JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 31 No. 11, pp. 343-359, November 2026
https://doi.org/10.9708/jksci.2026.31.11.343

Spatial Optimization-Based Input-Level Fusion for Computationally Efficient Bearing Fault Diagnosis

Yoonsik Cho*, Cheol Hong Kim**

*M.S. Student, Dept. of Computer Science and Engineering, Soongsil University, Seoul, Korea
**Professor, School of Computer Science and Engineering, Soongsil University, Seoul, Korea

[Abstract]

In resource-limited industrial edge environments, diagnosing variable-speed bearing fault is challenging. To address single-sensor limitations, feature-level fusion overhead, temporal data leakage from overlapping splits, this paper proposes a lightweight Multi-sensor Input-level fusion framework. Heterogeneous Acoustic Emission(AE) and vibration signals are synchronized at 0.1-second intervals and converted into 2D Short-Time Fourier Transform(STFT) spectrograms via session and file-level independent partitioning with 0% overlap. These spectrograms are then horizontally concatenated at the input stage, spatially optimized to minimize extra computation, and fed into MobileNet-V3 Small backbone. Experiments demonstrate that Zero-padding fusion method achieves 98.77% accuracy, consistently outperforming single-sensor models with robust performance in highly distorted roller-faults classification and in t-Distributed Stochastic Neighbor Embedding (t-SNE) clustering.

▸ **Key words**: Bearing fault diagnosis, Multi-sensor fusion, Input-level fusion, Temporal data leakage, Lightweight Neural Network

[요 약]

연산 자원이 제한된 산업용 엣지 컴퓨팅 환경에서 가변 속도 베어링 고장을 효율적으로 진단하기 위해, 본 연구는 단일 센서 정보 한계, Feature-level fusion의 연산량 과부하, 오버랩 분할로 인한 시간적 데이터 누수를 개선하는 경량 Multi-sensor Input-level fusion 프레임워크를 제안한다. 음향 방출(Acoustic Emission, AE) 및 진동 신호를 0.1초 단위로 동기화한 후, 데이터 누수를 방지하기 위해 세션·파일 단위의 독립적 분할과 0% 오버랩 조건을 적용하여 2D 단시간 푸리에 변환(Short-Time Fourier Transform, STFT) 스펙트로그램으로 변환한다. 생성된 스펙트로그램은 입력 단계에서 수평으로 병합되고, 연산량 증가를 최소화하는 공간 최적화를 거쳐 MobileNet-V3 Small 백본에 입력된다. 실험 결과, 제안한 Zero-padding 융합 모델은 98.77%의 정확도로 단일 센서 모델을 상회했으며, 왜곡이 심한 롤러 결함 분류와 t-SNE 특징 군집화에서도 강건성을 확인하였다.

▸ **주제어**: 베어링 고장 진단, 다중 센서 융합, 입력 융합, 데이터 누수, 경량 신경망

• First Author: Yoonsik Cho, Corresponding Author: Cheol Hong Kim
*Yoonsik Cho (cys3553@soongsil.ac.kr), Dept. of Computer Science and Engineering, Soongsil University
**Cheol Hong Kim (cheolhong@ssu.ac.kr), School of Computer Science and Engineering, Soongsil University
• Received: 2026. 08. 27, Revised: 2026. 10. 31, Accepted: 2026. 11. 11.

Copyright © 2026 The Korea Society of Computer and Information
http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945
www.kci.go.kr

I. Introduction

스마트 팩토리, 발전 터빈 등 다양한 산업에서 필수적인 장치로 지속적으로 활용되고 있는 회전 기계 시스템에서 베어링은 축의 하중을 지지하고 마찰을 최소화하는 핵심 부품이다[1]. 장시간의 고속 회전과 복합적인 하중 변동 환경에 노출되는 베어링은 내륜, 외륜, 롤러 등 다양한 부위에 미세 균열이나 박리와 같은 기계적 결함이 발생하기 쉽다[2]. 베어링 결함은 시스템 전체의 이상 진동과 소음을 유발하고 유지보수가 이루어지지 않을 경우 예기치 못한 가동 중단과 막대한 경제적 손실로 이어질 수 있다. 따라서 회전 기계의 신뢰성과 안전성을 보장하기 위해 베어링의 이상 징후를 조기에 감지하고 고장 유형을 정확하게 분류하는 고장 진단 기술의 중요성이 더욱 커지고 있다[3].

사물인터넷 센서 기술 및 딥러닝 알고리즘이 비약적으로 발전하면서, 데이터를 기반으로 한 고장 진단 기술이 널리 활용되고 있다. 전통적인 신호 처리 및 전문가 시스템 기반 진단 방식은 복잡한 도메인 지식과 수동적인 특징 추출에 의존해야 하는 한계가 있었다[1][4]. 반면, 합성곱 신경망(Convolutional Neural Network, CNN)으로 대표되는 딥러닝 모델은 수집된 측정 신호로부터 고장 특징을 자동으로 추출하고 패턴을 학습함으로써 높은 진단 정확도를 제공해 왔다[5]. 하지만 실제 산업 현장에서 딥러닝 기반 고장 진단 시스템을 성공적으로 운용하기 위해서는 여전히 극복해야 할 한계점들이 존재한다.

첫째는 단일 센서 기반 고장 진단이 가지는 정보의 한계이다. 고장 진단에 널리 사용되는 진동(vibration) 신호는 거시적인 구조적 결함을 포착하는 데 유리하지만 초기 미세 균열 감지에 취약하며 저주파 배경 소음에 쉽게 간섭받는다[6]. 반면, 음향방출(Acoustic Emission, AE) 신호는 미세 결함을 민감하게 잡아내지만 고장 후반부에 신호가 포화되는 단점이 있다[2]. 따라서 복잡한 실전 환경에서는 상호 보완적인 이기종 센서 신호를 융합하여 진단의 강건성을 확보하는 것이 필수적이다[7][8]. 둘째는 다중 센서 융합 모델의 연산 비대화 문제이다. 기존의 다중 센서 융합 연구들은 각 센서 신호를 처리하기 위해 독립적인 신경망 분기를 여러 개 병렬로 구성하는 Feature-level fusion 기법을 주로 채택했다[8]. 그러나 이는 모델의 파라미터 수와 부동소수점 연산량을 크게 증가시키기 때문에 가용 자원이 제한된 산업용 엣지 디바이스 환경에 적용하는 것을 사실상 불가능하게 만든다[9][10]. 셋째는 학습 데이터 분할 시 발생하는 데이터 누수(Data leakage)로 인한 모델 성능의 과대평가 문제이다[11]. 최근 다수의 고장 진단 연구들이 데이터셋을 구성할 때 하나의 시계열 신호를 오버랩시켜 분할한 뒤 무작위로 섞어 train 및 test 세트를 분리한다[11]. 이 경우 동일 시간대의 고주파 잡음 특성이 양쪽 세트에 그대로 공유되는 시간적 데이터 누수(Temporal data leakage) 현상이 발생하여, 딥러닝 모델이 결함의 본질적 주파수를 학습하지 않고 정답을 암기하게 됨으로써 실제 일반화 성능이 왜곡되는 부작용을 낳고 있다[12][13].

이러한 문제점들을 해결하기 위해 본 논문에서는 연산 자원이 제한된 산업용 환경에서도 가변 속도로 구동되는 회전 기계의 베어링 고장을 효율적이고 정밀하게 진단할 수 있는 경량 Multi-sensor Input-level fusion 프레임워크를 제안한다. 제안 기법에서는 진동과 AE 신호가 지닌 상호 보완적인 물리적 특성을 모델 내부 연산이 개입하기 전인 전처리 단계에서 단일 텐서로 결합하는 Input-level fusion 방식을 채택하였다. 또한 이기종 2D 스펙트로그램 결합 시 발생하는 해상도의 비대칭성을 해결하기 위해 딥러닝의 연산 오버헤드를 유발하지 않는 구조적 텐서 최적화 기법인 Zero-padding 방식과 Tiling 방식을 적용하여 그 실효성을 비교 분석한다. 또한, 모델 평가의 객관성을 확보하기 위해 시계열 신호 분할 시 오버랩 비율을 0%로 통제하여 시간적 데이터 누수 현상을 전처리 단계에서 봉쇄하는 신뢰성 있는 평가 프로토콜을 수립하였다[11][12][13]. 더 나아가, 모델이 학습 과정에서 한 번도 관측하지 못한 미지의 구동 속도 데이터(cross-speed)를 테스트 환경에 인가함으로써, 주파수 왜곡과 smearing 현상이 심한 롤러 결함과 같은 조건에서도 제안 기법이 탁월한 식별 강건성을 발휘할 수 있음을 입증하고자 한다.

본 논문의 구성은 다음과 같다. 2장에서는 회전 기계의 고장 진단에 사용되는 진동 및 AE 신호의 물리적 특성과 시간-주파수 변환을 위한 Short-Time Fourier Transform(STFT) 기법을 설명하고, 기존 Multi-sensor fusion 연구의 한계점 등 본 연구의 기반이 되는 이론적 배경과 관련 연구 동향을 고찰한다. 3장에서는 본 연구의 핵심인 Multi-sensor Input-level fusion 프레임워크의 전체 구조를 제안한다. 이기종 신호의 위상 동기화 및 전처리 과정, 그리고 텐서 규격을 맞추기 위해 고안된 두 가지 공간 최적화 기법(Zero-padding 및 Tiling 기법)의 구체적인 아키텍처를 상세히 서술한다. 4장에서는 제안하는 진단 프레임워크의 성능을 검증하기 위한 실험 환경을 제시하고, 결과를 다각도로 분석한다. 시간적 데이터 누수를 배제한 미학습 가변 속도 도메인 일반화 평가

www.kci.go.kr

(Cross-speed domain generalization evaluation) 시나리오 하에서 제안 기법과 대조군들의 고장 진단 정확도를 비교하며, 혼동 행렬, t-Distributed Stochastic Neighbor Embedding (t-SNE) 및 Grad-CAM 분석을 통해 결함 식별 성능과 특징 학습 경향을 다각도로 살펴보고자 한다. 마지막으로 5장에서는 본 연구의 핵심 결과 및 기여도를 종합적으로 요약한다.

