

서포트벡터머신 기반 PVDF 센서의 결함 예측 기법

김승욱* · 이상민**

Fault Detection Technique for PVDF Sensor Based on Support Vector Machine

Seung-Wook Kim* · Sang-Min Lee**

요약

본 연구에서는 구조물 건전성 모니터링에 널리 활용되고 있는 PVDF(Polyvinylidene fluoride) 센서에 나타날 수 있는 결함을 실시간으로 분류 및 예측하기 위한 방법론을 제안하였다. 센서 부착 환경에 따라 나타나는 센서의 결함 유형을 분류하였고, 임팩트 해머를 이용한 충격 시험을 수행하여 결함 유형에 따른 출력 신호를 획득하였다. 결함 유형에 따른 출력 신호간의 차이를 식별하기 위해 이들의 시간영역 통계 특징을 추출하여 데이터 집합을 구축하였다. 머신러닝 기반 분류 알고리즘들 중 센서 결함 유형 감지에 가장 적합한 알고리즘 선정을 위해 구축한 데이터 집합의 학습 및 이에 따른 결과를 분석하였고, 이들 중 SVM(Support vector machine)이 가장 높은 성능을 보임을 확인하였다. 선정된 SVM 알고리즘의 추가적인 정확도 향상을 위해 하이퍼 파라미터 최적화 작업을 수행하였으며, 결과적으로 92.5%의 정확도로 센서 결함 유형을 분류하였고 이는 타 분류 알고리즘에 비하여 최대 13.95% 높은 정확도를 보였다. 본 연구에서 제안한 센서 결함 예측 기법은 PVDF 센서뿐만 아니라 실시간 구조물 건전성 모니터링을 위한 다양한 센서의 신뢰성을 확보하기 위한 기반 기술로 활용될 수 있을 것으로 사료된다.

ABSTRACT

In this study, a methodology for real-time classification and prediction of defects that may appear in PVDF(Polyvinylidene fluoride) sensors, which are widely used for structural integrity monitoring, is proposed. The types of sensor defects appearing according to the sensor attachment environment were classified, and an impact test using an impact hammer was performed to obtain an output signal according to the defect type. In order to clearly identify the difference between the output signal according to the defect types, the time domain statistical features were extracted and a data set was constructed. Among the machine learning based classification algorithms, the learning of the acquired data set and the result were analyzed to select the most suitable algorithm for detecting sensor defect types, and among them, it was confirmed that the highest optimization was performed to show SVM(Support Vector Machine). As a result, sensor defect types were classified with an accuracy of 92.5%, which was up to 13.95% higher than other classification algorithms. It is believed that the sensor defect prediction technique proposed in this study can be used as a base technology to secure the reliability of not only PVDF sensors but also various sensors for real time structural health monitoring.

키워드

Fault Classification, Fault Detection, Support Vector Machine, PVDF Sensor, Sensor Fault
결함 분류, 결함 감지, 서포트 벡터 머신, PVDF 센서, 센서 결함

* 교신저자 : 한양대학교 융합공학부

** 고려대학교 기계공학부(sangmin727@korea.ac.kr)

• 접수일 : 2023. 08. 14

• 수정완료일 : 2023. 09. 12

• 게재확정일 : 2023. 10. 17

• Received : Aug. 14, 2023, Revised : Sep. 12, 2023, Accepted : Oct. 17, 2023

• Corresponding Author : Seung-Wook Kim

Dept. of Convergence Defense Hanyang University,

Email : seungwook526@naver.com

I. 서 론

PVDF(Polyvinylidene fluoride) 센서는 폴리머로 제작된 필름형 압전센서로 균열 또는 파손과 같이 진동 및 충격에 의해 나타날 수 있는 손상 유형을 진동 신호를 통해 구조물 상태를 실시간으로 감지하는 목적으로 사용된다. 기존 세라믹 기반 압전센서와 비교할 때, PVDF 센서는 신호를 받아들이는 반응성이 뛰어나고 뿐만 아니라 그 부피가 작아 이를 건축구조물, 의공학, 로봇, 자동화 시스템 등 다양한 분야에 적용하기 위한 연구가 활발히 진행되고 있다[1-2].

한편, 실시간 모니터링 장비의 센서에 결함이 존재할 경우 장비의 오작동이 발생할 수 있으며, 이는 막대한 경제적 손실뿐만 아니라 인명피해 또한 유발할 수 있다. 따라서 이러한 실시간 모니터링 장비의 센서가 출력하는 신호에 대한 신뢰성 확보가 필수적이며 센서 자체의 결함을 예측할 수 있는 연구가 필요하다. 이러한 센서의 신뢰성 확보를 위하여 D. Tibaduiza 등은[3] 구조 건전성 모니터링(Structure Health Monitoring, SHM)에 사용되는 센서의 고장에 관한 연구를 수행하였다. SHM 시스템의 오작동을 방지하기 위해 주성분 분석(Principle component analysis, PCA) 지수와 손상 지수 개념을 도입하였고, 이를 기반으로 센서의 결함 유형을 감지하는 방법론을 제시하였다. G. H. Park 등은[4] 센서의 결함 유형을 구분하는 연구를 수행하였다. 센서의 어드미턴스(Admittance)와 임피던스(Impedance)를 기반으로 압전센서의 결함 유형인 debonding, breakage, scratch를 구별하는 방법론을 제시하였다. T. C. Huynh 등은[5] 센서의 건전성을 진단하는 연구를 수행하였다. 압전센서의 결함 유형으로 breakage, debonding, complete failure를 선정하고, 센서 결함으로 인한 임피던스 변화를 분석하여 센서의 건전성을 진단하는 방법론을 제시하였다. 이와 같이 기존 연구사례들은 센서의 임피던스 변화에 기반하여 결함을 예측하였다. 하지만 이러한 연구사례들은 방대한 계산과 분석으로 많은 시간이 소요되고, 구조물의 손상 감지 작업에 유효한 주파수 대역을 결정하는데 매우 큰 어려움이 있어 실시간 모니터링에 적용이 제한된다.

