

머신러닝을 이용한 CNC 가공 불량 발생 예측 모델

한용희

송실대학교 벤처중소기업학과 부교수

Prediction Model of CNC Processing Defects Using Machine Learning

Yong Hee Han

Associate Professor, Department of Entrepreneurship and Small Business, Soongsil University

요약 본 연구는 최근 가공 불량 예측 방법으로 주목받고 있는 머신러닝 기반의 모델을 이용하여 CNC 가공 불량 발생의 실시간 예측을 위한 분석 프레임워크를 제안하고, 해당 프레임워크에 기반하여 XGBoost, CatBoost, LightGBM, 랜덤 포레스트, Extra Trees, SVM, k-최근접 이웃, 로지스틱 회귀 모델을 CNC 설비에 기본 내장된 센서들로부터 추출된 데이터에 적용 및 분석하였다. 분석 결과 XGBoost, CatBoost, LightGBM 모델이 동일하게 가장 우수한 정확도, 정밀도, 재현율, F1 점수, AUC 값을 보였으며, 이 중 LightGBM 모델이 소요 실행 시간이 가장 짧은 것으로 나타났다. 이러한 짧은 소요 실행 시간은 실 시스템 구축 비용 절감, 빠른 불량 예측에 따른 CNC 장비 파손 확률 감소, 전체적인 CNC 활용률 증가 등의 실무적 장점을 가지므로 LightGBM 모델이 기본 센서들만 설치된 CNC 설비에 적용 시 가공 불량 예측에 가장 효과적으로 판단된다. 또한 소요 실행 시간 및 컴퓨팅 파워의 제약이 없는 상황에서는 LightGBM, Extra Trees, k-최근접 이웃, 로지스틱 회귀 모형으로 구성된 앙상블 모델을 적용할 경우 분류 성능이 최대화됨을 확인하였다.

주제어 : 머신러닝, 인공지능, CNC, 불량, 밀링 가공

Abstract This study proposed an analysis framework for real-time prediction of CNC processing defects using machine learning-based models that are recently attracting attention as processing defect prediction methods, and applied it to CNC machines. Analysis shows that the XGBoost, CatBoost, and LightGBM models have the same best accuracy, precision, recall, F1 score, and AUC, of which the LightGBM model took the shortest execution time. This short run time has practical advantages such as reducing actual system deployment costs, reducing the probability of CNC machine damage due to rapid prediction of defects, and increasing overall CNC machine utilization, confirming that the LightGBM model is the most effective machine learning model for CNC machines with only basic sensors installed. In addition, it was confirmed that classification performance was maximized when an ensemble model consisting of LightGBM, ExtraTrees, k-Nearest Neighbors, and logistic regression models was applied in situations where there are no restrictions on execution time and computing power.

Key Words : Machine Learning, Artificial Intelligence, CNC, Defect, Milling

*Corresponding Author : Yong Hee Han(amade@ssu.ac.kr)

1. 서론

선반, 밀링 머신 등의 기존의 범용 설비에서는 수동 핸들에 의해 공구의 조작이 수행되는 반면 CNC(computerized numerical control) 가공은 컴퓨터로 제어되는 고속 회전 공구를 사용하여 대상을 자동으로 절삭 성형하는 방식으로[1], 재질과 형상의 자유도가 높은 성형 가공이 가능하고 가공 품질 및 정밀도가 매우 높은 장점 때문에 다양한 산업 분야(항공기, 선박, 자동차 등)에서 사용된다[2].

CNC 밀링 가공에서 재료의 가공으로 인한 마찰 또는 절삭력의 순간 변경으로 인한 파손으로 인해 공구의 마모 한계치에 도달하면 가공 정밀도의 급격한 감소로 인해 불량 발생하며 절삭력 증가로 인해 공작 기계의 수명도 단축되므로 생산 제품의 품질 균일화 및 재현성 향상을 위해 공구 수명에 따른 가공 불량 예측은 반드시 필요하다[2]. 따라서 절삭 가공 중 공정 신호 모니터링 결과를 분석해 공구의 마모량을 예측하고 이를 기반으로 공구 교체 시기를 최적화함으로써 생산성을 향상시키는 공구 마모 모니터링 기술은 설비 지능화의 주요 기술 중 하나이다[3].