II. Related Work

2.1 Domain-Specific Physical Characteristics of Vibration and AE Signals

회전 기계의 고장 진단 시스템을 설계할 때, 진단 대상인 롤러 베어링의 고장 상태를 명확하게 보여줄 수 있는 물리적 신호를 선택하는 것은 신뢰성을 확보하기 위해 매우 중요하다. 기계 고장 진단 분야에서 널리 활용되는 신호로는 진동과 AE 신호가 있다[1]. 진동 신호는 가속도계를 구동축 주변 기계 하우징 구조물에 부착하여 베어링 내부 손상 부위와 롤러 요소 간의 마찰, 파손 충격으로 인해 유발되는 기계적 탄성 변위 및 저주파 대역 가속도 진폭을 측정하는 신호이다. 베어링의 외륜이나 내륜에 손상이 가해지면 회전체 특성에 따라 고유한 주파수 특성이 저주파 및 중주파 대역에 걸쳐 주기적인 구조적 진동 패턴으로 매핑된다[1]. 따라서 기계의 거시적 결함 식별이나 파손 상태를 식별하는 데 높은 안정성을 보인다. 그러나 진동 기반 상태 모니터링은 주변 환경의 저주파 기계 잡음 및 주기적인 배경 소음에 간섭을 받기 쉬우므로 신호 대 잡음비(Signal-to-Noise Ratio, SNR)가 상대적으로 낮다는 한계가 있다[6]. 특히 초기 미세 균열이나 표면 박리와 같이 충격 진폭이 미미한 시점에서는 구조적 진동이 약하게 발생하여 초기 결함을 감지하는 데 어려움이 존재한다. 이러한 진동 센서의 감도 장벽과 저주파 배경 잡음 매몰 한계를 극복하기 위해 비파괴 검사 기법에서 유래한 AE 신호가 상태 감지에 다각도로 활용되고 있다[7]. AE 신호는 베어링 내부 부품 표면에 미세 균열이 발생하거나 변형이 시작되는 나노 또는 마이크로 스케일의 파괴 진전 단계에서, 재료 내부의 변형 에너지가 탄성파의 형태로 급격히 방출되는 물리 현상을 포착한다. 일반적으로 AE 센서의 유효 측정 주파수 대역은 100 kHz에서 수 MHz에 이르는 초고주파 대역에 국소화되어 있어, 구동 전동기의 주기적인 기계 잡음이나 기어 맞물림 등 저주파 노이즈 성분이 존재하는 대역과 물리적으로 격리되어 있다. 따라서 외부 잡음 환경에 간섭을 받지 않고 고장 초기의 초동 파동 성분을 높은 감도로 탐지하는 데 우수성을 지닌다[2]. 그러나 AE 신호는 고주파 대역을 연속 샘플링해야 하므로 전처리 메모리 버퍼 오버헤드가 극심하고, 결함이 심화되어 표면이 마모되거나 파손 부위가 커질 경우 연속적인 마찰 신호로 전이되어 파형이 포화되는 현상이 발생해 거시적 손상 추정력이 약화된다. 이처럼 진동 신호의 구조 역학적 저주파 특성과 AE 신호의 물리적 고주파 특성은 고장 식별에 있어 상호 보완성을 지니고 있으므로, 최근에는 두 모달리티를 동시에 활용하여 진단의 일반화 강건성을 높이려는 다중 센서 융합 연구가 활발히 진행되고 있다[8][14].

2.2 Short-Time Fourier Transform(STFT)

1차원 시계열 형태의 진동 및 AE 신호는 시간에 따라 통계적 주기가 변화하는 대표적인 non-stationary 신호이다. 특히 가동 속도가 수시로 급변하는 가변 속도 조건에서는 회전 주파수에 종속되는 베어링 고유의 결함 진폭 성분이 시간에 따라 동적으로 압축되거나 신장하는 비선형 변조 특징을 겪게 된다[15]. 따라서 전통적인 이산 푸리에 변환(Discrete Fourier Transform, DFT)을 통해 시간축 정보를 완전히 유실하고 정적인 주파수 스펙트럼 성분만 도출하는 방식은 주파수 성분의 시간적 변동성을 추적할 수 없어 일반화 정확도가 크게 하락한다.

이러한 한계를 극복하기 위해 시간 도메인의 비정상 신호를 주파수 축의 강도 변화와 동시에 국소화하여 표현할 수 있는 STFT가 표준적인 전처리 기법으로 널리 활용된다[16]. STFT의 기본적인 수학적 매핑 정의식은 [Eq. (1)]과 같이 표현된다.

$$X(t,f) = \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f\tau} d\tau \qquad (1)$$

$X(t,f)$: 시간 t와 주파수 f에 대한 2차원 시간-주파수 스펙트로그램
$x(\tau)$: 입력된 1차원 시계열 신호
$w(\tau - t)$: 시간 t를 중심으로 하는 분석 창 함수
f : 주파수 변수
τ : 시간축 상의 이동 변수

여기서 $x(\tau)$는 1차원 raw 시계열 신호이며, $w(\tau - t)$는 시간축 상에서 특정 구간의 신호만을 추출하기 위해 적용되는 분석 창 함수(analysis window function)이다. STFT는 연속적인 1차원 신호 전체에 일괄적으로 변환을

적용하는 것이 아니라, 창 함수 $w(\tau - t)$를 시간 t만큼 전방 갱신시키며 신호를 국소화한 상태에서 독립적인 이산 주파수 분석을 연속적으로 진행한다. 이때 창 함수의 물리적인 크기(window length)와 프레임 간의 이동 간격을 나타내는 hop size는 주파수 해상도와 시간 해상도의 상충 관계(resolution trade-off)를 제어한다. 기존의 머신러닝 및 딥러닝 기반 고장 진단 연구들에서는 train 데이터셋의 양을 인위적으로 증강하기 위해 윈도우 슬라이싱 과정에서 hop size를 줄여 오버랩을 상당 비율 허용하는 기법이 사용되어왔다. 그러나 최근 연구들에 따르면, 시간 영역 신호에서 오버랩을 허용하여 생성된 스펙트로그램 데이터셋은 train과 test 사이에 동일한 시간대의 고주파 잡음 특성이나 위상 정보가 그대로 공유되는 시간적 데이터 누수를 유발한다[11][13]. 이러한 데이터 누수가 발생할 경우 딥러닝 모델이 결함의 본질적인 주파수 패턴을 학습하는 것이 아니라, 중첩된 배경 노이즈의 shortcut을 기억하게 되어 진단 성능 지표가 비정상적으로 과대평가되는 심각한 오류를 초래한다[13]. 따라서 모델의 도메인 일반화 성능을 객관적으로 평가하기 위해서는, 데이터 분할 시 오버랩을 완전히 배제하는 슬라이싱을 통해 시계열 특징의 시간적 독립성을 확보해야 할 필요성이 커지고 있다[12].

2.3 Technological Trends in Multi-Sensor Information Fusion

이기종 다중 센서(heterogeneous sensors) 데이터를 유기적으로 통합하는 정보 융합 기술은 수집 및 인코딩이 수행되는 네트워크 계층에 따라 크게 Decision-level fusion, Feature-level fusion, 그리고 Data-level 또는 Input-level fusion 방식으로 분류된다[3].

최하단 연동 방식에 해당하는 Decision-level fusion은 각 센서 도메인별로 완전히 독립되게 설계된 분류기들이 도출한 최종 클래스별 확률 벡터나 진단 스코어를 결합하는 방식으로, 대표적으로 다수결 투표 규칙(Voting rules), 베이지안 추론(Bayesian inference), 뎀프스터-셰퍼(Dempster-Shafer) 증거 이론 기반의 앙상블 모델이나 개별 센서 출력을 유클리드 거리로 매핑하는 전통적인 k-최근접 이웃(k-Nearest Neighbors, k-NN) 결합 아키텍처 등이 이에 속한다[7][8]. 이 Decision-level 방식은 특정 센서에 하드웨어적 단선이나 신호 유실이 발생하더라도 다른 분류 분기가 독립적으로 작동하여 진단 시스템 전체의 기능 마비를 막아주는 높은 복구 강건성을 지니지만, 결함 정보 추출의 초기 단계에서 발생하는 이기종 신호 간의 비선형적인 교차 상관관계(Cross-modal correlation)를 포착하지 못한 채 표면적인 분류 결과만을 취합하므로 회전 속도가 복잡하게 변하는 변동 도메인 환경에서는 결함 식별 일반화 성능을 확보하기 어렵다는 단점이 있다[8].

이러한 한계를 극복하기 위해 다중 센서 기반 고장 진단 분야에서 가장 보편적으로 채택한 것은 Feature-level fusion이며, 대표적인 모델 군으로는 AE 신호와 진동 신호를 각각 별도의 파이프라인으로 처리하는 다중 분기 합성곱 신경망(Multi-branch CNN), 병렬 레이어 파이프라인 결합 모델, 혹은 다차원 은닉층 단계에서 추출된 개별 특징 맵들을 연결하거나 교차 어텐션(cross-attention) 메커니즘으로 엮어내는 심층 신경망 아키텍처가 활용된다. Feature-level fusion은 각 센서 특징에 최적화된 독립적 커널 필터를 통해 깊이 있는 다차원 고장 패턴을 풍부하게 탐지할 수 있어 정적 가동 조건 하에서 매우 우수한 진단 적응력을 보여주지만, 활용되는 센서 모달리티가 확장될수록 전체 네트워크의 레이어 깊이와 가중치 파라미터 수, 그리고 부동소수점 연산량이 기하급수적으로 증가하는 병목 현상을 유발한다[14]. 특히 은닉 레이어 내부에서 고차원 특징 텐서들을 동적으로 연산하는 과정은 하드웨어의 메모리 대역폭 점유율을 급격히 높여 심각한 연산 지연을 초래하므로, 가용 자원이 극도로 제한된 산업용 엣지 디바이스 환경이나 경량 임베디드 진단 모듈에 모델을 가볍게 배포하여 고장 진단을 효율적으로 수행하기에는 구조적 실효성이 현저히 떨어진다는 한계가 존재한다[9].

이에 반해, 하드웨어 연산 비용의 과부하를 방지할 수 있는 대안으로 조명받고 있는 Data-level 혹은 Input-level fusion 기법은 딥러닝 모델 내부의 복잡한 연산이 개입하기 전, 전처리 단계에서 서로 다른 센서의 이상 정보를 단일한 다차원 텐서 형태로 물리적 통합을 완결하는 방식이며, 대표적인 파이프라인 구조로는 raw data를 다채널로 적재하는 조기 융합 아키텍처나 다중 센서의 스펙트로그램을 전단 버퍼에서 병합하여 인가하는 단일 스트림(single-stream) 백본 네트워크 모델 등이 포함된다[7]. 입력 융합 기법은 이기종 센서가 제공하는 raw 주파수 및 시간 도메인의 결함 징후를 정보 손실 없이 온전하게 보존할 수 있다는 물리적 장점이 있을 뿐만 아니라, 입력 전처리 수준에서 특징 공간의 통합이 완결되므로 후속 분류기 신경망의 내부 구조를 다중 분기로 확장하거나 수정할 필요가 없어 MobileNet-V3 Small과 같은 단일 스트림 경량 백본 본연의 저연산·저메모리 사양과 부동소수점 연산의 효율성을 확보할 수 있다는 측면에서 구조적 이점을 제공한다[14]. 다만, 주파수 응답 특성과 샘플링 비율

www.kci.go.kr

Table 1. Comparison of representative bearing fault diagnosis studies and the proposed framework.