본 연구에서는 많은 시간이 소요된다는 한계점과 구조물 손상 감지 작업에 유효한 주파수 대역 도출의

어려움으로 인해 현장에 도입하기 어렵다는 한계점을 극복하기 위하여 머신러닝 알고리즘에 기반한 센서 결함 유형 추정 방법을 제안하였다.

본 논문은 다음의 순서로 구성된다. II장에서는 센서가 출력하는 신호로부터 얻을 수 있는 특징들에 대한 설명과 머신러닝 기법 종류에 대해 기술하였다. III장은 센서 결함 유형 분류 및 실험 조건에 대해 설명한 후, IV장에서 머신러닝 알고리즘별 센서 결함 유형 분류 결과 비교 및 분석을 진행하였다. 마지막으로 V장에서 연구의 결론을 맺는다.

II. 기반 이론

2.1 시간영역 특징 추출

시간영역 특징 추출 기법은 센서가 출력하는 신호로부터 신호의 특징을 추출하여 목적에 맞게 신호를 분류하는 방법 중 하나이다. 그림 1은 유사한 파형과 출력값을 보이는 두 신호로부터 시간영역 특징인 Max와 RMS를 추출하여 그래프로 도식화한 것이다.

이를 통해 유사한 파형과 출력값을 보여도 신호 특징 추출을 통해 신호를 구분할 수 있음을 나타낸다. 정상 및 센서 결함 유형을 분류하기 위해 유형별 센서 출력 신호로부터 시간영역 통계 특징들을 추출하는 작업을 수행하였다. 이를 통해 유사한 파형과 출력값을 보여도 신호 특징 추출을 통해 신호를 구분할 수 있음을 나타낸다. 따라서 정상 및 센서 결함 유형을 분류하기 위해 각 유형별 센서 출력 신호로부터 시간영역 통계 특징들을 추출하는 작업을 수행하였다.

표 1은 머신러닝 기반 분류 알고리즘에서 주로 사용되며, 신호로부터 시간영역에서 추출할 수 있는 12개의 특징들을 나타낸다[6-7]. 여기서 x_i 는 $i=1, 2, 3, \dots, n$ 일 때의 신호 시계열을 의미하며, n 은 데이터 포인트의 숫자를 의미한다. 표 1에서 Mean은 신호 출력값의 평균, Absolute mean은 신호 출력값의 절댓값 평균, Max는 출력값 중 극댓값, Root mean square는 파형 신호의 실효값, Kurtosis value는 신호에서 이상값이 발생하기 쉬운 정도를 수치화한 값, Skewness value는 파형 신호의 비대칭성을 수치화한 값, Peak to peak value는 파형 신호에서 극댓값과 극소값의 차이값, Shape factor는 신호의 형태와 관련

있는 인자로 실효값을 절댓값 평균으로 나눈 값, impulse factor는 신호 파형의 최대값을 절댓값 평균으로 나눈 값, crest factor는 신호 파형의 최대값을 실효값으로 나눈 비, skewness factor는 skewness value를 실효값의 세제곱으로 나눈값, kurtosis factor는 kurtosis value를 실효값의 세제곱으로 나눈값을 의미한다.

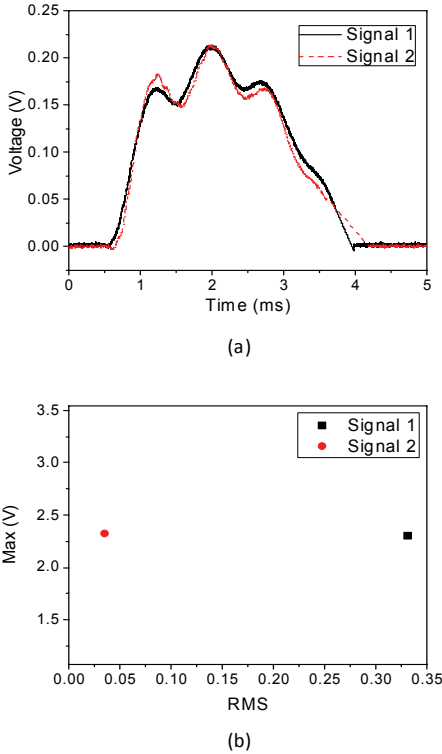


그림 1. 신호 특징을 기반으로 한 유사한 모양의 신호 분류

Fig. 1 Signal classification of similar shapes based on signal features.

본 연구에서는 위와 같은 특징들을 결함 유형별로 모두 추출하여 머신러닝 학습에 활용할 데이터 집합을 구축하였다.

2.2 머신러닝 기반 분류 알고리즘

분류 알고리즘은 머신러닝 중 지도 학습(Supervised Learning)의 일종으로, 기존 입력된 데이터와 클래스(Class)와의 관계를 학습하여 새로 입력된

데이터의 클래스를 판별하는 기법이며, 다량의 데이터를 빠르고 정확하게 분류한다는 장점이 있다. 본 연구에서는 이러한 장점을 가진 머신러닝을 연계하여 총 400회의 실험 데이터로부터 12개의 시간영역 통계 특징들을 추출하는 다량의 계산 작업을 빠르게 수행하고 센서 결함 유형을 분류하였다.