최근 센서, 정보통신, 인공지능 기술의 발전에 따라 절삭 가공 관련 다양한 공정 신호 측정 및 분석을 통한 실시간 공구 마모량 예측 기술이 개발되고 있으며, 절삭 가공 시 공구 상태를 모니터링하는 방법은 직접 및 간접 측정 방법으로 구분된다[4]. 직접 측정 방법은 광학적 측정, 레이저 측정, 전기 저항 변화, 마멸면 화상 인식 등을 통해 공구 마모량을 측정하는 방법으로[2], 측정 정확도가 높은 반면 사용되는 센서 및 측정 시스템의 높은 가격[5] 때문에 일부 고급 CNC 설비에만 설치되어 있는 관계로 높은 추가 비용이 발생하며, 가공 중 마모량의 실시간 측정이 불가능하다는 단점이 존재한다. 간접 측정 방법은 실제 가공 중 발생하는 진동, 절삭력, 주축 부하, 동력, 음향, 절삭 온도, 회전력, 가공물의 표면 거칠기 등의 공정 신호 측정 및 분석을 통해 공구 마모량을 추정하는 방법으로, 직접 측정 방법 대비 정확도가 낮은 반면 모니터링 시스템의 구성이 직접 측정 방법 대비 저렴하며 간단하지만[6,7], 제조 현장에서 일반적으로 사용되는 CNC 설비의 경우 기본적으로 내장된 센서들은 위치, 속도, 가속도, 전류, 전압, 전력 센서 정도이므로 간접 측정 방법을 적용하기 위한 추가적인 센서 설치가 필요한 경우가 대부분이다. 결과적으로 CNC 설비가 설치된 대부분의 생산 현장에서는 CNC 가공 중 품질 불량

발생 여부를 판단하기 어려운 관계로 가공 완료 후 오프라인 품질 검사를 시행하여 불량 발생 여부를 확인하므로 이에 따른 큰 손실 및 비용이 발생하는 상황이다.

따라서 본 연구는 다양한 머신러닝 기반 가공 불량 예측 알고리즘들을 일반적으로 CNC 설비에 기본 내장된 센서들로부터 추출된 데이터에 적용 및 분석하고, 그 결과에 기반하여 기본 센서들만 설치된 CNC 설비에 적용 시 가공 불량 예측에 가장 효과적인 머신러닝 모델을 제안한다.

최근 제조 분야에서 CNC 설비에 관한 연구가 활발히 진행 중이며[8], 제조 공정 지능화를 위해 센서를 활용한 공정 모니터링 데이터의 분석 및 기계학습 기법 적용이 확대되고 있다. 절삭 가공의 경우 공정 모니터링 데이터를 활용하여 실시간 이상 상태 진단 및 공구 수명 예측, 품질 예측 등의 지능화 기술 개발이 활발히 진행되고 있다. 특히 설비에 센서 및 CNC 정보를 실시간 모니터링하고 데이터를 수집, 분석하는 스마트 기능의 탑재 수요가 증가하고 있으며, 최근 4차 산업으로 산업 구조가 전환되는 추세에 따라 기존의 고정밀, 고능률화 위주가 아닌 ICT 스마트 가공 및 스마트 자동화 관련 다양한 연구가 진행되고 있다[2].

CNC 설비 모니터링 연구는 분석 수준에 따라 설비 운영 시스템 분석 방법론과 공정 데이터 분석 방법론으로 분류할 수 있다[1]. 설비 운영 시스템 분석 방법론은 고도화된 S/W를 통한 설비의 효율적 제어 관련 연구로 대표 사례로 개방형 플랫폼 연구가 있다[8]. [1]은 스마트폰 금속 프레임의 CNC 설비 기반 가공 공정에 소요되는 비용 최소화를 위해 공구 상태를 파악하였으며, 다변량 모니터링 방안을 통계 기법들에 기반하여 개발하여 중단 공정과 지연 공정의 이상 원인을 확인하였고, 중단 공정은 주로 Z축 모터, 지연 공정은 주로 스핀들 모터가 원인임을 확인하였다. 공정 데이터 분석 방법론은 설비의 센서 혹은 컨트롤러를 통해 수집된 데이터 분석을 통한 공정 상태 파악 관련 연구로 가공품의 표면 거칠기, 가공 공구 수명 등을 파악하는 연구가 대표적으로, 대부분의 연구는 CNC 절삭 가공 시 생산되는 제품의 품질 균일화 및 재현성 향상을 위한 다양한 공구 수명 예측 관련 연구에 집중되었다.