Study	Sensor Fusion Stage	Input Representation	Backbone Structure	Computational Cost	Data-Leakage Control	Cross-Speed Evaluation
Zhang et al. [6]	Single (Vib)	1D raw signal	1D WDCNN	Low	X (Overlapped)	O
Pham et al. [15]	Single (Vib)	2D STFT	VGG-16	High	X	O
Wan et al. [14]	Early layer (Vib+Sound)	Signal + 2D fusion layer	Multiscale CNN	Moderate-to-High	X	O
Ong et al. [8]	Feature-level (Vib+Acoustic)	Dual 2D STFT (flattened)	Dual-branch CNN	High (2x)	X	X
Proposed	Input-level (Vib+AE)	Concat 2D STFT + Zero-padding / Tiling	Single MobileNet-V3 Small	Ultra-low (0.055515 GFLOPs)	O (File-disjoint, 0% overlap)	O (Unseen RPMs)

이 상이한 이기종 센서 간의 해상도 불일치나 물리적 비율 왜곡 문제를 신경망 외적인 부담 없이 해결해야 하므로, 모델 연산 비용을 늘리지 않으면서도 합성곱 커널의 수용 영역(receptive field) 내에서 공간 여백을 제어하고 유효 결함 패턴의 노출 빈도를 최적화할 수 있는 전처리 단계의 구조적 텐서 제어 및 공간 최적화 메커니즘이 안정적으로 선행되어야만 고장 진단의 신뢰성과 강건성을 제공할 수 있다[7].

2.4 Comparison with Prior Studies and Differentiation of the Proposed Framework

다중 센서 기반 베어링 고장 진단에서 입력단 융합(Input-level/Early fusion)은 다중 분기 네트워크를 사용하는 특징단 융합(Feature-level fusion) 대비 연산 효율성이 높다는 장점이 있으나, 이기종 센서의 2D 시간-주파수 스펙트로그램을 단일 백본 네트워크의 표준 입력 규격(224x224)으로 결합하는 방식에 따라 특징 표현력과 연산 오버헤드가 크게 달라진다. 기존의 초기 융합 연구들은 주로 다중 신호를 채널 방향으로 단순 적층(channel-wise stacking)하거나 네트워크 최전단에 별도의 2차원 융합 합성곱 계층(2D fusion layer)을 추가하여 결합 특징을 추출하는 방식을 취해왔다[7][8][14]. 그러나 음향 방출(AE, 250 kHz)과 진동(vibration, 65.5 kHz) 신호처럼 샘플링 주파수와 물리적 응답 대역이 상이한 이기종 스펙트로그램을 채널 방향으로 적층할 경우, 동일한 2차원 공간 좌표상에 서로 다른 물리적 주파수 대역과 시간 분해능 격자가 중첩되는 주파수 정렬 불일치(spectral misalignment)가 발생하여 합성곱 커널의 공간-주파수 국소성을 훼손하게 된다. 또한 수평 병합된 직사각형 스펙트로그램(112x224)을 입력 규격(224x224)에 맞추기 위해 종횡비를 강제로 늘리는 비균일 보간을 적용하면 베어링 결함 고유의 공진 대역폭과 과도 충격 주기의 기하학적 비율이 왜곡되는 한계가 존재한다. 이러한 기존 입력단 융합의 구조적 한계를 극복하기 위해, 본 연구는 두 센서의 (112x112) 스펙트로그램을 수평 병합(112x224)하여 각 센서 고유의 시간-주파수 해상도를 원형 그대로 보존한 상태에서, 추가적인 레이어나 연산량(Floating Point Operations, FLOPs) 증가 없이 정방형 입력 규격과의 공간적 불일치를 해소하는 두 가지 공간 배치인 Zero-padding과 Tiling을 설계하고 그 작용 메커니즘을 비교 규명한다. 특히 비교 전략으로 도입한 Tiling 기법은 컴퓨터 비전 분야의 소형 객체 검출 및 인스턴스 분할에서 제한된 입력 캔버스 내에 타겟 패턴을 복제 배치하여 합성곱 수용장(receptive field) 내 특징 활성화 밀도와 공간 이동 불변성(translation invariance)을 향상시키는 데이터 증강 원리[17][18][19]에 기반을 두고 있다. 2D 스펙트로그램을 입력받는 CNN은 시간 및 주파수 축 모두에 대해 2차원 공간 합성곱과 다운샘플링을 수행하므로, 병합된 112x224 스펙트로그램을 상하로 복제하여 224x224 텐서를 구성하면 단일 순전파 내에서 과도 충격 및 공진 주파수가 합성곱 필터에 포착될 공간적 확률이 높아져 초기 수렴을 가속할 수 있다는 경험적 가설이 성립한다. 반면, 자연 이미지의 일반 객체와 달리 스펙트로그램의 수직축은 절대적인 물리적 주파수 인덱스를 의미하며 배경 변조 잡음을 포함하고 있으므로, 패턴의 인위적 반복(Tiling)이 후기 미세 최적화 단계에서 배경 노이즈의 중복 노출이나 상·하단 경계면의 주파수 불연속성을 유발할 수 있는지에 대해, 잔여 여백을 영행렬로 공간 격리(spatial isolation)하여 물리적 척도를 순수하게 보존하는 Zero-padding 전략과의 실증적 비교 검증이 요구된다. [Table 1]은 대표적인 베어링 고장 진단 선행 연구 4편[6][8][14][15]과 본 연구의 제안 프레임워크를 센서 융합 단계(Sensor Fusion Stage), 입력 표현(Input Representation), 백본 구조(Backbone Structure), 연산 비용(Computational Cost), 데이터 누수 통제(Data-Leakage Control), 그리고 미학습 속도 교차 평가(Cross-Speed Evaluation)의 6가지 핵심 차원에서 요약 비교하여 보여준다.

[Table 1]에 보이는 바와 같이, 기존의 단일 센서 기반

www.kci.go.kr

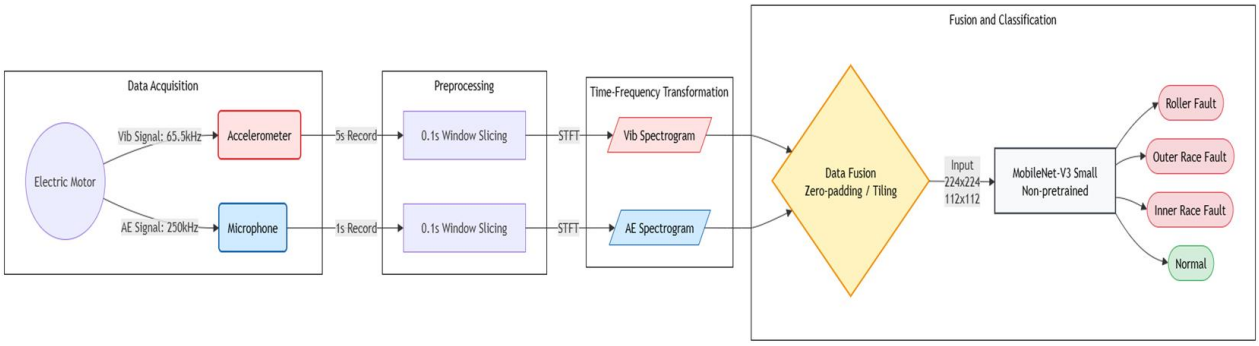


Fig. 1. Schematic diagram of the proposed diagnostic framework

연구들[6][15]은 중첩 슬라이싱으로 인해 시간적 데이터 누수(Temporal data leakage)에 취약하거나 VGG-16과 같은 고비용 백본에 의존하면서도 단일 물리 도메인의 한계를 지니며, 다중 센서 융합 연구들[8][14]은 별도의 2D 융합 계층이나 병렬 특징 추출 분기(dual-branch)를 추가하여 연산 비용이 증가하고 데이터 누수 통제나 미학습 속도 일반화 검증이 제한적이었다. 이와 달리 본 연구는 1) 별도의 융합 레이어나 다중 분기 확장 없이 이기종(Vib+AE) 스펙트로그램의 물리적 주파수 척도를 보존하면서 공간 배치(Zero-padding 및 Tiling)만으로 입력 텐서를 최적화하여 다중 센서 융합에 따른 추가 연산 오버헤드를 0%(0.055515 Giga Floating Point Operations, GFLOPs)로 억제하고, 2) 세션·파일 단위 독립 분할 및 0% 비중첩 슬라이싱을 통한 엄격한 데이터 누수 원천 차단과 3) 학습에 사용되지 않은 미지의 회전 속도(unseen cross-speed) 일반화 평가를 단일 경량 백본(MobileNet-V3 Small) 내에서 동시에 달성하였다는 점에서 기술적 차별성을 갖는다.

III. Proposed Method

3.1 Overall System Framework

본 연구에서 제안하는 Multi-sensor Input-level fusion 기반 베어링 고장 진단 프레임워크는 연산 및 메모리 자원이 제한적인 산업용 엣지 디바이스 환경에서 가변 속도 구동축의 비선형적 변동 속에서도 높은 수준의 분류 정확도를 달성하는 것을 목적으로 설계되었다. 기존의 다중 센서 융합 기반 연구들은 각 센서 신호별로 독립적인 신경망 분기를 병렬로 배치한 후, 네트워크의 심층 레이어에서 추출된 특징 맵을 결합하는 Feature-level fusion 방식을 주로 채택해왔다[8][14]. 그러나 다중 분기 네트워크 구조는 활용되는 센서의 도메인이 확장될수록 모델의 총 파라미터 수와 부동소수점 연산량, 그리고 특징 맵의 메모리 점유 공간이 배수 단위로 증가하여 고장 진단 시스템의 비대화를 초래한다[9]. 이를 극복하기 위해 본 논문에서는 이기종 센서 데이터의 물리적 특성을 모델 전단의 입력 버퍼 단계에서 다차원 공간상으로 선제 통합하는 Input-level fusion 기법을 제안한다. 제안 기법은 이기종 신호 간의 상호 보완적인 고장 특징을 단일 채널의 2D 텐서 공간 버퍼로 병합하므로 후속 분류기 네트워크가 추가적인 아키텍처 확장이나 연산 오버헤드 없이 단일 경량 백본 네트워크만으로 결함 특징을 전역적으로 인코딩할 수 있도록 유도한다. 제안하는 Multi-sensor Input-level fusion 고장 진단 프레임워크의 전체적인 파이프라인 및 시스템 구조도는 [Figure 1]과 같다.

제안 기법은 3상 유도 전기 모터 구동 환경에서 수집된 250 kHz의 AE 신호와 65.5 kHz의 진동 신호를 기반으로 작동하며, raw data로부터 최종 상태를 분류하기까지 크게 네 단계의 파이프라인을 거쳐 수행된다. 첫째, 서로 다른 측정 단위를 가진 이기종 신호를 0.1초 단위로 분할하여 시간축을 동기화한다. 둘째, 동기화된 각 1차원 시계열 프레임에 STFT를 적용하여 시간-주파수 도메인의 2D 스펙트로그램으로 매핑한다. 셋째, 생성된 진동 및 AE 스펙트로그램을 수평으로 병합한 뒤, 연산 비용의 증가 없이 모델의 입력 규격을 맞추기 위해 영행렬을 채우는 Zero-padding 방식이나 패턴을 상하로 복제하는 Tiling 방식의 공간 최적화를 수행한다. 마지막으로, 전처리가 완결된 단일 텐서를 경량 신경망인 MobileNet-V3 Small에 인가하여 최종적으로 베어링의 4가지 상태(정상, 내륜 결함, 외륜 결함, 롤러 결함)를 분류한다.

