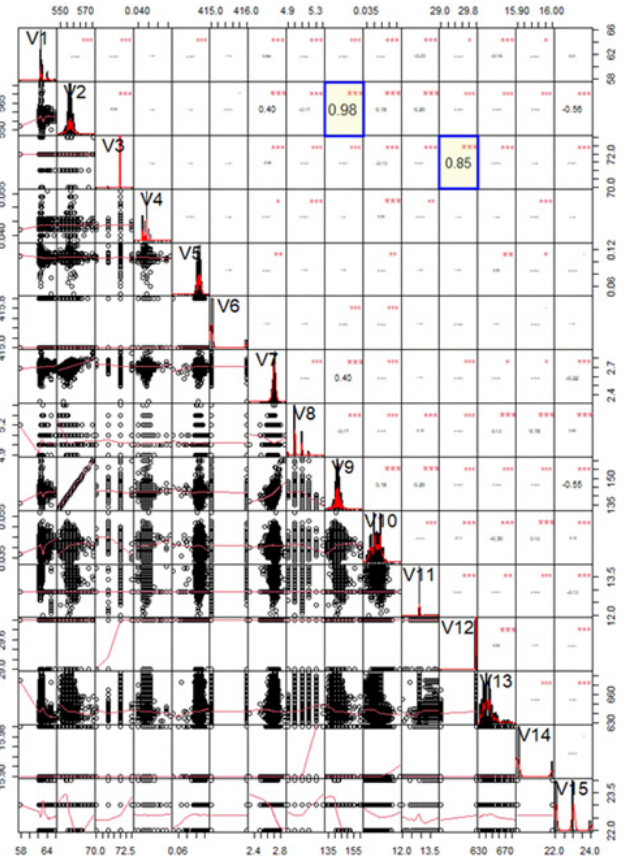
One-Class SVM, Autoencoder 등이 있으며 여기서는 열처리 작업에서 확보된 불량률을 활용하여 1차로 Isolation Forest 군집분석 알고리즘을 통해 양품과 불량품 그룹으로 분류하였다.



<Figure 3> Validation Analysis with 1st Time Block Data

즉, Isolation Forest 알고리즘에서 하이퍼파라미터인 Contamination rate에 불량률을 적용하면 양품 군집, 불량품 군집과 같이 2개의 군집으로 분류할 수 있다. <Table 4>는 Isolation Forest 알고리즘을 이용한 양품/불량품의 2-Class 라벨링 결과를 보여주고 있다. 표에서 Accuracy는 오분류 비율 값이며 수량 분류정확도를 보여주고 있다.

한편 불량발생 예측시 불량타입 판별을 위해서는 불량품 군집을 대상으로 불량타입별로 군집을 소분류 할 필요가 있다. 여기서는 하이퍼파라미터로 군집수를 할당할 수 있는 K-Means 알고리즘을 이용해서 4가지 불량타입으로 군집분석을 수행하였다. <Table 5>는 K-Means를 이용해서 4가지 불량타입별 군집분석을 수행한 결과를 보여주고 있다. 수량 분류정확도와 현장 전문가의 의견을 참고하여 분류 결과를 활용하기에 큰 문제가 없음을 확인하였다.



<Figure 4> Result of Correlation Analysis

<Table 4> 2-Class Labelling Result

Category	Good	Defective	Total
Actual(A)	20,136	601	20,737
Predicted(B)	20,132	605	20,737
Accuracy	0.02%	0.67%	-

<Table 5> 4-Class Labelling Result with Defectives

Category	LUB HOEL	PR Valve & Blow	TCC Valve & Blow	General Bubble	Total
Actual(A)	140	119	38	304	601
Predicted(B)	183	126	41	255	605
Accuracy	14.4%	5.9%	7.9%	10.2%	

### 3.2 머신러닝 학습모형 개발

먼저, 이상치 탐지를 통해 확보한 양품/불량품 군집 데이터에는 양적 불균형(Data Imbalance)이 존재하고 있다. 즉, 양품 데이터 개수에 비교해서 불량 데이터 개수가 심하게 적기 때문에 이로 인한 기계학습 편향(Bias)이 발생할 수 있으므로 이를 제거하기 위해 Synthetic Minority Oversampling Technique(SMOTE)을 이용해서 오버샘플링

을 수행하였고 데이터셋의 균형을 맞추었다.