표 1. 시간영역 통계 특징
Table 1. Time-domain features

Feature	Feature Definition
Mean	$T_{mean} = \frac{1}{n} \sum_{i=1}^n x_i$
Absolute mean	$T_{absm} = \frac{1}{n} \sum_{i=1}^n x_i $
Max	$T_{max} = \max(x_i)$
Root mean square	$T_{rms} = [\frac{1}{n} \sum_{i=1}^n x_i^2]^{1/2}$
Kurtosis value	$T_{ku} = \frac{\sum_{i=1}^n (x_i - T_{absm})^4}{(n-1) (\frac{1}{n-1} \sum_{i=1}^n (x_i - T_{absm}))^4}$
Skewness value	$T_{sk} = \frac{\sum_{i=1}^n (x_i - T_{absm})^3}{(n-1) (\frac{1}{n-1} \sum_{i=1}^n (x_i - T_{absm}))^3}$
Peak to peak value	$T_{ppv} = \max(x_i) - \min(x_i)$
Shape factor	$T_{sf} = \frac{T_{rms}}{T_{absm}}$
Impulse factor	$T_{if} = \frac{x_{max}}{T_{absm}}$
Crest factor	$T_{cf} = \frac{T_{max}}{T_{rms}}$
Skewness factor	$T_{skf} = \frac{T_{sk}}{T_{rms}^3}$
Kurtosis factor	$T_{kuf} = \frac{T_{ku}}{T_{rms}^4}$

분류 알고리즘인 결정 트리(Decision Tree), kNN(k-Nearest Neighbours), Naive Bayes, SVM을

사용하여 성능 비교를 통해 센서 결함 유형 감지에 적합한 알고리즘을 찾는 과정을 수행하였다. 사용한 알고리즘에 대한 설명은 다음과 같다.

2.2.1 결정트리

결정 트리는 데이터들 사이에 존재하는 패턴을 학습하여 대상 변수의 값을 예측하는 모델이다[8]. 이는 통계학과 머신러닝에서 사용하는 예측 모델링 중 하나이다. 결정 트리는 자료를 정규화하거나 임의의 변수 및 제거하는 등의 자료를 가공할 필요가 없으며 간략한 설명만으로 결과를 해석하고 이해하기 쉽다는 장점이 있다. 그림 2는 결정 트리의 예를 나타낸다.

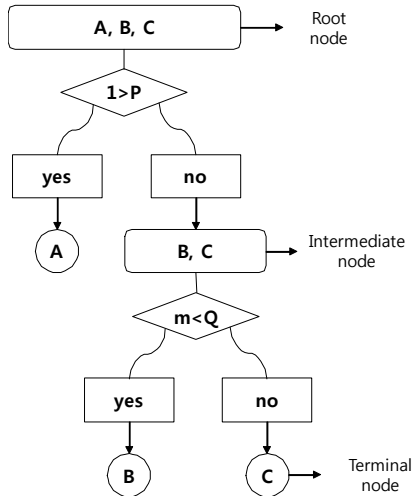


그림 2. 결정 트리 예시
Fig. 2 The example of decision tree

그림 2에서 뿌리 노드(Root node)는 결정 트리에서 가장 위에 존재하는 것을 의미하며, 노드와 노드 사이를 연결하는 것을 가지(Branch)라 한다. 가지 밑으로 더 생성되지 않는 노드를 말단 노드(Terminal node)라 부르며 뿌리 노드와 말단 노드를 제외한 노드들은 전부 중간 노드(Intermediate node)라 부른다.

불순도(Impurity)란 결정 트리를 통해 분류한 범주 안에 서로 다른 데이터가 얼마나 섞여 있는지를 의미하며 그 식은 식 (1)과 같다.

$$Entropy = - \sum_i (p_i) \log_2(p_i) \quad \dots (1)$$

위 식에서 Entropy는 불순도를 수치화한 척도이며, p_i 는 한 영역 안에 존재하는 데이터 중 범주 I에 속하는 데이터 비율을 의미한다.

2.2.2 kNN

kNN이란 새로운 데이터가 주어졌을 때 훈련된 데이터 사이에 가장 근접한 k개의 데이터 정보를 이용하여 새로운 데이터를 분류하는 알고리즘이다[9]. kNN에서 근접하다는 의미는 유클리드 거리(Euclidean Distance)로 정의하며 그 식은 (2)와 같다.

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \dots (2)$$

위 식에서 d는 유클리드 거리를 의미하며 x_1, x_2, y_1, y_2 는 그림 3에서의 좌표를 의미한다.

kNN은 오류 데이터가 결과에 영향을 미치지 않아 높은 정확도를 보인다는 장점이 있으나 데이터양이 증가함에 따라 다량의 데이터 활용을 위한 고용량의 메모리를 사용하며 처리시간이 증가한다는 단점이 있다. 그림 3은 kNN의 예시이다. 그림 3에서 실선은 k값이 3일 때의 분류 결과, 점선은 k값이 5일 때의 결과를 나타낸다. k를 3으로 설정했을 경우 새로운 입력 데이터는 삼각형으로 분류되지만, k를 5로 설정했을 경우 사각형으로 분류된다.

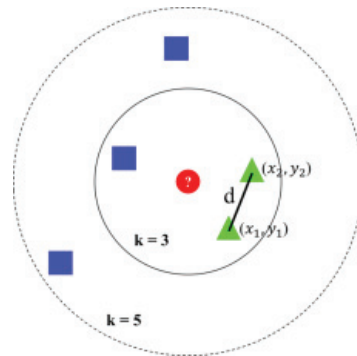


그림 3. kNN 예시
Fig. 3 The example of kNN

2.2.3 Naive Bayes

Naives Bayes는 두 확률 변수의 사전 확률(prior

probability)과 사후 확률(posterior probability) 사이의 관계를 나타낸 Bayes 정리를 기반으로 한 통계적 분류 기법이며 지도 학습 중 한 종류이다[10]. Naive Bayes는 다량의 데이터에 대해 빠른 처리 속도와 높은 정확도를 보인다는 장점이 있다. Naive Bayes 관련 식은 식 (3)과 같다.

$$P(a|b) = \frac{P(b|a)P(a)}{P(b)} \quad \dots (3)$$

식 (3)에서 $P(a|b)$ 는 특정 개체 b가 그룹 a에 포함될 사후 확률을 의미하고 Naive Bayes를 이용하여 구한다. $P(a)$ 는 그룹 a가 발생할 빈도, $P(b)$ 는 그룹 b가 발생할 빈도를 의미한다.

2.2.4 SVM

SVM은 머신러닝 분야 중 하나로 두 개의 클래스를 가진 데이터들이 존재할 때, 서포트 벡터를 이용하여 두 클래스간 거리를 최대화한 초평면을 찾는 이진 분류기법이다[11].