3.2 Time Synchronization and Frequency Transformation of Heterogeneous Signals

서로 다른 물리적 특성을 지닌 진동 센서와 AE 센서는 각기 다른 sampling rate로 데이터가 수집되므로 신호 간

www.kci.go.kr

의 동시성 확보가 필수적이다. 본 연구에서 사용한 데이터셋의 구조에서 진동 신호는 65.5 kHz의 샘플링 비율로 5초 단위의 단일 파일로 저장되는 반면, AE 신호는 250 kHz의 샘플링 비율로 1초 단위로 저장되는 시간적 불일치를 지닌다. 두 신호 간의 물리적 발생 시점을 오차 없이 정렬하기 위해 본 연구에서는 파일 레벨과 샘플 레벨로 구성된 2단계 계층적 동기화 전처리 프로세스를 수행한다. 첫 번째 단계인 파일 레벨 동기화(file-level synchronization)에서는 5초 길이의 진동 신호 파일 1개에 대응하는 동일 측정 시점의 연속된 1초 길이 AE 신호 파일 5개를 순차적으로 찾아 선형 병합을 수행한다. 이를 통해 두 이기종 센서 신호가 가변 속도 구동축의 동일한 5.0초의 측정 시간 윈도우를 공유하도록 물리적 시간축을 병합한다. 두 번째 단계인 샘플 레벨 분할(sample-level slicing)에서는 통합 완료된 5초 길이의 1D 시계열 신호를 0.1초 크기의 윈도우 단위로 분할한다. 동기화 위상의 정밀도를 보장하기 위해 0.1초 구간에 포함되는 각 센서 고유의 주파수 대역별 샘플 개수를 정확히 계산하여, 진동 신호는 6,550개의 데이터 포인트, AE 신호는 25,000개의 샘플 데이터 포인트 인덱스 단위로 슬라이싱을 진행한다. 특히, 시계열 신호 분할 과정에서 프레임의 이동 간격인 hop size를 윈도우 크기와 일치하는 0.1초로 설정함으로써 오버랩 비율을 0%로 설정한다. 이는 연속된 프레임 간에 데이터 중복성과 인접 시간 데이터의 혼입을 차단하여, train 도메인의 노이즈 특성이 test 도메인으로 전이되는 시간적 데이터 누수 문제를 전처리 단계에서부터 방지하기 위함이다[11][12][13].

더 나아가 본 연구에서는 기존 선행 연구들의 한계점인 데이터셋 분할 오류를 차단하기 위한 데이터 격리 프로토콜을 수립한다. 기존의 대다수 딥러닝 기반 고장 진단 연구들은 수집된 시계열 신호를 윈도우 단위로 잘게 쪼갠 후, 전체 세그먼트 풀 내에서 샘플 단위로 무작위로 섞어 train과 test 데이터 셋을 구축해왔다. 이러한 무작위 분할 방식은 동일한 파일 혹은 동일한 구동 세션 내에서 파생된 인접 시간대의 스펙트로그램들이 두 데이터셋에 분산되어 섞여 들어가는 현상을 야기한다. 이 경우, 물리적으로 독립되지 않은 두 데이터셋 간의 특정 측정 시점의 고유한 배경 노이즈 및 환경적 잡음 등의 위상 특성이 고스란히 공유되는 치명적인 시간적 데이터 누수를 유발한다[11][13]. 결과적으로 딥러닝 모델은 베어링 결함 고유의 본질적인 주파수 특성을 주도적으로 학습하는 것이 아니라, 특정 파일의 배경 소음 패턴과 정답 간의 인위적인 shortcut을 기억하고 암기하게 됨으로써, 실제 일반화 성능과 무관하게 지표상의 테스트 정확도가 비정상적으로 부풀려지는 방법론적 오류를 낳는다. 이러한 신뢰성 왜곡 문제를 근본적으로 해결하기 위해 본 논문에서는 세션 및 파일 독립적 데이터 분할 (session-independent dataset partitioning) 기법을 도입한다. 제안 기법의 데이터셋 구축 방식은 단일 5.0초 동기화 파일 그룹에서 생성된 2D 스펙트로그램 이미지 전체를 오직 train 데이터셋에만 포함시키거나, 혹은 전량 test 데이터셋에만 할당하는 원칙을 지킨다. 즉, 임의의 한 파일에서 추출된 데이터 포인트 파편들이 분할 계층을 가로질러 교차 혼입되는 것을 방지한다. 이를 통해 모델은 테스트 과정에서 배경 소음에 대한 암기 효과를 발휘할 수 없으며, 오직 가변 속도 도메인 하에서 동적으로 변화하는 고장 징후 대역의 순수한 주파수 스펙트럼 변조 특성만을 기반으로 일반화 유효성을 객관적으로 평가받게 된다.

동기화가 완결된 0.1초 단위의 1D 시간 영역 신호는 최종적으로 STFT를 거쳐 시간-주파수 도메인의 2D 스펙트로그램 이미지로 변환된다. 변환 과정을 통해 비정상성이 강한 신호 내에 복잡하게 혼재되어 있던 결함 정보들이 딥러닝 커널 필터가 기하학적 공간 특징 정보로 인식하고 전역적으로 인코딩할 수 있는 시각적인 시간-주파수 2D 텐서 공간 형태로 재구성된다[16].

3.3 Spatial Optimization-Based Input-Level Spectrogram Fusion

독립적인 STFT 변환 과정을 통해 생성된 진동 및 AE 스펙트로그램의 특징 공간을 결합하기 위해 제안 기법은 신경망 모델의 파라미터나 레이어 구조를 확장하지 않고 연산 비용 증가율을 0%로 유지하는 특징 융합 기법을 제안한다. [Figure 2]는 제안하는 Input-level fusion 기법의 텐서 차원 전이 및 공간 여백 제어 메커니즘을 보이고 있다. 융합 파이프라인의 전반부에서는 독립적으로 생성된 진동 스펙트로그램 특징 텐서($H \times W_1 \times C$)와 AE 스펙트로그램 특징 텐서($H \times W_2 \times C$)를 시간 축(Horizontal axis)의 일관성을 확보한 채 수평 병합(Horizontal Concatenation)하여 $H \times (W_1 + W_2) \times C$ 차원을 지닌 단일 중간 융합 텐서를 생성한다.

본 연구에서는 경량성과 필터 추출 효율을 위해 각 단일 텐서 스펙트로그램의 해상도를 112x112 픽셀 크기로 규격화하므로, 수평 병합을 통해 완성된 중간 융합 텐서는 112x224의 비대칭 해상도를 지니게 된다. 이후 표준 합성곱 신경망 백본의 입력 해상도인 224x224 규격을 충족시키기 위해, 수직축 주파수 도메인 상에 남는 112 픽셀 공

www.kci.go.kr

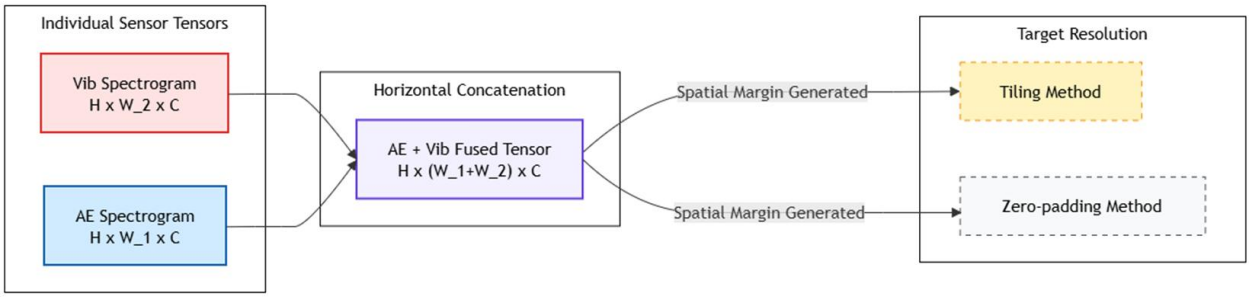
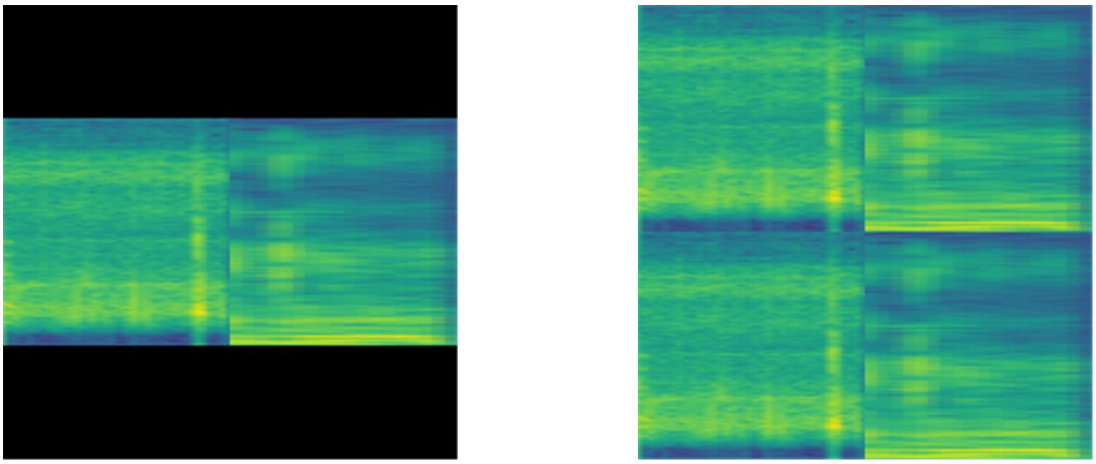


Fig. 2. Proposed spectrogram fusion method



(a) Zero-padding method (b) Tiling method

Fig. 3. Results of actual tensor transformation (input images)

간을 제어하는 두 가지 공간 최적화 특징 증강 기법을 분리하여 제안한다. 입력 단계에서 생성되는 실제 데이터 텐서의 구조적 변환 양상은 [Figure 3]과 같이 정의된다. 첫 번째 기법인 Zero-padding 방식([Figure 3(a)])은 수평 병합된 112x224 특징 텐서 상·하단에 남는 수직축 112 픽셀 여백 공간의 모든 원소 값을 노이즈가 없는 영행렬(Zero-matrix)로 채워 넣는 기법이다. 이 방식은 이기종 센서가 병합된 스펙트로그램 고유의 시간-주파수 물리적 척도와 픽셀 비율을 원본 그대로 왜곡 없이 보존할 수 있으며, 백본 신경망 필터가 불필요하게 가중치를 분산시키지 않고 여백 내에 존재하는 핵심 고장 주파수 선형 패턴에만 집중할 수 있도록 유도하는 공간 격리 이점을 제공한다. 두 번째 기법인 Tiling 방식([Figure 3(b)])은 수평 병합을 통해 완성된 112x224 해상도의 중간 다중 센서 융합 텐서 패턴 자체를 수직축 여백 공간에 상하로 복제하여 224x224 규격을 맞추는 기법이다. 이는 컴퓨터 비전 도메인의 인스턴스 분할 영역에서 객체 식별 일반화 성능을 높인 복사-붙여넣기(Copy-Paste) 메커니즘을 시계열 시간-주파수 도메인 특징 공간에 차용한 접근법이다[17][18][19]. 이 기법은 모델 추론 단계에서 파라미터나 부동소수점 연산량의 오버헤드를 유발하지 않으면서도, 합성곱 신경망 필터의 수용 영역 내에 유효 결함 주파수 성분을 수직축 상으로 중복해서 드러냄으로써 특징 공간상에서의 이동 불변성(translation invariance)을 극대화할 수 있다.