또한 CNC 절삭 공구 관련 연구는 국내와 해외 연구로 구분할 수 있으며, 해외 연구의 경우 [9]는 CNC 설비 모니터링 관련 논문 29건을 다양한 기준으로 분류한

결과 84%가 인공신경망 혹은 퍼지 논리 기반의 알고리즘에 기반하여 독립 및 종속 변수 간 관계를 파악하는 것으로 나타났다. 최근에는 기계학습을 공구 모니터링 과정에 적용하는 연구가 진행되고 있으며[3], 다양한 머신러닝 기법들을 주파수, 공구 이동 속도, 전력, 진동 등의 데이터에 적용한 연구가 다수 존재한다. [10]는 스피너들의 진류 변화를 측정하여 기계학습 기법의 하나인 인공신경망 적용을 통해 공구 마모 정도를 예측하였으며, [11,12]는 음향 방출 센서를 이용하여 신호를 수신하여 공구 상태를 모니터링하는 연구를 수행하였고, [13]는 CNC 설비 내 센서에서 수집된 데이터를 AI 기법을 이용하여 분석하는 다양한 논문들을 정리하였다.

반면 국내 연구는 CNC 설비에 온도, 소음, 음향, 진동 센서를 설치하여 진동, 소음 등의 설비 관련 데이터를 수집하는 연구가 대부분으로, 이러한 연구들은 수집된 데이터를 일반적으로 회귀분석과 실험계획법을 사용하여 분석하였다[8]. [14]는 공구에 센서를 설치하여 데이터를 수집하여 건식 제약조건에서의 절삭력 예측 관련 연구를 수행하였으며, 실험 결과는 편차가 크게 나타났다. [3]은 절삭력 측정 결과와 자가연상신경망을 공구 마모량 대신 이용하여 공구 마모도를 예측하는 모델을 제안하고 그 성능을 검증하였지만, 국내 연구는 센서 등을 이용하여 데이터를 수집 및 분석하는 방법에 초점을 맞추는 연구가 대부분으로, 다양한 특성을 사용한 연구(특히 머신러닝 방법론을 적용한 연구)가 부족한 상황이다[8].

앞에서 정리한 선행 연구의 절대다수는 공구 마모 예측에 집중하였고, CNC 가공품 불량률의 주된 원인이 공구 마모이지만 일부 불량은 설비 설정 오류 등의 기타 원인에 의해 발생한다. 또한 어느 정도 공구 마모가 진행된 후 CNC 가공품 불량률이 발생하는지 제조 현장에서 기존 불량품 발생 데이터 등을 분석하여 파악해야 하는 현실적 어려움이 존재한다. 따라서 본 연구에서는 다양한 머신러닝 기반 가공 불량 예측 알고리즘들을 일반적으로 CNC 설비에 기본 내장된 센서들로부터 추출된 데이터에 적용 및 분석하고, 그 결과에 기반하여 기본 센서들만 설치된 CNC 설비에 적용 시 가공 불량 예측에 가장 효과적인 머신러닝 모델을 제안한다.

2. 연구 방법 및 분석 결과

머신러닝을 이용한 CNC 가공 불량 발생의 실시간 예측 및 불량 발생에 영향을 미치는 변수 분석이 본 연

구의 목적이며, 이를 위해 본 연구는 Fig. 1과 같은 분석 프레임워크를 제안하고 해당 프레임워크에 기반하여 분석을 수행하였다.