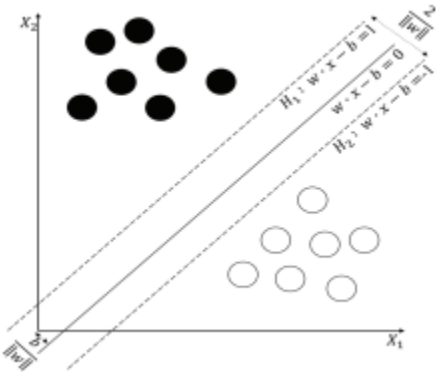


그림 4. SVM 기반 선형 분류
Fig. 4 SVM linear classification

SVM은 초평면으로부터 두 클래스 그룹간의 사이가 클수록 명확한 특징구분이 가능하며 데이터를 판별하는데 높은 신뢰성을 줄 수 있다. 그림 4는 두 개의 클래스를 선형으로 분류하는 예를 보여주며, 최적의 초평면 식은 식 (4)와 같다.

$$w^T x + b = 0 \quad \dots (4)$$

위 식에서 w 는 초평면의 가중치 벡터, x 는 데이터 벡터 그리고 b 는 절편을 의미한다. 서포트 벡터 위에 있는 초평면인 H_1 과 H_2 를 기준으로 입력데이터 x 가 속한 클래스는 각각 식 (5)와 식 (6)과 같다.

$$H_1: w^T x + b \geq 1 \text{ for } y_i = +1 \quad \dots (5)$$

$$H_2: w^T x + b \leq -1 \text{ for } y_i = -1 \quad \dots (6)$$

또한, 두 초평면 사이의 거리는 마진(margin)이라 부르며, 이는 $\left\| \frac{2}{w} \right\|$ 으로 표현된다.

III. 실험 방법

본 연구에서는 많은 시간이 소요된다는 한계점과 구조물 손상 감지 작업에 유효한 주파수 대역 도출의 어려움으로 인해 현장에 도입하기 어렵다는 한계점을 극복하기 위하여 머신러닝 알고리즘에 기반한 센서 결함 유형 추정 방법을 그림 5와 같이 제안하였다.

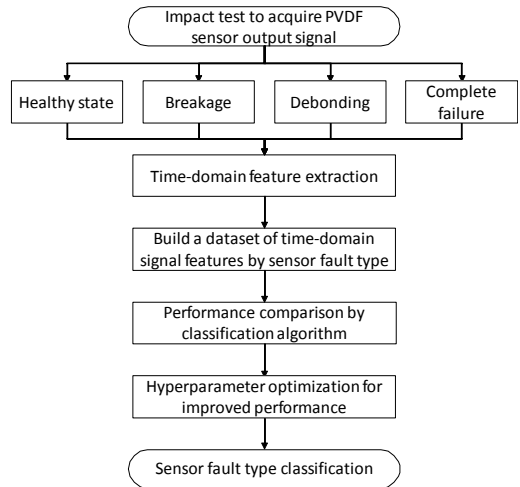


그림 5. 센서 결함 감지 프로세스
Fig. 5 The process of sensor fault detection technique

이를 위해 센서 결함 유형을 분류하였고, 임팩트 해머를 이용한 충격 시험을 수행하여 결함 유형에 따

른 출력 신호를 획득하였다. 획득된 신호의 시간영역 통계 특징을 추출하여 데이터 집합을 구축하였고, 이를 기반으로 총 4 가지의 머신러닝 기반 분류 알고리즘 중 센서 결함 유형 감지에 가장 적합한 알고리즘을 선정하였다. 선정된 알고리즘의 정확도를 추가로 향상시키기 위하여 베이지안 최적화 기법을 활용한 하이퍼 파라미터 최적화(Bayesian Optimization) 작업을 수행하였고, 결론적으로 높은 정확도로 센서 결함 유형을 분류할 수 있는 PVDF 센서 결함 예측기법을 개발하였다.

3.1 센서 결함 유형 분류

압전센서는 사용 목적에 따라 주변 환경에 노출되는 경우가 발생하며, 이 때 외부로부터의 충격 또는 작동 부주의 등 주변 환경에 의해 센서에 결함이 발생한다[12]. 그림 6은 정상 센서와 주된 센서 결함 유형인 debonding 유형, breakage 유형 그리고 complete failure 유형을 나타낸다[5]. debonding 유형은 센서와 부착부 사이에 발생한 접촉 불량 상태, breakage 유형은 센서의 일부가 절단된 상태 그리고 complete failure 유형은 센서가 완전히 손상되어 신호를 출력하지 못하는 상태를 의미한다.

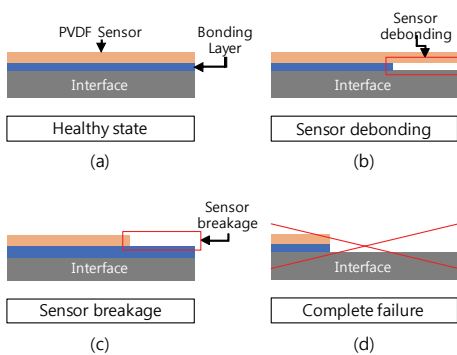


그림 6. 정상센서 및 센서 결함 유형 설명 : (a) 센서 정상 상태, (b) 센서 접촉 불량, (c) 센서 파손 유형, (d) 센서 고장 유형

Fig. 6 Illustration of healthy sensor and sensor fault type : (a) sensor healthy state, (b) sensor debonding case, (c) sensor breakage case, (d) sensor complete failure case

그림 7은 본 연구에서 고려한 센서 유형을 나타낸다. 참고문헌을 기반으로 센서의 결함 유형을 debonding 유형 (36%, 64%, 84% debonding 그림 7(a) 3가지, breakage 유형 (36%, 64%, 84% breakage 그림 7 (b)) 3가지 그리고 complete failure 총 7가지의 센서 결함 유형을 설정하였다.