3.4 Lightweight Diagnostic Network Design Considering Computational Efficiency

본 연구는 연산 능력과 메모리 버퍼 자원이 제한된 실제 공장의 엣지 디바이스 환경에서도 구동되는 프레임워크를 수립하는 데 목적이 있으므로 네트워크의 내부 레이어 깊이와 하드웨어 친화적인 블록 구조를 유기적으로 배치하여, 총 파라미터 수 및 연산량 대비 특징 추출 효율을 향상시킨 경량 CNN 모델인 MobileNet-V3 Small 아키텍처를 백본 네트워크로 채택한다[9]. 제안하는 전체 시스템 아키텍처는 3.3절에서 기술한 전처리 입력 전반의 특징 공간 최적화를 통해 이기종 센서의 다차원 주파수 정보가 변환된 단일 2D 텐서(224x224x3)만을 백본에 인가한다. 결과적으로 신경망 모델 내부에서 복잡하게 연산 자원을 소모하며 다중 채널 분기를 확장하거나 레이어 중간 단계에서

www.kci.go.kr

파라미터를 희생하며 특징을 합칠 필요성을 배제한다. 다중 이기종 센서의 강점을 모두 수용하면서 MobileNet-V3 Small 경량 구조 사양과 부동소수점 연산량의 효율성을 모델 설계 단계에서부터 활용할 수 있어, 연산 자원 최적화와 결함 분류 극대화의 trade-off를 해결한 연산 효율적 아키텍처를 설계할 수 있다.

IV. Experimental Design and Result Analysis

4.1 Experimental Dataset and Cross-Speed Domain Generalization Scenarios

본 연구에서 제안하는 Multi-sensor Input-level fusion 기법의 실효성을 확인하기 위해, 3상 유도전기 모터 구동 환경에서 수집된 베어링 진동 및 AE 데이터를 활용하였다. 실험 대상이 되는 고장 클래스는 정상(Normal), 내륜 결함(Inner), 외륜 결함(Outer) 그리고 롤러 결함(Roller) 총 4가지로 분류하였으며, 결함 크기 12 mm 데이터를 사용하였다. 전체 데이터셋은 6가지 회전 속도 조건(250, 300, 350, 400, 450, 500 RPM)과 4가지 고장 클래스의 조합으로 이루어진 총 24개의 독립적인 구동 세션(operating session)으로 구성된다.

각 구동 세션(단일 RPM 및 단일 클래스) 내에서 진동 데이터는 65.5 kHz의 샘플링 비율로 5.0초 단위의 raw 파일 총 94개로 수집되었으며, AE 데이터는 250 kHz의 샘플링 비율로 1.0초 단위의 raw 파일 총 470개로 수집되었다. 이기종 센서 간의 물리적 위상을 일치시키기 위해, 연속 수집된 1.0초 단위의 AE raw 파일 5개를 순차 결합하여 5.0초 단위의 진동 raw 파일 1개와 매칭함으로써 세션당 94개의 5.0초 동기화 파일 그룹을 구축하였다. 이후 동기화된 5.0초 시계열 신호를 2D 스펙트로그램으로 변환하기 위해 0.1초 크기의 윈도우 분할(진동 6,550 samples, AE 25,000 samples)을 수행하였다. 세그먼트 간 분할의 hop size를 세그먼트 길이와 동일한 0.1초로 설정하여 세그먼트 간 오버랩 비율을 0%로 하였다. 이에 따라 1개의 5.0초 동기화 파일 그룹(진동 파일 1개 및 AE 파일 5개)에서는 인접 샘플과 신호를 공유하지 않는 50개의 비중첩(non-overlapping) 0.1초 세그먼트가 생성되며, 각 구동 세션당 총 4,700장의 샘플이 균일하게 구축되었다.

분할된 각각의 0.1초 세그먼트는 STFT를 통해 2D 스펙트로그램으로 변환된다. 여기서 원신호를 0.1초 단위로 나눌 때 적용한 세그먼트 간 0% 오버랩과, 개별 0.1초 세그먼트 내부에서 시간-주파수 스펙트럼을 계산할 때 적용하는 STFT 내부 창 오버랩은 엄격히 구분된다. 세그먼트 간 분할은 샘플 간 데이터 누수 방지를 위해 비중첩으로 수행된 반면, 이미 격리된 각 0.1초 세그먼트 내부에서는 스펙트럼 누설 억제와 시간축 연속성 확보를 위해 길이 1,024 samples(nperseg=1024)의 Hann window를 25% 오버랩(noverlap=256, hop length 768 samples) 조건으로 이동시키며 1,024-point FFT(nfft=1024)를 수행하였다. STFT 연산은 단일 0.1초 세그먼트 경계 내부에서만 수행되므로 세그먼트 간 정보 혼입을 유발하지 않는다. 산출된 복소 스펙트럼은 선형 진폭 추출 후 미세 결함 에너지 보존을 위해 $10\log_{10}(|Z_{xx}|+10^{-10})$을 적용하여 데시벨(dB) 스케일로 로그 압축되었다. 이후 로그 진폭 행렬은 Gouraud 보간(Piecewise bilinear shading) 기반 렌더링과 양선형 보간(Bilinear interpolation)을 거쳐 축과 여백이 제거된 112x112 규격의 3채널(RGB) 스펙트로그램으로 생성된다. 최종적으로 제안 기법에서는 각각 112x112x3으로 생성된 진동 및 AE의 RGB 스펙트로그램이 3.3절의 수평 병합(112x224x3)과 공간 확장(Zero-padding 또는 Tiling)을 통해 224x224x3 RGB 텐서로 구축되며, 모델 입력 시에는 패딩 영역의 0 값 보존과 원본 스펙트럼 대비 유지를 위해 별도의 평균·표준편차 표준화 없이 8-bit 픽셀값[0, 255]을 [0.0, 1.0] 구간으로 선형 정규화(min-max rescaling)하였다.

딥러닝 모델이 특정 회전 속도에 편향되는 것을 방지하고 시간적 데이터 누수로 인한 성능 과대평가를 차단하기 위해, 본 연구에서는 구동 세션 독립성(session-independence)과 파일 독립성(file-independence)을 계층적으로 결합한 미학습 가변 속도 도메인 일반화 평가(cross-speed domain generalization evaluation) 시나리오를 설계하였다. 첫째, 세션 독립적 분할(session-independent partitioning)은 학습/검증 도메인과 테스트 도메인 간의 물리적 구동 환경 격리를 의미한다. 전체 24개 구동 세션 중 250, 350, 450 RPM에서 수집된 12개 구동 세션(56,400장)은 train 및 validation 풀로 할당하고, 모델이 학습 과정에서 단 한 번도 관측하지 못한 300, 400, 500 RPM의 12개 구동 세션(56,400장)은 전량 test 데이터셋으로 격리하였다. 둘째, 파일 독립적 분할(file-independent partitioning)은 동일한 회전 속도 세션을 공유하는 train 및 validation 데이터셋 내부에서 인접 시간대 세그먼트 간 배경 노이즈가 공유되는 현상을

www.kci.go.kr

Table 2. Summary of dataset partitioning across variable speed domains

Partition	Speed (RPM)	Sessions	Sync. File Groups (5.0s)	Spectrograms (0.1s)
Train	250, 350, 450	12	1,016 (~90.1%)	50,800
Validation	250, 350, 450	12	112 (~9.9%)	5,600
Test	300, 400, 500	12	1,128 (100%)	56,400
Total	6 RPMs	24	2,256	112,800

방지하기 위한 분할 원칙이다. 무작위로 개별 0.1초 스펙트로그램을 섞는 기존 방식을 배제하고, 5.0초 동기화 파일 그룹의 고유 식별자를 최소 분할 단위로 설정하여 고정 난수 시드(Seed=42) 기반의 9:1 분할을 수행하였다. 단일 5.0초 동기화 파일 그룹에서 파생된 50장의 스펙트로그램 세그먼트 전체는 쪼개지지 않고 오직 train 또는 validation 중 하나의 파티션에만 일괄 배정된다. 결과적으로 train/validation 풀의 전체 1,128개 동기화 파일 그룹 중 1,016개 그룹(50,800장)은 train 데이터셋에, 독립된 112개 파일 그룹(5,600장)은 validation 데이터셋에 배정되었으며, 전체 데이터셋의 분할 구조는 [Table 2]에 보이는 바와 같다.

4.2 Model Training Environment

본 실험은 이미지넷 사전 학습된 가중치의 영향을 배제하고 베어링 고유의 시간-주파수 특징만을 학습시키기 위해 MobileNet-V3 Small 네트워크를 처음부터 학습시켰다. 네트워크 가중치 최적화를 위해 L2 정규화(weight decay) 특성이 우수한 AdamW 옵티마이저를 채택하였으며, 배치 사이즈(batch size)는 32, 총 학습 epoch은 50으로 설정하였다. 이때 학습률과 총 학습 epoch을 포함한 모든 학습 조건은 미학습 test 데이터셋(300, 400, 500 RPM)을 전혀 참조하지 않고, 오직 학습 도메인(250, 350, 450 RPM) 내에서 파일 단위로 독립 분할된 train 데이터셋(50,800장)과 validation 데이터셋(5,600장)만을 이용한 예비 실험을 통해 사전 확정하였다. 본 연구의 train 데이터셋(50,800장)과 배치 사이즈(32) 조건에서는 1 epoch당 1,588회의 미니배치 가중치 업데이트가 수행되어 50 epoch 동안 총 79,400회의 반복 최적화(iteration)가 이루어진다. 최적의 학습률(learning rate) 선정을 위해 train 및 validation 데이터셋만을 대상으로 1e-4부터 1e-7까지의 로그 스케일 탐색 범위를 비교한 결과, 1e-4와 같은 일반적인 학습률을 적용할 경우 1 epoch 내의 1,588회 업데이트만으로도 학습 및 검증 정확도가 95% 이상으로 급격히 치솟는 조기 과포화(premature saturation) 현상이 발생하였다. 이러한 급격한 초기 수렴은 모델이 학습 도메인의 지엽적 특징에 과적합(overfitting)된 것인지 혹은 불완전하게 수렴된(underfitting) 상태인지 validation 지표만으로 판별하기 어렵게 만들 뿐만 아니라, epoch 진행에 따른 공간 최적화 기법 간 검증 손실 거동과 후기 진동(oscillation) 여부를 정밀하게 확인하기 어려운 문제를 발생시킨다. 따라서 초기 기울기 급변과 조기 포화를 방지하고 점진적이고 안정적으로 수렴이 이루어지도록 학습률을 2e-7로 확정하였으며, 총 학습 epoch을 train loss와 validation loss가 모두 발산이나 과적합 징후 없이 평탄한 수렴 구간에 도달하는 50 epoch으로 고정하였다. 각 실험 조건당 총 5회의 독립적인 실험을 수행하였다. 이때 dataloader의 무작위 시드를 42로 고정하여 데이터가 모델에 입력되는 순서를 5회 모두 동일하게 설정하였다. 반면, 네트워크 초기 가중치는 고정하지 않고 무작위성을 부여하였다. 이는 데이터 입력 순서라는 외부 환경이 동일하게 통제된 조건 속에서, 모델이 불리한 무작위 가중치에서 출발하더라도 제안하는 융합 특징을 통해 안정적으로 높은 진단 성능에 도달함을 확인하기 위함이다.