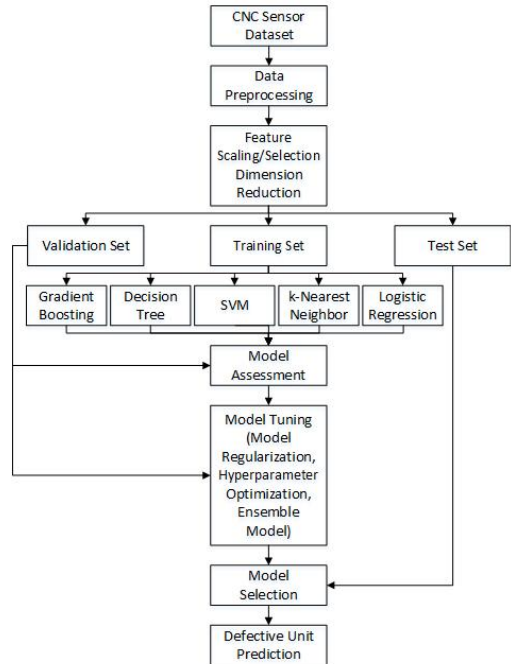


Fig. 1. Proposed Analytical Procedure

데이터 전처리 단계에서는 데이터셋의 기술 통계치 조사 등을 통한 데이터셋 특성 파악, 데이터 표준화 작업 등을 수행하며, 데이터셋 분리 단계에서는 수집된 데이터셋을 선정한 모형들의 불량 예측 능력 검증을 위하여 전체 분석 데이터를 학습용, 검증용, 시험용 데이터로 구분한다. 이후 학습용 데이터를 이용하여 다수의 머신러닝 모델(본 프레임워크에서는 랜덤 포레스트(random forest), Extra Trees, XGBoost, CatBoost, LightGBM, SVM(support vector machine), k-최근접 이웃(k-Nearest Neighbor), 로지스틱 회귀)을 각 모델의 기본값(default value)을 사용하여 훈련시키고 검증용 데이터를 이용하여 각 머신러닝 모델 별 규제(regularization) 및 하이퍼파라미터 최적화를 수행한다. 머신러닝은 기존 샘플링 자료에 대한 통계적인 분석을 통해 획득하기 어려운 새로운 통찰을 획득하는 것을 목적으로 하며, 통계적인 기법에서는 분석 과정에서 연구자의 주관 개입 확률이 높은 반면, 머신러닝 기반 분

석에서는 연구자의 주관 대신 데이터 기반으로 분석이 진행되기 때문에 기존 통계적 기법 대비 연구의 객관성을 확보하기 용이하다.

단일 결정 트리(single decision tree) 모델은 하나의 결정 트리를 이용하여 데이터를 부분 집합으로 분할하고 비용 함수가 최소로 얻어지는 최상의 분할을 검색하여 각 분할에 속해 있는 값들의 불순도가 최소가 되는 방향으로 학습하는 모델로, 원리가 단순하여 실행 시간이 빠르고, 의사결정 과정을 직관적으로 이해할 수 있는 점이 장점인 반면 과적합 발생, 즉 일반 결정 트리의 경우 설명 변수의 수가 증가하면 결정 트리의 가지의 수도 또한 증가하여 학습 데이터 자체에 대한 예측 정확도만 증가하고 일반적인 데이터에 대한 예측 정확도는 감소하는 현상이 발생하기 쉬운 단점이 존재한다. 랜덤 포레스트 모델은 단일 결정 트리의 단점을 보완하기 위해 하나의 결정 트리 생성 시 설명 변수 일부만을 임의로 샘플링하여 사용하며 복원 추출(bagging)을 통해 다수의 결정 트리를 생성하여 그 결과를 종합하는 모델로, 단일 결정 트리 모델 대비 예측 성능이 더 우수한 반면 과적합 발생은 더 적은 양상불 계열의 모델이다. 랜덤 포레스트 모델에서 트리 생성 시 각 노드가 무작위로 특성의 서브셋(subset)을 생성하여 분할에 사용하는 데 비해 Extra Trees 모델은 트리 생성의 무작위성을 위해 최적의 임계값을 검색하지 않고 후보 특성을 사용한 무작위 분할 후 최상의 분할을 선택하므로 랜덤 포레스트 모델 대비 소요 실행 시간이 단축된다. 그래디언트 부스팅(Gradient Boosting) 모델은 다수 모델들을 결합해 단일 모델 대비 더 우수한 예측 성능을 갖는 양상불 계열의 모델로, 다수 결정 트리 모델들의 예측 결과들을 종합해 예측한다는 점에서는 랜덤 포레스트 모델이나 Extra Trees 모델과 동일하지만 이전 결정 트리 모델이 잘못 예측하였던 에러를 다음 결정 트리 모델이 학습하여 점진적으로 에러를 줄여가는(즉, 다음 결정 트리 모델이 이전 결정 트리 모델의 잔여 오차를 대상으로 학습하는) 순차적인 학습을 진행한다는 점이 가장 큰 차이점이다. 다양한 그래디언트 부스팅 모델 중 일반적으로 가장 많이 사용되는 모델은 XGBoost, Catboost, LightGBM 모델이다. XGBoost 모델은 원본 그래디언트 부스팅 모델 대비 예측 성능이 우수하며, 계산 효율성이 높고, 병렬 처리가 가능하여 계산 속도가 빠르며, 자체적으로 과적합 규제 기능을 제공하는 장점을 가진