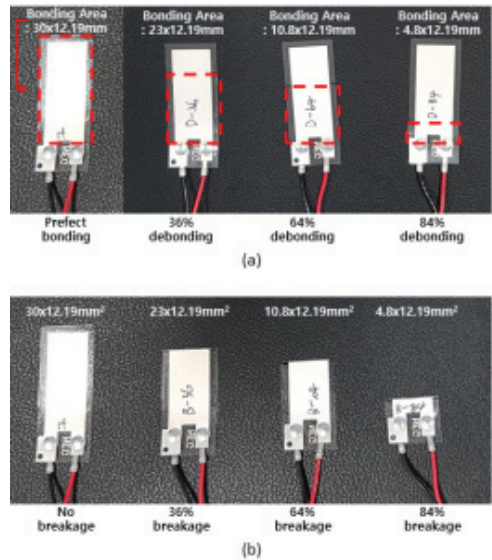


그림 7. PVDF 센서 결함 유형 : (a) 센서 접촉 불량 유형, (b) 센서 파손 유형

Fig. 7 PVDF sensor fault type : (a) sensor debonding case, (b) sensor breakage case

3.2 충격 시험

물체에 충격을 가하여 진동을 발생시키고, 이때 발생한 진동신호를 기반으로 물체의 건전성 평가를 수행하는 방법은 많이 사용되어왔던 방법이다[13]. 따라서 센서 결함 유형별 출력 신호 데이터 수집을 위해 센서 및 SHM 분야의 건전성 평가에 사용되고 있는 임팩트 해머[14]를 사용한 충격 시험을 수행하였다. 그림 8과 같이 센서가 부착된 폴리카보네이트 판에 임팩트 해머를 사용하여 충격을 가하는 실험을 구성하였고, 정상 센서와 결함이 존재하는 센서 간의 출력 신호 차이 비교를 위하여 표 2와 같이 총 8 가지의 실험조건을 설정하였다. 이 때 센서를 통하여 출력된 신호를 측정하였다.

출력 데이터는 DAQ(Data Acquisition) 장비와 PVDF 센서(LDT1-028K)를 사용하여 측정하였다. 머신러닝 알고리즘의 학습을 위한 신뢰성 있는 데이터 획득 및 알고리즘의 정확도 향상을 위하여 각 조건의 실험을 50번씩 반복하여 수행하였으며, 따라서 표 2에 나타난 바와 같이 총 400 번의 충격 실험을 수행하였다. 또한, 각 충격 실험에 따라 출력되는 신호로부터 시간영역 통계 특징들을 모두 추출함으로써 센서 결함 유형 분류를 위해 사용될 데이터 집합을 구축하였다.

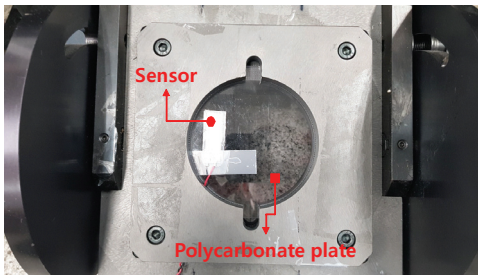


그림 8. 데이터 수집을 위한 실험 구축
Fig. 8 Experimental setup for data collection

표 2. 센서 결함 유형별 실험 진행 횟수
Table 2. Number of experiments by sensor fault type

Sensor Fault Type	Number of Experiments
Healthy State	50
36% Breakage	50
64% Breakage	50
84% Breakage	50
36% Debonding	50
64% Debonding	50
84% Debonding	50
Complete failure	50

IV. 실험 결과 및 고찰

4.1 센서 결함 유형별 출력 데이터 분석 결과

그림 9는 센서 결함 유형별 출력 데이터를 나타낸다. 대체로 비슷한 형상의 파형 신호를 출력하는 것을 확인할 수 있었고, 센서의 결함 유무 및 정도에 따라 센서의 출력 신호값이 점점 작아지는 것을 확인할 수

있으나 센서 결함 유형 중 다른 결함 유형과 구분짓기 어려운 신호가 존재하는 것을 확인할 수 있었다. 이처럼 센서 출력 신호만으로 결함 유무 또는 유형 파악에 어려움이 있음을 확인하였다. 따라서 센서 결함 유무 및 유형 파악을 위해 각 결함 유형별 신호의 특징을 추출하는 작업이 필수적임을 확인하였다.

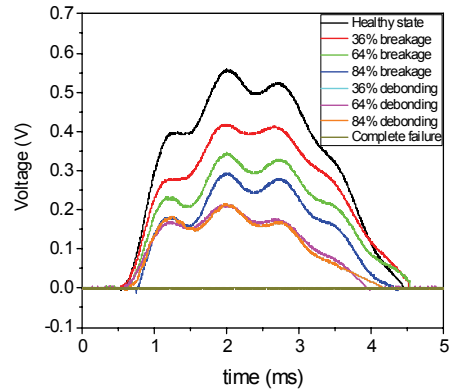


그림 9. 센서 결함 유형에 따른 신호 출력
Fig. 9 Output signal graph by sensor fault type

4.2 시간영역 통계 특징 분석결과

mRMR(minimum Redundancy Maximum Relevance)은 머신러닝 기반 분류에 활용되는 변수들의 상관성과 중복성 두 가지를 고려하여 평가하여 특징들의 중요도 순위를 선정하는 기법이다[15]. 본 연구에서는 센서 결함 유형 분류에 활용된 시간영역 특징들이 중복성에 의해 특징들 중 일부가 실제 분류 학습에서의 성능에 비해 고평가되는 것을 방지하기 위해 mRMR 알고리즘을 활용하여 불필요한 특징들을 배제함으로써 분류 알고리즘의 성능을 향상시키고 학습 데이터에 대해서는 정확하게 예측하지만 실제 데이터가 입력되었을 때 예측에 실패하는 과적합(over-fitting) 현상을 방지하고자 하였다.