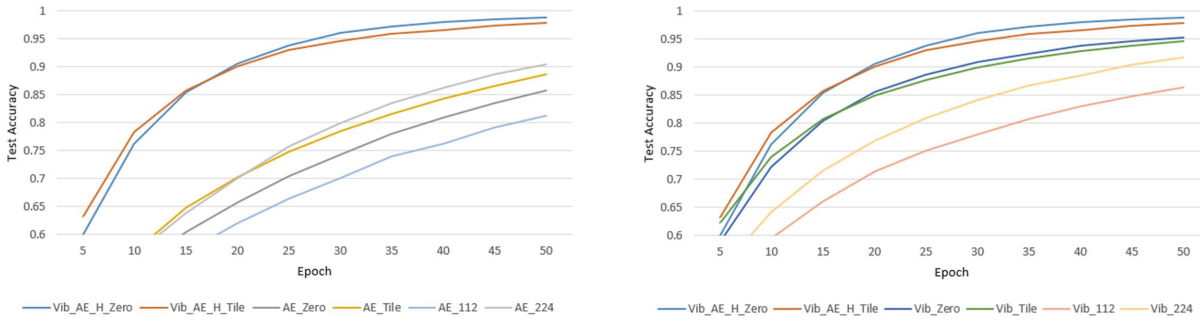
평가 프로토콜의 독립성을 보장하기 위해, 본 연구에서는 조기 종료(early stopping)를 사용하지 않고, 11개 전체 실험 조건과 5회 반복 시행 모두에서 위와 같이 train 및 validation 수렴 기준으로 최종 50 epoch 시점의 모델을 최종 평가 모델로 동일하게 채택하였다. 본 학습이 진행되는 동안 미학습 test 데이터셋은 사용되지 않은 채 매 epoch의 가중치만 저장되었으며, 50 epoch 학습이 모두 완료된 이후에야 저장된 체크포인트들을 대상으로 일괄 사후 평가를 수행하였다.

제안하는 Multi-sensor Input-level fusion과 공간 최적화 기법(Tiling / Zero-padding)의 실효성을 단계적으로 검증하기 위해 대조군을 포함한 총 11가지의 실험 조건별 구체적인 특징 공간 조건은 [Table 3]에서 보이는 바와 같다.

www.kci.go.kr

Table 3. 11 input data experimental conditions for evaluating the proposed method

Group	Condition	Sensor Modality	Spatial Optimization Method	Input Resolution
Information-Controlled	AE_112	AE	None	112x112
	Vib_112	Vib	None	112x112
FLOPs-Controlled	AE_224	AE	None	224x224
	Vib_224	Vib	None	224x224
Feature-Level Baseline	Vib_AE_Feature_level	Vib+AE	None	224x224
Single-Sensor Extension	AE_Zero-padding	AE	Zero-padding	224x224
	AE_Tile	AE	Tiling	224x224
	Vib_Zero-padding	Vib	Zero-padding	224x224
	Vib_Tile	Vib	Tiling	224x224
Multi-Sensor Fusion (Proposed)	Vib_AE_H_Tile	Vib+AE	Tiling	224x224
	Vib_AE_H_Zero-padding	Vib+AE	Zero-padding	224x224



(a) AE-based Baselines vs. Proposed Fusion (b) Vib-based Baselines vs. Proposed Fusion

Fig. 4. Test accuracy curves over epochs

4.3 Comparative Analysis of Fault Diagnosis Performance

[Figure 4]는 50 epoch 학습 종료 후 저장된 체크포인트들을 사후 평가하여 도출한 실험 조건별 변화 추이를 음향방출(AE) 기반 대조군([Figure 4(a)])과 진동(Vib) 기반 대조군([Figure 4(b)])으로 나누어 시각화한 결과이며, [Table 4]는 각 조건의 연산량(GFLOPs)과 학습 초기(10 epoch), 중기(30 epoch), 말기(50 epoch)의 주요 마일스톤 단계에서 산출된 5회 반복 실험의 분류 정확도(평균 ± 표준편차)를 상세 정리한 결과이다.

정확도 분석에 앞서, 제안 기법의 목적 중 하나인 연산 효율성과 제한된 엣지 하드웨어 환경에서의 구동 적합성을 확인하기 위해 제안된 2개의 기법들(Vib_AE_H_Zero-padding, Vib_AE_H_Tile)과 동일한 입력 규격(224x224)을 갖는 단일 센서 비교군(AE_224, Vib_224) 및 특징단 융합 비교군(Vib_AE_Feature_level)을 대상으로 모델 파라미터 수(Parameters), 연산량(GFLOPs), 피크 추론 메모리 점유율(Peak Inference Memory), 그리고 데스크톱 CPU 환경(Windows 11 AMD64)과 실제 ARM64 기반 임베디드 엣지 디바이스(Linux aarch64, Raspberry Pi 5 BCM2712, 단일 코어 CPU_Threads=1, Batch_Size=1)에서의 신경망 순전파 추론 지연 시간(forward-pass inference latency)을 측정하였다. 이때 지연 시간은 모델 구조에 따른 순수 연산 오버헤드를 비교하기 위해 STFT 변환, 스펙트로그램 렌더링 및 텐서 병합 등의 전처리 과정을 제외하고, 메모리에 적재된 224x224x3 입력 텐서에 대한 신경망 순전파 연산 구간만을 측정하였다. MobileNet-V3 Small 단일 백본을 사용하는 단일 센서 비교군(AE_224 최종 정확도 90.39 ± 1.67%, Vib_224 91.17 ± 2.05%)은 1.52M(1,521,956개)의 파라미터와 0.055515 GFLOPs의 연산량, 그리고 7.77 MB(모델 가중치 5.85 MB, 입력 텐서 0.57 MB, 피크 활성화 메모리 1.34 MB)의 피크 메모리를 요구하며, 엣지 디바이스 상에서 각각 18.19 ± 0.12 ms(데스크톱 6.84 ± 0.20 ms)와 18.18 ± 0.07 ms(데스크톱 6.88 ± 0.17 ms)의 추론 지연 시간을 기록하였다. 반면, 이기종 센서의 특징 추출을 위해 두 개의 백본 분기를 병렬로 구성한 특징단 융합 비교군(Vib_AE_Feature_level, 최종 정확도 97.77 ± 0.15%)은 파라미터 수(3.04M, 3,043,908개)와 연산량(0.111039 GFLOPs)이 단일 백본 대비 약 2.00배로 증가하고 피크 메모리 사용량이 14.19 MB로 약 1.83배 증가하였으며, 이로 인해 실제 엣지 디바이스에서의 순전파 지연 시

www.kci.go.kr

Table 4. Comparison of test accuracy in the variable-speed domain according to training progress (average of 5 runs)

Experimental Condition	10 Epochs Acc. (%)	30 Epochs Acc. (%)	50 Epochs Acc. (%)	Computational Cost (GFLOPs)
AE_112	50.26 ± 2.26	70.11 ± 2.62	81.26 ± 1.82	0.015943
Vib_112	59.47 ± 2.10	78.02 ± 1.52	86.43 ± 0.56	0.015943
AE_224	56.49 ± 5.28	80.01 ± 2.50	90.39 ± 1.67	0.055515
Vib_224	64.22 ± 4.58	84.18 ± 3.66	91.17 ± 2.05	0.055515
AE_Zero-padding	53.32 ± 2.64	74.32 ± 3.72	85.73 ± 3.08	0.055515
AE_Tile	56.93 ± 2.65	78.52 ± 2.48	88.69 ± 2.66	0.055515
Vib_Zero-padding	72.29 ± 7.74	90.86 ± 1.94	95.37 ± 0.71	0.055515
Vib_Tile	74.01 ± 5.60	89.92 ± 2.71	94.61 ± 1.40	0.055515
Vib_AE_H_Tile	78.30 ± 4.17	94.59 ± 1.26	97.82 ± 0.52	0.055515
Vib_AE_H_Zero-padding	76.21 ± 6.67	96.06 ± 0.60	98.77 ± 0.13	0.055515
Vib_AE_Feature_level	66.45 ± 4.52	93.09 ± 0.72	97.77 ± 0.15	0.111039

간 또한 36.26 ± 0.38 ms(데스크톱 13.50 ± 0.26 ms)로 약 2배 증가하는 연산 병목 현상을 보였다. 이와 달리 본 연구에서 제안하는 기법인 Vib_AE_H_Zero-padding(최종 정확도 98.77 ± 0.13%)과 Vib_AE_H_Tile(97.82 ± 0.52%)은 전처리 단계에서 두 센서의 112x112 스펙트로그램을 단일 224x224 텐서로 병합하여 인가하므로, 단일 센서 비교군(AE_224, Vib_224)과 완전히 동일한 1.52M 파라미터, 0.055515 GFLOPs 연산량, 7.77 MB 피크 메모리만을 소모한다. 나아가 엣지 디바이스 상에서의 단일 샘플 신경망 순전파 지연 시간 역시 Vib_AE_H_Zero-padding이 18.13 ± 0.10 ms(데스크톱 6.69 ± 0.02 ms), Vib_AE_H_Tile이 18.11 ± 0.13 ms(데스크톱 6.66 ± 0.02 ms)로 측정되어 추가 지연 오버헤드 없이 단일 센서 모델과 동일한 수준의 순전파 속도를 유지하였으며, 특징단 융합 비교군 대비 절반의 순전파 시간과 메모리 자원만으로도 더 높은 분류 정확도를 기록하였다. 비록 본 측정에는 STFT 및 스펙트로그램 생성 전처리 시간이 포함되지 않아 종단간 실시간 처리 성능을 직접 단정할 수는 없으나, 0.1초(100 ms) 신호 윈도우 주기 중 신경망 추론이 차지하는 비중을 약 18.1% 수준으로 억제하여 특징단 융합(약 36.3%) 대비 전처리 연산을 위한 시간적 여유를 충분하게 확보함으로써 제한된 엣지 하드웨어에 유리한 모델 차원의 연산 경량성을 확인하였다.

epoch별 테스트 정확도 시각화 곡선인 [Figure 4]와 주요 시점별 추적 데이터인 [Table 4]를 살펴보면, 단일 도메인 모델의 본질적인 정보량 한계점과 제안하는 이기종 Input-level fusion 기법의 일반화 역학 메커니즘을 다음과 같이 확인할 수 있다.