본 연구에서 적용업체에서는 불량품을 판별해 내는 것이 중요하다고 판단하고 있으므로 양품/불량품과 같이 2-클래스 분류모형에서는 불량품을 정확하게 불량품으로 판별하는 평가지표인 특이도(Specificity)를 평가지표로 채택하였다. <Table 6>에서는 2-클래스 분류모형별 평가지표 결과를 보여주고 있다. 특이도 측면에서는 SVM 모형이 가장 좋은 결과를 보여주고 있으므로 이 모형을 양불 판단 모형으로 채택하였다.

<Table 6> 2-Class Classification Result

Model	Accuracy	Specificity	Recall	Precision	F1-score
Logistic Regression	0.9581	0.8387	0.9642	0.9908	0.9773
AdaBoost	0.9287	0.8710	0.9374	0.9858	0.9609
SVM	0.9665	0.9355	0.9888	0.9758	0.9822
Random Forest	0.9919	0.8065	0.9709	0.9709	0.9709

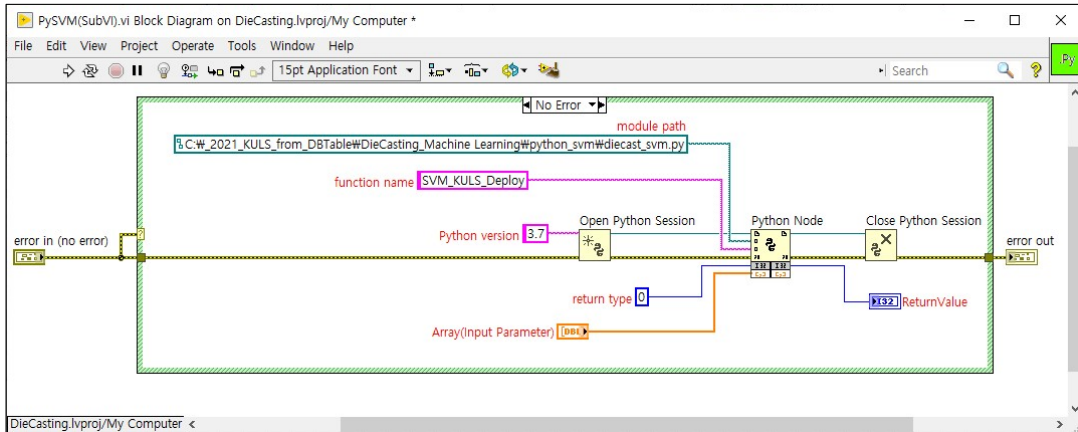
한편 불량타입별 머신러닝 분석에서는 불량타입별로 정확하게 분류하는 것이 중요하므로 분류정확도(Accuracy)를 평가지표로 채택하였다. <Table 7>은 각 모형별 분류정확도 평가지표 결과를 보여주고 있다. 따라서 분류정확도 측면에서는 Logistic Regression, SVM, AdaBoost 등의 알고리즘이 정확하게 불량타입을 예측하고 있음을 확인할 수 있고 이들 중 하나를 불량타입별 예측모형으로 채택할 수 있다.

#### 4. 모니터링 시스템 구축

본 연구에서는 앞에서 설명한 것과 같이 주조설비의 컨트롤러를 통해 확보한 주조 파라미터 데이터와 확보된 불량률 지표를 이용해서 라벨링 작업을 수행하였고, 이 결과를 이용해서 머신러닝 모형(Logistic Regression, Adaboost, RandomForest, SVM 등)을 개발하여 현장에 적용하여 실시간 모니터링을 수행할 수 있는 LabView와