그림 10은 mRMR(minimum Redundancy Maximum Relevance)을 통해 센서 결함 분류에 사용된 시간영역 통계 특징들의 중요도를 나타낸다. 그림 10을 통해 시간영역 통계 특징들 중 mean이 센서 결함 감지에 가장 큰 영향을 미치는 것을 확인할 수 있었고 특징 중 kurtosis value, skewness factor, impulse factor 그리

고 crest factor는 결합 유형 감지에 기여하지 못한다는 것을 알 수 있었다.

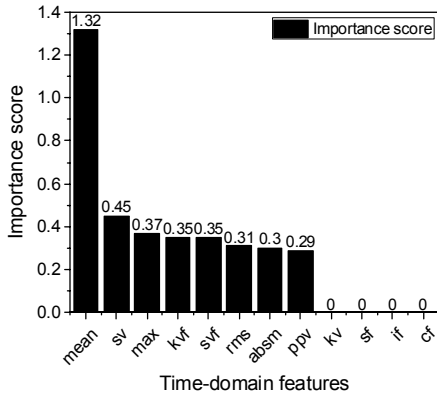


그림 10. 시간영역 특징들 중 중요도 순위
Fig. 10 The importance score of time-domain features

따라서 결합 유형별 출력 신호로부터 중요도 점수가 0점인 4개의 특징들을 제외한 나머지 8개의 특징들을 추출하여 분류 알고리즘 학습을 위한 데이터 집합을 구축하였다.

4.3 분류 알고리즘 성능 평가

각 분류 알고리즘의 성능 평가를 위해 구축한 데이터 집합 중 70%를 분류 알고리즘 모델의 훈련용 데이터로 설정하였으며, 나머지 30%를 테스트용 데이터로 사용하였다. 또한, 분류 알고리즘의 성능 평가에 있어서 가장 많이 사용되고 있는 방법인 10-fold 교차 검증[16]을 통해 분류 알고리즘의 성능 평가 및 검증을 수행하였다. 본 연구에서는 센서 결합 유형 감지에 적합한 알고리즘 선정을 위해 분류 알고리즘별 정확도를 비교하였다.

표 3은 분류 알고리즘별 센서 결합 유형 감지 정확도를 나타낸다. 표 3을 통해 SVM이 가장 높은 정확도로 센서 결합 유형을 분류하는 것을 알 수 있다. 이를 통해 SVM은 센서 결합 유형 분류에 가장 적합한 알고리즘임을 확인하였다. 표 3에서 정확도는 표 4와 같이 분류 알고리즘의 성능을 시각화할 수 있는 혼동 행렬(Confusion Matrix)을 통해 계산되었으며 계산식은 식 (7)과 같다.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad \dots (7)$$

식(7)에서 TP(True Positive)는 Positive를 Positive로 예측한 값, TN(True Negative)은 Negative를 Negative로 예측한 값, FN(False Negative)은 Positive를 Negative로 예측한 값, FP(False Positive)는 Negative를 True로 예측한 값을 의미한다.

표 3. 알고리즘 별 센서 결합 유형 분류 정확도
Table 3. Accuracy of sensor fault type classification by algorithm

Classification Algorithm	Accuracy
Decision Tree	82.4%
kNN	85.0%
Naive Bayes	81.2%
SVM	88.2%

표 4. 혼동 행렬
Table 4. Confusion matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive	False Positive
	Negative	False Negative	True Negative

4.4 하이퍼 파라미터 최적화

본 연구에서는 센서 결합 유형 분류에 높은 정확도를 보였던 SVM의 추가 성능 향상을 위해 커널 함수(Kernel Function)를 도입하였다. 커널 기법은 입력된 데이터를 고차원 공간으로 투영(mapping)시켜 고차원 공간에서 데이터 분류에 적합한 초평면을 찾는 방법이다. 커널 함수의 대표적인 예로는 다항식(Polynomial), RBF(: Radial Basis Function), Sigmoid가 있다. 본 연구에서는 시간영역 통계 특징들로 구성된 고차원 데이터 집합을 커널 함수를 활용하여 센서 결합 유형 감지에 적합한 초평면을 찾고자 하였다. 이를 위해 불필요한 하이퍼 파라미터 반복 탐색을 감소시키고 빠른 시간 내에 최적 하이퍼 파라미터 조합을 찾고, 그 효율성이 뛰어나 머신러닝 분야에서 하이퍼 파라미터 최적화에 많이 활용되고 있는 베이지안 최

표 5. H_SVM 및 분류 알고리즘들별 정확도
Table 5 Accuracy of H_SVM and classification algorithm

Classification Algorithm	Classification accuracy by sensor fault type (%)							
	1 (Healthy)	2 (36% Breakage)	3 (64% Breakage)	4 (84% Breakage)	5 (36% Debonding)	6 (64% Debonding)	7 (84% Debonding)	8 (Complete failure)
Decision Tree	88	88	85	79	62	67	83	100
kNN	84	94	94	75	77	78	75	100
Naive Bayes	82	88	79	75	58	83	88	100
SVM	90	98	91	83	81	72	80	100
H_SVM	90	98	97.1	90.4	90.6	88.9	85	100

적화[17]를 수행하였다. 38회 이상부터 동일한 결과값이 나왔으며, 도출된 하이퍼 파라미터는 커널 함수, 상자 제약 조건(box constraint level), 다중클래스 (multi-class) 방법이며, 결과값은 다음과 같다. 커널 함수는 선형(linear), 상자 제약 조건 수준은 354.15, 다중클래스 방법은 일대일(one-vs-one) 방식이다. 본 연구에서는 이를 적용한 H_SVM을 제안하였다.

그림 11과 12는 각각 하이퍼 파라미터들을 최적화한 SVM으로 센서 결함 유형을 분류한 결과와 분류 결과에 대한 혼동행렬이다.