다. CatBoost 모델은 순서형 TBS(target-based statistics)와 순서형 부스팅(ordered boosting)을 통해 범주형 변수를 효과적으로 처리하며, 하이퍼파라미터 최적화에 민감하지 않고, 시간에 따라 데이터의 분포가 변화하는 경우에도 예측 성능 저하가 낮은 점이 장점인 반면, LightGBM 대비 계산 속도가 느리고, 범주형 특성이 적은 경우 계산 속도나 계산 속도 측면에서 타 그래디언트 부스팅 모델 대비 이점이 없다는 점이 단점이다. LightGBM 모델은 히스토그램 기반 그래디언트 부스팅 모델로, 장점은 타 그래디언트 부스팅 모델 대비 학습 시간이 가장 짧고, 메모리 사용량이 가장 낮은 점이며, 단점은 데이터가 적은 경우 과적합 발생 확률이 상승한다는 점이다. SVM 모델은 학습 데이터를 비선형 매핑(nonlinear mapping)에 의한 고차원 변환을 통해 새로운 차원에서의 최적 분리를 위한 선형 초평면(hyperplane)을 찾는 방식으로 분류를 수행하며, 타 머신러닝 알고리즘 대비 상대적으로 적은 데이터에서도 예측 성능 및 일반화 능력이 우수한 점이 장점인 반면, 특성 스케일링(feature scaling)이라는 데이터 가공 과정이 필요하고, 데이터가 많은 경우 학습 속도가 느린 점이 단점이다. k-최근접 이웃 모델은 현재 샘플(sample)로부터 가장 가까운 거리의 샘플 k개의 레이블(label) 중 가장 많은 수를 차지하는 레이블로 현재 샘플을 분류하는 모델로, 원리가 단순하여 직관적 이해가 가능하고, (횡수, 점수, 거리 등의) 모델의 사전 훈련이 불필요하여 수집된 새로운 훈련 데이터에 즉시 적용 가능한 점이 장점인 반면, 전체 샘플과의 거리를 계산하므로 타 머신러닝 모델 대비 예측 속도가 느리고, 가까운 샘플들만으로도만 계산하므로 편향된 결과가 가능하며, 예측 시 모든 훈련 데이터셋을 보관해야 하므로 데이터가 많은 경우 저장 공간 문제가 발생하며, 데이터 특이치(outlier)에 민감한 점이 단점이다. 로지스틱 회귀 모델은 설명 변수와 여러 범주가 있는 명목형 반응 변수 간의 선형 관계를 추정한다.

모델 선택 단계에서는 시험용 데이터를 이용하여 각 머신러닝 모델의 성능을 평가하며, 분류 문제에서 일반적으로 사용되는 정확도(accuracy), 정밀도(precision), 재현율(recall), F1 점수(F1 score), AUC(area under the curve of ROC(receiver operating characteristic)), 소요 실행 시간을 사용한다. 정확도는 전체 예측 건수 중 정확히 예측한 건수의 비율로, 정밀도는 전체 양성(positi

tive) 예측 건수 중 진짜 양성 건수의 비율로, 재현율은 전체 양성 건수 중 진짜 양성 건수의 비율로, F1 점수는 재현율 vs. 정밀도의 조화 평균 값으로, AUC는 모든 임계값에서의 분류 모델의 성능을 의미하는 ROC 곡선 아래 면적, 소요 실행 시간은 머신러닝 모델의 훈련, 예측, 평가에 소요된 총 시간으로 각각 정의되며, 머신러닝 모델이 예측한 값 (예측값) 및 실제값이 Table 1과 같을 때 정확도, 정밀도, 재현율, F1 점수는 다음과 같이 정의된다.

Table 1. Confusion Matrix

		Predicted Value	
		Negative (-)	Positive (+)
Actual Value	Negative (-)	TN (True Negative)	FP (False Positive)
	Positive (+)	FN (False Negative)	TP (True Positive)

$$\text{정확도} = \frac{TN + TP}{TN + TP + FN + FP}$$

$$\text{정밀도} = \frac{TP}{FP + TP}$$

$$\text{재현율} = \frac{TP}{FN + TP}$$

$$\text{F1 점수} = \frac{1}{\frac{1}{\text{정밀도}} + \frac{1}{\text{재현율}}}$$

본 연구에서 사용된 모든 지표는 소요 실행 시간을 제외하고 모두 [0, 1] 에 분포하며, 높은 값일수록 범주에 대한 분류 성능이 뛰어남을 의미한다[16].