학습 초기 단계인 10 epoch 시점에서 정보량이 통제된 해상도의 단일 센서 기준군인 [Figure 4(a)]의 AE_112와 [Figure 4(b)]의 Vib_112는 각각 50.26 ± 2.26%와 59.47 ± 2.10%라는 저조한 일반화 정확도를 기록하며 미학습 속도 도메인 적응에 지연을 겪고 있다. 연산량이 통제된 224x224 기준군인 [Figure 4(a)]의 AE_224와 [Figure 4(b)]의 Vib_224 역시 10 epoch 시점에 각각 56.49 ± 5.28%, 64.22 ± 4.58%에 머물렀다. 초기 가중치 조건에 따라 ± 4.58~5.28%에 달하는 큰 표준편차를 보였다. 또한, 최종 50 epoch 결과에서도 AE_224와 Vib_224는 각각 90.39 ± 1.67%와 91.17 ± 2.05%에 그치며 90~91% 대의 성능 한계에 부딪히게 된다. 이는 회전 속도가 수시로 변동하는 가혹한 구동 환경하에서 단일 물리 도메인의 정보만으로는 합성곱 필터가 결함 특징의 비선형적인 도메인 이동(domain shift)을 안정적으로 극복하지 못하고 외부 잡음에 쉽게 간섭받아 오분류를 야기함을 알 수 있다[11]. 반면, 본 연구에서 제안하는 Multi-sensor Input-level fusion 모델들은 [Figure 4(a)]와 [Figure 4(b)]의 최상단 곡선에서 나타나듯 학습 시작과 동시에 매우 가파른 상승 곡선을 그린다. [Table 4]의 10 epoch 구간 수치를 보면, Vib_AE_H_Tile (제안 기법 2)과 Vib_AE_H_Zero-padding (제안 기법 1)은 각각 78.30 ± 4.17%와 76.21 ± 6.67%라는 정확도를 조기에 달성하고 연산량이 약 2배에 달하는 Vib_AE_Feature_level (66.45 ± 4.52%)까지 큰 격차로 상회하는 초기 수렴을 보여주었다. 이 구간에서 주목할 점은 [Table 4]에서 수평 병합된 112x224 텐서를 수직축으로 복제하여 확장한 Vib_AE_H_Tile 기법(78.30 ± 4.17%)이 영행렬을 채운 Vib_AE_H_Zero-padding 방식(76.21 ± 6.67%)보다 평균 정확도가 약 2.09% 높을 뿐만 아니라, 표준편차 역시 약 2.50% 낮게 나타나, 학습 초기 단계에서 상대적으로 빠른 수렴 경향을 보인다는 점이다. 단일 센서 확장군에서도 10 epoch 기준 [Figure 4(a)]의 AE_Tile(56.93 ± 2.65%)이 AE_Zero-padding(53.32 ± 2.64%)보다 높은

www.kci.go.kr

Table 5. Class-wise and macro-averaged diagnostic performance metrics of the proposed Vib_AE_H_Zero-padding method on the unseen variable-speed test dataset (Mean ± Std over 5 runs)

Fault Class	Precision (%)	Recall (%)	F1-Score (%)	Support (Samples)
Normal	99.11 ± 0.22	99.67 ± 0.24	99.39 ± 0.13	14,100
Inner Fault	98.33 ± 0.67	98.24 ± 0.39	98.28 ± 0.40	14,100
Outer Fault	99.73 ± 0.19	98.80 ± 0.64	99.26 ± 0.28	14,100
Roller Fault	97.96 ± 0.77	98.40 ± 0.73	98.18 ± 0.22	14,100
Macro-Average	98.78 ± 0.13	98.78 ± 0.13	98.78 ± 0.13	56,400

정확도를 보이며, [Figure 4(b)]의 Vib_Tile(74.01 ± 5.60%) 역시 Vib_Zero-padding (72.29 ± 7.74%) 대비 높은 평균 정확도와 낮은 편차를 기록하였다. 이는 백본 신경망의 학습 가중치가 성숙하지 않은 초기 최적화 단계에서 결함 주파수 유효 패턴을 반복 노출시킴으로써, 경량 CNN 커널의 수용 영역(Receptive field) 내 특징 추출 빈도를 강제로 증폭하는 Tiling 분기 기법이 초기 특징 공간 형성에 긍정적으로 작용했을 가능성을 보인다[17][18][19].

백본 신경망의 가중치 최적화가 심화되고 내부 필터가 고도화되는 중기(30 epoch)에 접어들면서, [Figure 4(a)]와 [Figure 4(b)] 상단의 융합 기법 학습 곡선이 서로 교차하며 일반화 성능이 역전되는 현상이 관측된다. [Table 4]에 따르면 30 epoch 마일스톤 시점에서 잔여 공간을 무소음 영역으로 격리한 제안 기법 1 (Vib_AE_H_Zero-padding)은 96.06 ± 0.60%의 정확도를 기록하며, Tiling 기반의 제안 기법 2(Vib_AE_H_Tile, 94.59 ± 1.26%)보다 약 1.47% 높은 평균 정확도와 절반 이하로 감소한 표준편차를 보인다. 이후 [Figure 4]에서 제안 기법 1의 곡선은 fluctuation 없이 평탄함을 유지하며 최종 50 epoch에서 98.77 ± 0.13%로 전체 실험 조건 중 가장 높은 평균 진단 정확도와 가장 낮은 표준편차를 기록하였다. 비록 조건 간의 별도의 통계적 가설 검정을 수행하지는 않았으나, 2배의 연산 자원을 소모하는 Vib_AE_Feature_level (97.77 ± 0.15%) 및 Vib_AE_H_Tile(97.82 ± 0.52%) 대비 5회 반복 실험에서 일관되게 더 높은 평균 정확도와 좁은 편차 범위를 나타냈다. 이는 무작위 셔플로 인한 shortcut 암기 효과가 배제된 독립 파일 분할 환경에서도 제안 기법이 초기 가중치 변동에 덜 민감하게 수렴하며 안정적인 일반화 성능을 유지하는 경향이 있음을 보여준다. 이러한 학습 후반부 성능 역전의 매커니즘으로, 이기종 센서 결합을 통해 이미 결함 특징 정보가 충분히 확보된 다중 텐서 구조에서는, 미세 최적화 단계로 진입할수록 동일한 패턴을 중복 배치하는 Tiling 방식(Vib_AE_H_Tile, 97.82 ± 0.52%) 유효 신호뿐만 아니라 스펙트로그램 내 배경 잡음 성분까지 반복 노출시켜 미세 가중치 갱신 과정에서 부분적인 과적합 요인으로 작용했을 가능성을 보인다. 반면, Zero-padding 기법은 원본의 시간-주파수 물리적 척도를 변형 없이 유지하고, 잔여 공간을 영행렬 버퍼 영역으로 격리함으로써, 컨볼루션 필터가 불필요한 여백 활성화 없이 중앙의 유효 주파수 대역에 주의를 집중하는 데 유리한 환경을 제공한 것으로 해석된다. 이에 대한 보조적 시각화 분석은 4.6절에서 상세히 다룬다.

4.4 Confusion Matrix-Based Identification Capability Verification by Fault Class

제안 기법의 클래스별 분류 성능을 세부적으로 확인하기 위해, 전체 실험 조건 중 가장 우수한 도메인 일반화 성능을 달성한 Vib_AE_H_Zero-padding 조건의 최종 미학습 가변 속도 test 데이터셋(300, 400, 500 RPM, 클래스당 14,100장, 총 56,400장)에 대한 혼동 행렬(confusion matrix)을 [Figure 5]에 보이고 있다. 정규화된 비율만으로는 직관적으로 파악하기 어려운 실제 오류 규모와 클래스별 정량적 식별 성능을 종합적으로 검증하기 위해 5회 독립 반복 실험의 혼동 행렬로부터 산출된 클래스별 정밀도(Precision), 재현율(Recall), F1-점수(F1-score), 매크로 평균(Macro-average) 및 평가 샘플 수(Support)를 [Table 5]에 제시하였다.

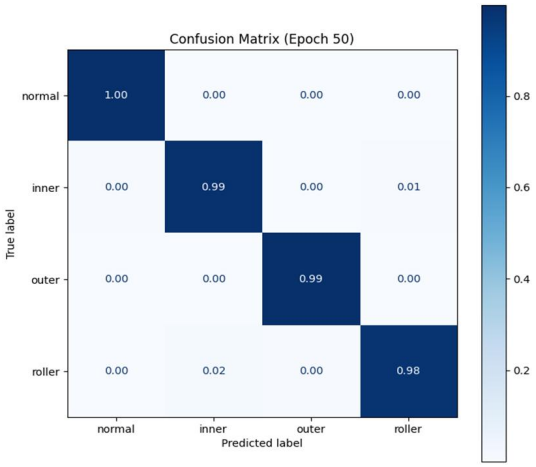


Fig. 5. Confusion matrix at Epoch 50

www.kci.go.kr

[Figure 5]의 혼동 행렬과 [Table 5]의 지표를 종합하여 분석한 결과, 주 대각선에 대다수의 데이터가 밀집하여 4가지 고장 클래스 전반에 걸쳐 편향 없이 높은 분류 성능을 제공함을 알 수 있다. 5회 반복 실험의 실제 샘플 수(sample count) 기준으로 살펴보면, 각 클래스당 14,100개의 미학습 속도 테스트 샘플 중 정상 상태(Normal)는 평균 14,053.4개(평균 오분류 46.6개), 내륜 결함(Inner fault)은 평균 13,851.6개(평균 오분류 248.4개), 외륜 결함(Outer fault)은 평균 13,931.0개(평균 오분류 169.0개), 롤러 결함(Roller fault)은 평균 13,874.4개(평균 오분류 225.6개)가 정확히 분류되어, 전체 56,400장의 테스트 데이터 중 평균 55,710.4장(평균 총 오분류 689.6장)을 올바르게 판별하였다. 이를 클래스별 평가지표로 살펴보면, 정상 상태에서는 정밀도 99.11 ± 0.22%, 재현율 99.67 ± 0.24%, F1-점수 99.39 ± 0.13%를 달성하여 정상 구동 데이터를 고장으로 오경보하는 비율이 극히 낮음을 확인하였으며, 외륜 결함과 내륜 결함에서도 각각 99.26 ± 0.28%와 98.28 ± 0.40%의 높은 F1-점수를 기록하였다. 이에 따라 4개 클래스 전체의 종합적인 분류 성능을 나타내는 Macro-Precision, Macro-Recall, Macro-F1 모두 98.78 ± 0.13%를 기록하여, 반복적인 가중치 초기화 조건에서도 낮은 편차와 클래스 간 균형 잡힌 식별 성능을 일관되게 유지함을 확인하였다. 여기서 주목할 점은 베어링 진단 영역에서 가장 식별 난이도가 높은 것으로 알려진 롤러 결함에 대한 분류 강건성이다.