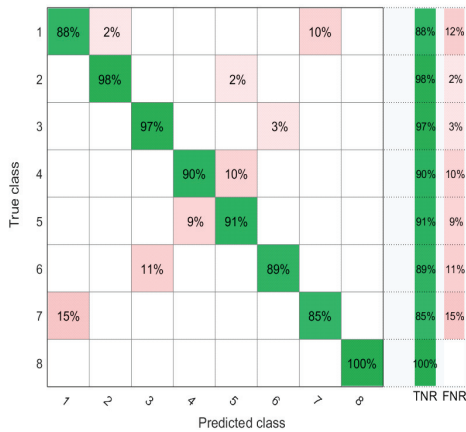


그림 11. H_SVM 분류 결과에 대한 혼동행렬
Fig. 11 Confusion matrix for H_SVM classification results

그림 11에서 좌측과 하단에 있는 숫자 1부터 8은 각각 Healthy state, 36% Breakage, 64% Breakage, 84% Breakage, 36% Debonding, 64% Debonding,

84% Debonding을 의미한다. TPR(True Positive Rates)은 실제 클래스 당 올바르게 분류된 관측치의 비율을 의미하며 FNR(False Negative Rates)은 실제 클래스 당 잘못 분류된 관측치의 비율이다. 예를 들어, 그림 11에서 클래스 6(64% Debonding)은 88.9%가 클래스 6(64% Debonding)으로 분류되었고 나머지 11.1%는 클래스 3(64% Breakage)으로 분류되었음을 나타낸다.

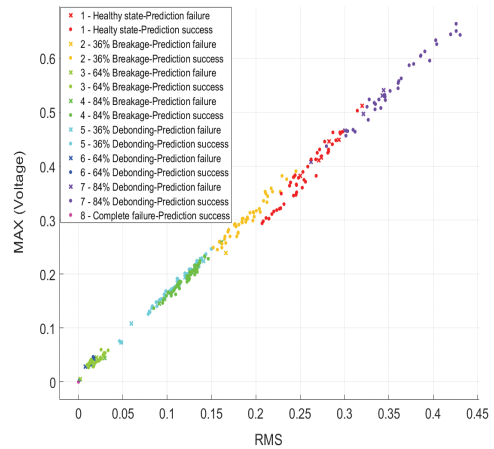


그림 12. H_SVM의 센서 결함 유형 분류 결과
Fig. 12 Result of sensor fault type classification by H_SVM.

그림 11을 통해 가장 분류 정확도가 낮은 클래스는 클래스 7(84% Debonding)로써 85%의 분류 정확도를 보이며, 그림 12에서 나타난 바와 같이 클래스 1(Healthy state)과 클래스 7(84% Debonding)가 특징

공간에서 겹쳐지는 공간이 발생하여 식별에 어려움이 있음을 알 수 있다. TPR과 FNR을 구하는 식은 각각 식 (8), 식 (9)와 같다.

$$TPR = \frac{TP}{TP + FN} \quad \dots (8)$$

$$FNR = \frac{FN}{FN + TP} \quad \dots (9)$$

그림 11의 혼동행렬을 기반으로 정확도를 평가한 결과 92.5%의 정확도를 보였다.

표 5는 센서 결합 유형 진단에 사용된 분류 알고리즘 및 하이퍼 파라미터를 최적화한 SVM의 정확도를 나타낸다. 표 5을 통해 기존 분류 알고리즘들에서 낮은 정확도를 보였던 5(36% Debonding) 클래스와 6(64% Debonding) 클래스가 하이퍼 파라미터를 최적화한 SVM에서는 각각 최대 28.6%, 21.9%의 정확도 향상이 있었다. 이를 통해 하이퍼 파라미터 최적화를 통한 분류 알고리즘의 정확도 향상을 확인하였다.

그림 13은 분류 알고리즘의 성능을 평가하기 위한 분류 알고리즘별 ROC 커브(Receiver Operating Characteristic Curve)를 나타낸다. ROC 커브 아래의 면적 즉, AUC(Area Under ROC Curve) 값을 통해 알고리즘의 성능을 평가한다[18].

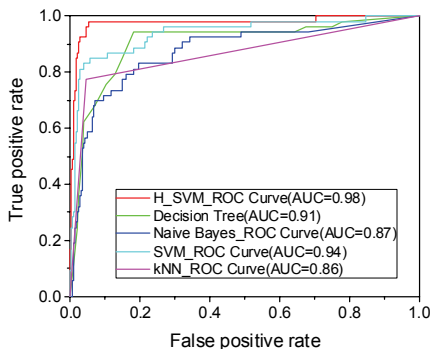


그림 13. H_SVM 과 기타 분류 알고리즘들의 ROC 커브

Fig. 13 ROC curve for H_SVM and classification algorithm

본 연구에서 활용한 분류 알고리즘들에서 낮은 정확도를 보였던 클래스 6(64% Debonding)의 데이터들이 Positive class로 사용되었으며 그 외 클래스들은 Negative class로 사용되었다. 그림 13에서

FPR(False Positive Rate)은 Negative를 Positive로 예측한 비율로 구하는 식은 식 (10)과 같다.

$$FPR = \frac{FP}{TN + FP} \quad \dots (10)$$

그림 13을 통해 H_SVM의 성능이 가장 우수함을 확인할 수 있었다. 그림 13에서 H_SVM의 그래프는 FPR이 0.5 정도에서 TPR 값이 1에 먼저 도달하여 기존 알고리즘들 보다 효과적인 모델임을 보였다. 또한 AUC 값이 기존 분류 알고리즘들과 비교하여 최대 13.95%이상 향상된 값인 0.98로 1에 매우 근접한 값이 나왔다. 따라서 본 연구에서 사용한 H_SVM이 기존의 분류 알고리즘보다 정확하고 안정된 예측을 할 수 있는 모델임을 입증하였다.