3. 분석 결과

본 연구에서 사용된 CNC 머신 AI 데이터셋[15]은 특정 자동차 부품 제조 업체의 특정 CNC 장비(두산공작기계 MD 6700)에 기본적으로 설치된 48개의 센서로부터 2020년 10월 19일에서 23일까지의 기간 동안 수집된 데이터이며, 주로 가공되는 부품의 x, y, z축, 스피들(spindle) 관련 변수들의 설정값 및 센서를 통해 실제로 측정된 값으로 구성되었다. 여기에 포함되는 변수는 위치, 속도, 가속도, 전류, 전압, 전력(전원), 토크 관성 전력, 수행 중인 가공 단계이다. 또한 토크 관성, 프로그램이 CNC에 나열된 번호, 실행 중인 G-code 라인, 스

핀들의 순간 공급 속도, 현재 수행 중인 가공 단계 데이터도 데이터셋에 포함되었다. 해당 데이터셋을 이용한 예측 대상은 가공이 완료된 제품의 양품/불량 여부 데이터는 작업자의 육안 검사를 통해 수집되었으며, 총 데이터 32,048개 중 양품 데이터는 22,645개, 불량 데이터는 9,403개이다.

데이터 전처리 단계에서는 데이터셋의 기술 통계를 조사한 결과 특성 간 범위의 차이가 크며 이상치(outlier)로 의심되는 샘플들이 일부 존재함을 확인하였으며, 이상치가 포함된 데이터셋에 적합한 사이킷런(Scikit-Learn) 라이브러리의 RobustScaler 클래스를 적용하여 데이터 표준화 작업을 수행하였다. 또한 양품 데이터와 불량 데이터 간 불균형 문제 해결을 위해 SMOTE(synthetic minority oversampling technique)에 의한 불량 데이터 오버샘플링(oversampling)으로 양품:불량 데이터 비율을 1:1로 조정하였다. 또한 본 연구를 위해 수집한 48개 특성 간 관계가 분명치 않아 CNC 가공 불량 예측 모형의 데이터로 모두 사용하는 것에는 비효율적인 측면이 존재한다. 따라서 다음 단계인 차원 축소 단계에서는 주성분 분석(PCA) 기법을 적용하여 전체 특성 분산의 98.3%를 설명하는 1개 주성분을 도출 후 분석에 사용하였다. 다음 단계인 데이터셋 분리 단계에서는 전체 분석 데이터를 60: 20: 20의 비율로 각각 학습용, 검증용, 시험용 데이터로 구분하였다.

Table 2는 머신러닝 모델별 성능 평가 결과이다. 그라디언트 부스팅 계열 앙상블 머신러닝 모델인 XGBoost, CatBoost, LightGBM 모델의 경우 각 모델별 소요 실행 시간 간 차이가 큰 반면 모두 동일하게 가장 우수한 정확도, 정밀도, 재현율, F1 점수, AUC 값을 보였으며, 결정 나무 계열 앙상블 모델인 랜덤 포레스트와 Extra Tree 모델의 경우 근소한 차이(0.01)로 낮은 지표값을 보였다. 그 외의 모델(SVM, k-최근접 이웃, 로지스틱 회귀 모델)의 경우 타 모델 대비 낮은 평가 척도 값을 보였다.

또한 본 연구에서는 직접 투표(hard voting) 방식에 기반하여 개별 모델의 예측 결과에 균등하게 가중치를 부여하여 최종 예측을 수행하는 앙상블 모델을 제안하고 해당 모델의 성능을 측정하였다. 일반적으로 구성 모델이 상이할수록 전반적인 분류 성능이 향상되는 앙상블 모델의 특성 상 앙상블 모델에 유사한 모델들이 과다하게 포함되는 것은 바람직스럽지 않다. 따라서 그라디언트 부스팅 계열 및 결정 나무 계열 모델 중 소요 실행

시간이 가장 적은 LightGBM 및 Extra Trees 모델이 이상불 모델에 포함되었고, 추가적으로 위의 두 모델과 작동 원리가 상이한 k-최근접 이웃, 로지스틱 회귀 모델이 포함되었다 (SVM 모델은 소요 실행 시간 문제로 미포함됨). 해당 이상불 모델의 성능 평가 결과 분류 성능은 약간 상승하였으나 소요 실행 시간이 크게 증가한 것으로 나타났다.