<Table 7> Classification Result by Defect Types

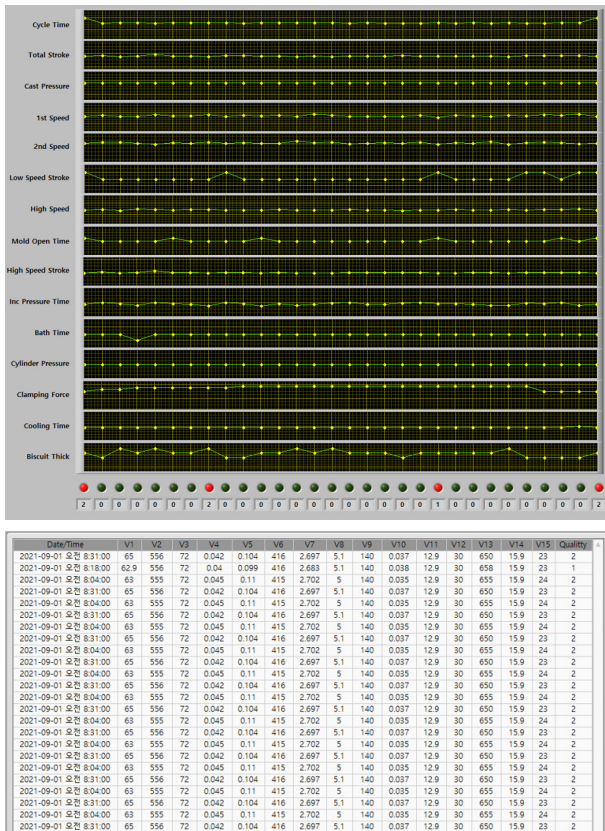
Model	Accuracy	Confusion Matrix					
		Actual					
		General	LUB	PR	TCC		
Logistic Regression	1.0000	predict	General	51	-	-	-
			LUB	-	37	-	-
			PR	-	-	26	-
			TCC	-	-	-	9
		Total	51	37	26	9	
SVM	1.0000	Predict	General	51	-	-	-
			LUB	-	37	-	-
			PR	-	-	26	-
			TCC	-	-	-	9
		Total	51	37	26	9	
Ada Boost	1.0000	Predict	General	51	-	-	-
			LUB	-	37	-	-
			PR	-	-	26	-
			TCC	-	-	-	9
		Total	51	37	26	9	
Random Forest	0.9919	Predict	General	51	1	-	-
			LUB	-	36	-	-
			PR	-	-	26	-
			TCC	-	-	-	9
		Total	51	37	26	9	



<Figure 5> SVM Execution Example on LabView

Python 기반의 시스템을 구현하였다. 즉, 실시간 불량예측 및 불량타입 모니터링 시스템의 구현을 위해서 LabView 환경에서 실시간으로 MES 데이터베이스에 접근하여 Python 언어로 구현한 머신러닝 모듈을 실행할 수 있도록 통합하였다. 즉, 구조조건에 대한 파라미터 데이터는 구조 설비 컨트롤러를 통해 수집되어 실시간으로 MES 데이터베이스에 저장된다. 구축된 데이터베이스를 LabView를

이용하여 실시간으로 접근한 후 Python 불량예측 모듈을 호출하여 진단하는 시스템을 구현하였다. 불량예측 머신러닝 시스템을 현장에 배치하고 실시간 모니터링되는 구조 공정 파라미터를 이용해서 불량예측을 수행하며 불량에 예측될 시 모니터링 화면에 표시 및 관리자에게 통지 등 알람 시스템 동작을 통해 예측결과를 활용할 수 있도록 하였다. <Figure 5>는 LabView 환경의 Python 노드에서 Python으로 구현한 SVM 머신러닝 모듈을 호출하는 코드 일부를 보여주고 있다.



<Figure 6> Implementation of Monitoring System

<Figure 6>은 LabView 환경에서 실시간 불량예측 모니터링 시스템의 실행결과를 보여주고 있다. 수집되는 공정 데이터에 대해 불량원인별 불량량의 발생 여부를 판단하는 화면이며 각 Shot에 대한 15개의 구조공정 파라미터로 구성된 데이터의 트렌드를 보여주고 있으며, 그림 중앙에는 각 Shot에 대한 불량 여부 판단결과를 보여주고 있다. 불량타입 0은 해당 Shot이 양품으로, 불량타입 1, 2, 3, 4는 해당 Shot이 각 불량타입으로 예측되었음을 의미한다. 그림 하단에서는 불량이 예측되는 Shot에 대한 구조공정 파라미터 정보와 불량타입 정보를 보여주고 있다.

### 5. 결론 및 추후연구

본 연구에서는 다이캐스팅 구조공정 설비에서 실시간으로 수집되는 구조조건 파라미터 데이터를 기반으로 머신러닝 기술을 이용하여 구조 시점에서 불량타입별 불량 발생을 실시간으로 예측할 수 있는 접근법을 제시하였다. 현재 구조공정에서 생산되는 제품은 열처리, 사상 등 후공정 작업에서 불량여부가 판별되기 때문에 구조 시점에서 미리 조치하지 못해 대량 불량 발생을 초래할 수 있다.



불량 타입별 불량 예측을 수행하기 위해서는 수집한 데이터에 대한 라벨링 작업이 선행되어야 한다. 본 연구에서는 첫째, 양불 라벨링 작업을 위해서 후가공 업체에서 식별되는 전체 불량률 정보를 Isolation Forest 이상치 탐지 모형의 하이퍼파라미터로 이용하여 양불 라벨링 작업을 수행하였다. 둘째, 분리된 불량 데이터셋을 대상으로 불량 타입 개수를 K-Means 군집분석 모형의 하이퍼파라미터로 이용하여 불량 타입별로 군집을 나눈후에 불량 타입 라벨링 작업을 수행하였다.