V. 결론

본 연구에서는 머신러닝 기반 분류 알고리즘 중 가장 높은 정확도를 보인 SVM에 기반하여 PVDF 센서의 결합 유형을 예측하기 위한 방법론을 제안하였다. 충격 실험을 통하여 정상 센서와 결합이 존재하는 센서들이 출력하는 신호를 수집하였고, 수집한 신호들의 시간영역 통계 특징들을 추출하여 데이터 집합을 구축하였다. 데이터 집합 분석을 위하여 하이퍼 파라미터를 최적화한 H_SVM을 제안하였고, 정상 센서 및 센서 결합 유형들을 예측하였다. 본 연구를 통해 얻은 주요한 결론은 다음과 같다.

정상 센서와 결합이 존재하는 센서들이 출력하는 신호로부터 시간영역 통계 특징들을 추출하여 데이터 집합을 구축하였고, 시간영역 통계 특징들로 구축된 데이터 집합을 SVM과 연계하여 센서의 유형을 분류하였다. 그 결과, 다른 분류 알고리즘들보다 최대 11.3% 높은 정확도를 보였다. 따라서 본 연구에서 제안한 센서 결합 유형 예측 기법은 향후 보다 다양한 센서 결합 유형 데이터를 추가적으로 학습시킴으로써 각종 센서의 신뢰성 구축에 활용될 수 있을 것으로 예상된다.

References

- [1] Y.-H. Huh, J.I. Kim, J.H. Lee, S.G. Hong and J.H. Park, "Application of PVDF Film Sensor to Detect Early Damage in Wind Turbine Blade Components," *Procedia Eng.*, vol. 10, 2011, pp. 3304-3309.
- [2] D.-W. Keum and J.-D. Kim, "Measurement of Apnea Using a Polyvinylidene Fluoride Sensor Inserted in the Pillow," *Journal of Sensor Science and Technology*, vol. 27, no. 6, 2018, pp. 8682-8690.
- [3] D. Tibaduiza, M. Anaya, E. Forero, R. Castro and F. Pozo, "A Sensor Fault Detection Methodology applied to Piezoelectric Active Systems in Structural Health Monitoring Application," *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 138, no. 1, 2016, p. 012016.
- [4] G. Park, C. R. Farrar, F. L. di Scalea, and S. Coccia, "Performance Assessment and Validation of Piezoelectric Active-Sensors in Structural Health Monitoring," *Smart Mater. Struct.*, vol. 15, no. 6, 2006, pp. 1673-1683.
- [5] T.-C. Huynh, T.-D. Nguyen, D.-D. Ho, N.-L. Dang, and J.-T. Kim, "Sensor fault diagnosis for impedance monitoring using a piezoelectric-based smart interface technique," *Sensor*, vol. 20, no. 2, 2020, pp. 510-530.
- [6] B. Samanta, "Gear fault detection using artificial neural networks and support vector machines with genetic algorithms," *Mech. Syst. Signal Process.*, vol. 18, no. 3, 2004, pp. 625-644.
- [7] T. W. Rauber, F. de A. Boldt, and F. M. Varejão, "Heterogeneous feature models and feature selection applied to bearing fault diagnosis," *IEEE Trans. Ind. Electron.*, vol. 62, no. 1, 2015, pp. 637-646.
- [8] N. S. Altman, "An Introduction to Kernel and Nearest-neighbor Nonparametric Regression," *The American Statistician*, vol. 46, no. 3, 1992, pp. 175-185.
- [9] L. Rokach and O. Maimon, "Top-down Induction of Decision Trees Classifiers-a Survey," *IEEE Transactions on Systems, Man and Cybernetics Part C : Applications and Reviews*, vol. 35, no. 4, 2005, pp. 476-487.
- [10] S.-J. Yu, "A Study on Recommendation Method for Real Estate Using Naive Bayes Classification," *The Journal of Korean Institute of Information Technology*, vol. 17, no. 10, 2019, pp. 115-120.
- [11] S. U. Jan, Y.-D. Lee, J. Shin, and I. Koo, "Sensor fault classification based on support vector machine and statistical time-domain features," *IEEE Access*, vol. 5, 2017, pp. 8682-8690.
- [12] D. Ai, H. Luo, and H. Zhu, "Diagnosis and validation of damaged piezoelectric sensor in electromechanical impedance technique," *Journal of Intelligent Material Systems and Structures*, vol. 28, no. 7, 2017, pp. 837-850.
- [13] A. M. AY, Y. Wang and S. Khoo, "Signal processing for time domain analysis of impact hammer test data," *8th European Workshop on Structural Health Monitoring, EWSHM*, vol. 1, 2016, pp. 628-637.
- [14] X. Deng and V. Giurgiutiu, "Impact Monitoring and Fault Detection Using Piezoelectric Transducers and Wavelet Analysis," *53rd Meeting of the Society for Machinery Failure Prevention Technology, Virginia Beach, VA.*, 1999 pp. 167-172.
- [15] H. Peng, F. Long, and C. Ding, "Feature Selection Based on Mutual Information: Criteria of Max-Dependency, Max-Relevance, and Min-Redundancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 8, 2005, pp. 1226-1238.
- [16] S. Arlot and A. Celisse, "A survey of cross-validation procedures for model selection," *Stat. Surv.*, vol. 4, 2010, pp. 40-79.
- [17] B. Shahriari, K. Swersky, Z. Wang, R. P. Adams and N. de Freitas, "Taking the Human out of the Loop: a Review of Bayesian Optimization," *Chinese Journal of Evidence-Based Medicine*, vol. 14,

no. 10, 2014, pp. 1270 - 1275.

- [18] T. Fawcett, "An Introduction to ROC Analysis," *Pattern Recognit. Lett.*, vol. 27, no. 8, 2006, pp. 861-874.

저자 소개



김승욱(Seung-Wook Kim)

2019년 2월 강원대학교 기계의용
공학과 졸업(공학사)

2021년 2월 한양대학교 대학원 융
합국방학과 졸업(공학석사)

2021년 ~현재 동남권원자력의학원 연구원

※ 관심분야 : 센서, 인공지능



이상민(Sang-Min Lee)

2021년 2월 한양대학교 기계공학
부 졸업(공학사)

2021년 ~현재 고려대학교 기계공학부 석박통합과
정 재학

※ 관심분야 : 인공지능