Table 2. Evaluation Results

	Accuracy	Precision	Recall	F1 Score	AUC	time (sec)
XGBoost	0.97	0.97	0.97	0.97	0.96	2.24
CatBoost	0.97	0.97	0.97	0.97	0.96	12.98
LightGBM	0.97	0.97	0.97	0.97	0.96	0.71
Random Forest	0.96	0.96	0.96	0.96	0.96	3.85
Extra Trees	0.96	0.96	0.96	0.96	0.96	1.90
SVM	0.80	0.79	0.80	0.79	0.72	47.34
k-Nearest Neighbor	0.87	0.87	0.87	0.86	0.80	3.93
Logistic Regression	0.73	0.72	0.73	0.72	0.64	0.42
Ensemble	0.98	0.98	0.98	0.98	0.97	7.91

또한 불량에 영향을 미치는 특성 변수들의 상대적 중요도를 LightGBM 모델로 분석한 결과, Table 3과 같이 스피들의 실제 위치, 스피들의 순간 공급 속도, x축의 출력 전압, 스피들의 실제 속도, z축의 실제 위치 순으로 불량 예측에 영향을 미치는 것을 확인하였다.

Table 3. Feature Importances

Feature Name	Importance
Actual Position of Spindle	428
Current Feedrate of Spindle	216
Output Current of x Axis	163
Actual Velocity of Spindle	138
Actual Position of z Axis	117
Output Current of Y Axis	115
Output Current of Spindle	108
Set Acceleration of Spindle	102
Actual Position of x Axis	98
Actual Position of y Axis	98

4. 결론

본 연구는 최근 가공 불량 예측 방법으로 주목받고 있는 머신러닝 기반의 모델 중 XGBoost, CatBoost, LightGBM, 랜덤 포레스트, Extra Trees, SVM, k-최근접 이웃, 로지스틱 회귀 모델을 적용하여 CNC 가공 완제품의 불량 여부를 예측하였다. 머신러닝 모델 간 정확도, 정밀도, 재현율, F1 점수, AUC 값, 소요 실행 시간을 비교한 결과 XGBoost, CatBoost, LightGBM 모델이 동일하게 가장 우수한 정확도(0.97), 정밀도(0.97), 재현율(0.97), F1 점수(0.97), AUC(0.96) 값을 보였으며, 이 중 LightGBM 모델이 소요 실행 시간이 0.71초로 가장 짧은 것으로 나타났다. 이러한 짧은 소요 실행 시간은 실 시스템 구축 비용 절감, 빠른 불량 예측에 따른 CNC 장비 파손 확률 감소, 전체적인 CNC 활용률 증가 등의 실무적 장점을 가지므로 LightGBM 모델이 CNC 가공 완제품의 불량 여부 예측에 가장 효과적으로 판단된다. 또한 소요 실행 시간 및 컴퓨팅 파워의 제약이 없는 상황에서는 LightGBM, Extra Trees, k-최근접 이웃, 로지스틱 회귀 모형으로 구성된 이상불 모델을 적용할 경우 분류 성능이 최대화됨을 확인하였다.

하지만 본 연구에서는 데이터 확보의 한계 때문에 특정 CNC 모델에 설치된 센서들의 제한된 기간(5일) 내 데이터만을 분석에 이용하였기 때문에 연구 결과의 일반화에 한계가 존재하며, 향후 연구에서는 다양한 CNC 모델 및 센서로부터 장기간 수집한 데이터를 이용할 경우 보다 일반적이고 정교한 CNC 가공 불량 예측 모델의 개발과 이를 통한 빠른 불량 예측이 가능할 것으로 판단된다.

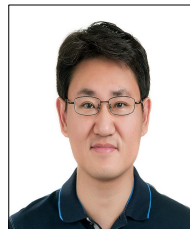
REFERENCES

- [1] S. H. Kang & S. B. Kim. (2016). Multivariate Monitoring of the Metal Frame Process in Mobile Device Manufacturing, *Journal of Korean Institute of Industrial Engineers*, 42(6), 395-403. DOI : 10.7232/JKIIIE.2016.42.6.395
- [2] J. S. Kong. (2018), Optimization of the Tool Life Prediction Using Genetic Algorithm, *Journal of the Korea Academia-Industrial cooperation Society*, 19(11), 338-343. DOI : 10.5762/KAIS.2018.19.11.338
- [3] D. J. Oh, B. S. Sim, & W. Lee. (2021). Tool Wear Monitoring during Milling Using an