위와 같이 라벨링 된 데이터셋에는 데이터 불균형이 존재하기 때문에 SMOTE 기법을 이용하여 데이터 균형을 맞춘후 불량 발생 및 불량 타입별 불량 예측 학습모형을 개발하였다. 불량발생 예측은 불량을 탐지하는 것이 중요하므로 특이도(Specificity) 지표가 가장 우수한 SVM 모형을 선정하였고, 불량 타입별 발생 탐지를 위한 학습모형은 분류정확도(Accuracy) 지표가 우수한 AdaBoost 모형을 선정하였다. 본 연구에서는 개발한 분류모형을 현장에 배치하여 실시간으로 수집되는 주조공정 파라미터를 이용해서 모니터링 작업을 수행할 수 있도록 LabView와 Python을 이용해서 프로토타입 시스템을 제시하였다. 본 연구결과는 인력확보의 어려움 등을 겪고 있는 국내 뿌리산업 제조업체의 제조 지능화에 기여할 수 있을 것으로 판단된다.

## Acknowledgement

이 논문은 2021~2022년도 창원대학교 자율연구과제 연구비 지원으로 수행된 연구결과임

## References

- [1] Abe, N., Zadrozny, B., and Langford, J., Outlier detection by active learning, In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ACM Press, New York, NY, USA, 2006, pp. 504-509.
- [2] Aggarwal, C.C., A Human-Computer Interactive Method for Projected Clustering, *IEEE Transactions on Knowledge and Data Engineering*, 2004, Vol.16. No.4, pp. 448-460.
- [3] Bouveyron, C., Brunet-Saumard, Camille, Model-based Clustering of High-dimensional Data: A Review, *Computational Statistics & Data Analysis*, 2014, Vol. 71, pp. 52-78.
- [4] Breunig, M.M., Kriegel, Hans-Peter, Ng, R.T., and Sander, J., LOF: Identifying Density-Based Local Outliers, *ACM SIGMOD Record*, 2000, Vol. 29, Issue 2, pp. 93-104.
- [5] Chandola, V., Banerjee, A, Kumar, A., Anomaly detection: A survey, *ACM Computing Surveys*, 2009, Vol. 41, Issue 3, pp. 1-58.
- [6] Chawla, N.V., Japkowicz, N., and Kotcz, A., Editorial: special issue on learning from imbalanced data sets, *SIGKDD Explorations* 6, 2004, 1, pp. 1-6.
- [7] Choi, E. S., Kim, J.H., Aziz, N., Lee, S.H., Kang, J.T., and Yoo, K.H., Detection of the Defected Regions in Manufacturing Process Data using DBSCAN, *The Journal of the Korea Contents Association*, 2017, Vol. 17, No. 7, pp. 182-192.
- [8] Choi, S. and Lee, D., Real-Time Prediction for Product Surface Roughness by Support Vector Regression, *Journal of Society of Korea Industrial and Systems Engineering*, 2021, Vol. 44, No. 3, pp. 117-124
- [9] Choo, Y.-S. and Shin, S.-J., Cost-optimal Preventive Maintenance based on Remaining Useful Life Prediction and Minimum-repair Block Replacement Models, *Journal of Society of Korea Industrial and Systems Engineering*, 2022, Vol. 45, No. 3, pp.18-30
- [10] Choi, N.-H., Oh, J.-S., Ahn, J.-R., Kim, K.-S., A Development of Defect Prediction Model using Machine Learning in Polyurethane Foaming Process for Automotive Seat, *Journal of the Korea Academia-Industrial cooperation Society*, 2021, Vol. 22, No. 6, pp. 36-42
- [11] Ezugwu, A.E., Ikotun, A.M., Oyelade, O.O., Abualigah, L., Agushaka, J.O., Eke, C.I., and Akinyelu, A.A., A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects, *Engineering Applications of Artificial Intelligence*, 2022, Vol. 10, Article 104743.
- [12] Ezugwu, A.E., Shukla, A.K., and Agbaje, M.B., Automatic clustering algorithms: A systematic review and bibliometric analysis of relevant literature, *Neural Computing and Applications*, 2021, Vol. 33, pp. 6247-6306.
- [13] Jain, A.K., Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, 2010, Vol. 31, Issue 8, pp. 651-666.
- [14] Joshi, M.V., Agarwal, R.C., and Kumar, V., Mining needle in a haystack: Classifying rare classes via two-phase rule induction, In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, ACM Press, New York, NY, USA, pp. 91-102.
- [15] Kim, J., Kang, H.S., and Lee, J.Y., Development of