- Autoassociative Neural Network, *Transactions of the Korean Society of Mechanical Engineers - A*, 45(4), 285-291.
DOI : 10.3795/KSME-A.2021.45.4.285
- [4] R. Teti, K. Jemielniak, G. O'Donnell, & D. Dornfeld. (2010). Advanced Monitoring of Machining Operations, *CIRP Annals*, 59(2), 717-739.
DOI : 10.1016/j.cirp.2010.05.010
- [5] Y. C. Liu, X. F. Hu, & S. X. Sun. (2019, July). Remaining Useful Life Prediction of Cutting Tools Based on Support Vector Regression, *IOP Conference Series: Materials Science and Engineering*, 576, 1-8.
DOI : 10.1088/1757-899X/576/1/012021
- [6] P. Stavropoulos, A. Papacharalampopoulos, E. Vasiliadis, & G. Chryssolouris. (2016). Tool Wear Predictability Estimation in Milling Based on Multi-sensorial Data, *The International Journal of Advanced Manufacturing Technology*, 82(1-4), 509-521.
DOI : 10.1007/s00170-015-7317-6
- [7] X. Li, A. Djordjevich & P. K. Venunivod. (2000). Current-sensor-based Feed Cutting Force Intelligent Estimation and Tool Wear Condition Monitoring, *IEEE Transactions on Industrial Electronics*, 47(3), 697-702.
DOI : 10.1109/41.847910
- [8] K. Lee, S. Park, S. Sung, & D. Park. (2019). A Study on the Prediction of CNC Tool Wear Using Machine Learning Technique, *Journal of the Korea Convergence Society*, 10(11), 15-21.
DOI : 10.15207/JKCS.2019.10.11.015
- [9] J. V. Abellan-Nebot & F. R. Subirón. (2010). A Review of Machining Monitoring Systems Based on Artificial Intelligence Process Models, *The International Journal of Advanced Manufacturing Technology*, 47(1), 237-257.
DOI : 10.1007/s00170-009-2191-8
- [10] C. Drouillet, J. Karandikar, C. Nath, A. C. Journeaux, M. El Mansori, & T. Kurfess. (2016). Tool Life Predictions in Milling Using Spindle Power with the Neural Network Technique, *Journal of Manufacturing Processes*, 22, 161-168.
DOI : 10.1016/j.jmapro.2016.03.010
- [11] da Silva, R. H. L., M. B. da Silva, & A. Hassui. (2016). A Probabilistic Neural Network Applied in Monitoring Tool Wear in the End Milling Operation via Acoustic Emission and Cutting Power Signals, *Machining Science and Technology*, 20(3), 386-405.
DOI : 10.1080/10910344.2016.1191026
- [12] J. A. Duro, J. A. Padget, C. R. Bowen, H. A. Kim, & A. Nassehi. (2016). Multi-sensor Data Fusion Framework for CNC Machining Monitoring, *Mechanical Systems and Signal Processing*, 66, 505-520.
DOI : 10.1016/j.ymsp.2015.04.019
- [13] A. J. Torabi, M. J. Er, X. Li, B. S. Lim, L. Zhai, R. J. Oentaryo, & J. M. Zurada. (2013). A Survey on Artificial Intelligence-based Modeling Techniques for High Speed Milling Processes, *IEEE Systems Journal*, 9(3), 1069-1080.
DOI : 10.1109/JSYST.2013.2282479
- [14] S. T. Jung, S. H. Kim, H. J. Kim, & S. Y. Baek. (2018). Prediction and Experiments of Cutting Forces in Down Milling of Hardened Mold Steel, *Journal of the Korean Society of Manufacturing Technology Engineers*, 27(4), 346-350.
DOI : 10.7735/ksmte.2018.27.4.346
- [15] Ministry of SMEs and Startups of Korea. (2020). *CNC Machine AI Dataset*. Korea AI Manufacturing Platform (KAMP). <https://kcamp-ai.kr>
- [16] A. Géron. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*, O'Reilly Media.

한 용 희(Yong Hee Han)

[정회원]



- 1995년 2월 : 한양대학교 산업공학과 (공학사)
- 1998년 8월 : Georgia Institute of Technology 산업공학과 (공학석사)
- 2004년 12월 : Georgia Institute of Technology 산업공학과(공학박사)
- 2011년 3월 ~ 현재 : 송실대학교 벤처중소기업학과 교수
- 관심분야 : 머신러닝, 융복합, 기업가정신
- E-Mail : amade@ssu.ac.kr