- Intelligence Data Analytics System for Quality Enhancement of Die-Casting Process, *Journal of the Korean Society for Precision Engineering*, 2020, Vol 37, No 4, pp. 247-254.
- [16] Kittur, J.K., Manjunath, P.G.C., and Parappagoudar, M.B., Modeling of Pressure die casting process: An Artificial Intelligence Approach, *International Journal of Metalcasting*, 2015, Vol. 10, Issue 1, pp. 70-87.
- [17] Kwon, Sehyug, Anomaly Detection of Big Time Series Data Using Machine Learning, *Journal of Society of Korea Industrial and Systems Engineering*, 2020, Vol. 43, No. 2, pp. 33-38.
- [18] Lee, Jong-Yeong, Choi, Myoung Jin, Joo, Yeongin, Yang, Jaekyung, Ensemble Method for Predicting Particulate Matter and Odor Intensity, *Journal of Society of Korea Industrial and Systems Engineering*, 2019, Vol. 42, No. 4, pp. 203-210.
- [19] Lee, J.H., Noh, S.D., Kim, H.J., and Kang, Y.S., Implementation of Cyber-Physical Production Systems for Quality Prediction and Operation Control in Metal Casting, *Sensors*, 2018, Vol. 18, No. 5, 1428.
- [20] Lee, J. and Lee, Y.C., Die-casting fault detection based on unsupervised deep-learning, *Proceeding of KSME Annual Meeting*, 2021, pp. 1027-1029.
- [21] Lee, S., Lee, S.C., Han, D.S., and Kim, N.S., Study on the Process Management for Casting Defects Detection in High Pressure Die Casting based on Machine Learning Algorithm, *Journal of Korea Foundry Society*, Vol. 41, No. 6, pp. 521-527.
- [22] Lee, J.S., Lee, Y.C., and Kim, J.T., Migration from the traditional to the smart factory in the die-casting industry: Novel process data acquisition and fault detection based on artificial neural network, *Journal of Material Processing Technology*, 2021, Vol. 290, 1735.
- [23] Mac Queen, J.E., Some methods for classification and analysis of multivariate observations, *Proceedings of the Fifth Berkley Symposium Math. Stat Prob*, 1967, Vol.1, pp. 281-297.
- [24] Mutar, Jinan Redha, A Review of Clustering Algorithms, *International Journal of Computer Science and Mobile Applications*, 2022, Vol.10, Issue. 10, pp. 44-50.
- [25] Park, C.S. and Bae, S.M., A Study on the Predictive Maintenance of 5 Axis CNC Machine Tools for Cutting of Large Aircraft Parts, *Journal of Society of Korea Industrial and Systems Engineering*, 2020, Vol. 43, No. 4, pp. 161-167.
- [26] Phua, C., Alahakoon, D., and Lee, V., Minority report in fraud detection: classification of skewed data, *SIGKDD Explorer Newsletter* 6, 2004, 1, pp. 50-59.
- [27] Seo, M.-K. and Yun, W.Y., Clustering-based Monitoring and Fault detection in Hot Strip Roughing Mill, *Journal of Korean Society for Quality Management*, 2017, Vol. 45, No.1, pp. 25-38.
- [28] Theiler, J. and Cai, D.M., Resampling approach for anomaly detection in multispectral images, In *Proceedings of SPIE 5093*, 2003, pp. 230-240.
- [29] Steinwart, I., Hush, D., and Scovel, C., A classification framework for anomaly detection, *Journal of Machine Learning Research*, 2005, Vol. 6, pp. 211-232.
- [30] Triantafillakis, A., Panagiotis Kanellis, Drakoulis Martakos, Data warehouse clustering on the web, *European Journal of Operational Research*, 2005, Vol. 160, No. 2, pp. 353-364.
- [31] Ur Rehman, A. and Belhaouari, S.B., Unsupervised outlier detection in multidimensional data, *Journal of Big Data*, 2021, Vol. 8, p.80
- [32] Vilalta, R. and Ma, S., Predicting rare events in temporal domains, In *Proceedings of the 2002 IEEE International Conference on Data Mining*, 2002, IEEE Computer Society, Washington, DC, USA, 474.

#### ORCID

Chul SoonPark | <http://orcid.org/0000-0003-3147-7613>

Heungseob Kim | <http://orcid.org/0000-0003-0090-5670>