일반적으로 롤러 결함은 구동 중 롤러 요소(rolling element)의 자전과 공전, 그리고 하중 구간 진입 및 이탈에 따른 미끄러짐 현상이 복합적으로 작용하여, 가변 속도 환경에서는 주파수 왜곡과 smearing이 가장 극심하게 발생하는 특성을 지닌다[1][14]. 이러한 고유의 비선형적 기구학 특성으로 인해 기존의 단일 센서 기반 진단 모델들은 롤러 결함의 특징을 타 결함이나 배경 노이즈로 쉽게 오분류하는 한계를 보였다[4]. 그러나 본 연구에서 제안하는 Multi-sensor 융합 기법은 진동 센서가 포착하는 기계적 구조 변위와 AE 센서가 포착하는 고주파 에너지를 단일 텐서 내에서 교차 도메인(cross-domain) 특징으로 보완 학습함으로써, [Table 5]에서 확인되듯 주파수 왜곡이 심한 롤러 결함 클래스에 대해서도 정밀도 97.96 ± 0.77%, 재현율 98.40 ± 0.73%, F1-점수 98.18 ± 0.22%의 높은 식별 성능을 유지하며 오분류를 효과적으로 억제할 수 있음이 확인된다.

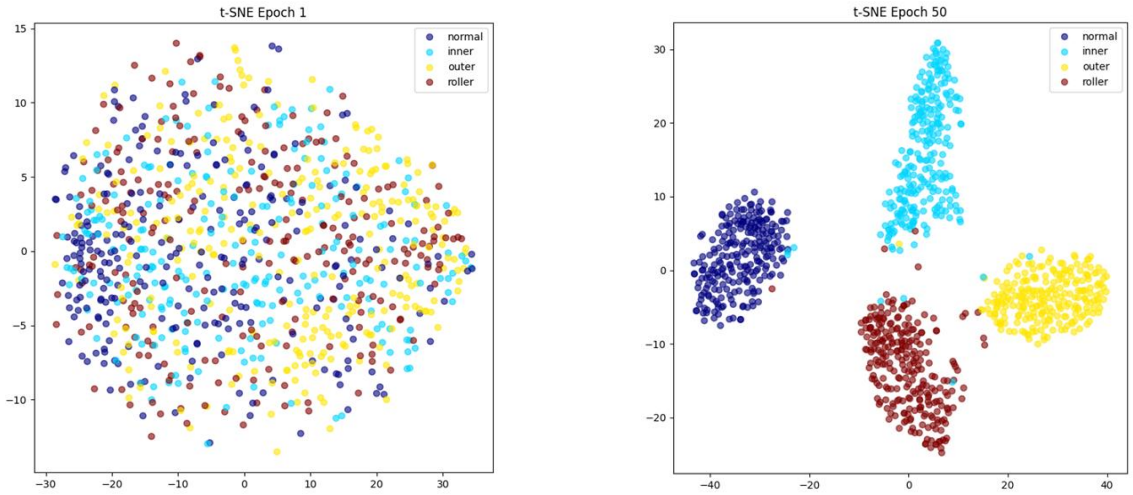
4.5 Visualization of Feature Space Cluster Separability Using t-SNE Dimensionality Reduction

제안 기법이 이기종 다중 센서 스펙트로그램으로부터 추출한 특징 벡터의 클래스 간 특징 분리도(feature separability)를 시각적으로 확인하기 위해, 고차원 특징 벡터를 저차원 공간으로 투영하는 t-SNE 기법을 활용한 보조적 분석을 수행하였다[20]. 딥러닝 기반 고장 진단 연구에서 최종 분류기 직전 레이어의 특징 분포를 2차원 t-SNE 공간상에 투영하여 학습 진행에 따른 군집 형성 양상을 관찰하는 방식이 널리 활용되고 있다[6].

본 연구에서는 가장 우수한 진단 성능을 기록한 제안 기법 1(Vib_AE_H_Zero-padding)을 대상으로, 마지막 컨볼루션 블록의 출력 특징 맵이 전역 평균 풀링(Global Average Pooling)과 헤드 투영 계층(conv_head)을 통과한 직후이자 최종 4-클래스 선형 분류 계층(final linear classifier) 직전 단계인 마지막 은닉(penultimate layer, pre_logits) 계층에서 1,024차원 특징 벡터를 추출하였다. 시각화의 재현성과 가독성을 확보하기 위해 미학습 가변 속도(300, 400, 500 RPM) 테스트 데이터셋에서 난수 시드(Seed=42)를 기준으로 무작위 비복원 추출된 1,000개의 샘플을 사용하였으며, t-SNE 하이퍼파라미터는 투영 차원 n_components = 2, 초기화 방식 init = 'pca', 혼란도 perplexity = 30, 학습률 learning_rate = 'auto', 반복 횟수 n_iter = 1,000, 난수 시드 random_state = 42로 설정하였다. [Figure 6]의 (a)와 (b)는 동일한 설정 하에서 도출된 학습 초기(1 epoch)와 학습 완료 시점(50 epoch)의 2차원 특징 공간 투영 결과를 비교하여 보여준다.

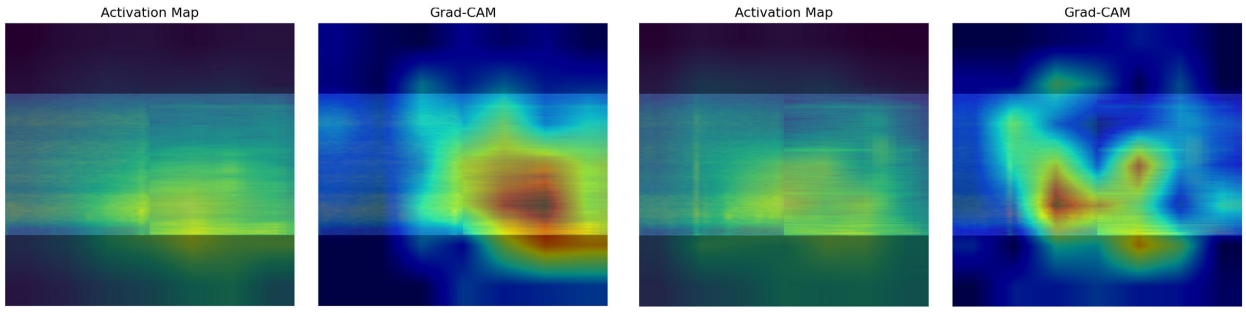
학습 초기 시점인 [Figure 6(a)]를 보면, 정상 상태를 포함한 총 4가지 클래스의 데이터 포인트들이 2차원 투영 공간 중심부에 무질서하게 혼재되어 분리되지 않는 분포를 나타낸다. 이는 가중치 최적화가 충분히 이루어지지 않은 학습 초기 단계에서 300, 400, 500 RPM의 다양한 회전 속도 조건을 가진 테스트 샘플들이 높은 클래스 내 분산(intra-class variance)을 보이며 클래스 간 경계를 형성하지 못했음을 보여준다. 그러나 50 epoch의 학습이 완료된 시점인 [Figure 6(b)]에서는 미학습 가변 속도 조건의 테스트 데이터가 혼합되어 있음에도 불구하고, 4개의 클래스가 각각 독립적인 군집(Cluster)을 형성하며 분리된 모습을 보여준다. 특히 학습 완료 시점의 분포는 초기 epoch 대비 향상된 군집 내 응집도(intra-class cohesion)와 군집 간 경계 여백이 확보되는 군집 간 분리도(inter-class separability)를 나타낸다. 비록 epoch 1

www.kci.go.kr



(a) t-SNE at Epoch 1 (b) t-SNE at Epoch 50

Fig. 6. t-SNE visualization of feature vectors at different training stages using the proposed fusion method



(a) normal condition (b) roller fault

Fig. 7. Visualization of Activation Maps and Grad-CAM heatmaps of proposed Vib_AE_H_Zero-padding method

과 epoch 50에서의 2차원 시각적 군집 형성 결과만으로 완전한 도메인 불변 특징(domain-invariant feature)의 입증을 단정할 수는 없으나, 이러한 분포 변화는 제안하는 공간 최적화 기반 입력단 융합 파이프라인이 미학습 속도 조건에서도 각 고장 유형을 구분할 수 있는 판별력 있는 특징 분리도를 효과적으로 학습하였음을 보여주는 보조적 시각 근거를 제공한다.

4.6 Visual Interpretation of Diagnostic Decision-Making via Activation Map and Grad-CAM

제안하는 공간 최적화 기반 입력단 융합 모델이 상태 판별 시 주목하는 시간-주파수 영역을 살펴보기 위해, 최종 컨볼루션 레이어 출력에 대한 활성화 맵(activation map)과 클래스별 기울기 가중치 기반 활성화 맵인 Grad-CAM (Gradient-weighted Class Activation Mapping) 시각화를 수행하였다. [Figure 7]의 (a)와 (b)는 최고 일반화 성능을 기록한 제안 기법 1(Vib_AE_H_Zero-padding) 모델에 미학습 가변 속도 테스트 데이터를 인가했을 때 도출된 정상 상태와 롤러 결함의 활성화 분포를 비교하여 보여준다. 각 이미지의 좌측은 AE 스펙트로그램, 우측은 진동 스펙트로그램에 해당한다.

[Figure 7(a)]와 [Figure 7(b)]의 activation map 및 Grad-CAM 분포를 살펴보면, 모델의 주요 활성화 에너지는 중앙의 유효 스펙트로그램 영역 내부에 집중되어 있으며 상·하단의 Zero-padding 여백 구간에서는 활성화가 대체로 억제(짙은 보라색 및 파란색)되어 나타난다. 이러한 분포는 앞선 4.3절에서 분석한 바와 같이, 영행렬 버퍼 영역이 불필요한 여백 활성화를 줄이고 중앙의 유효 주파수 대역에 주목하는 데 유리한 환경을 제공했을 것이라는 해석과 부합하는 정성적 경향을 보여준다. Activation map에서는 정상 상태와 롤러 결함 모두 신호 에너지가 높은 중앙 하단 경계부를 따라 완만한 노란색 활성화가 넓게 형성되어 분류기 직전 단계에서의 전반적인 특징 분포를 보여준다. 반면 Grad-CAM에서는 목표 클래스의 예측 점수에 대해 상대적으로 높은 기울기 민감도를 보이는 시간-

www.kci.go.kr

주파수 대역이 붉은색으로 강조되어 나타난다.

이러한 Grad-CAM의 활성화 분포를 비교하면, 입력 신호의 특성에 따라 모델이 상대적으로 높은 민감도를 보이는 센서 영역이 달라지는 경향이 관찰된다. 먼저 정상 상태를 나타내는 [Figure 7(a)]의 Grad-CAM에서는 고주파 음향 방출 에너지가 뚜렷하지 않은 좌측의 AE 스펙트로그램 영역의 활성화가 낮게 나타나는 반면, 우측 진동 스펙트로그램의 중·하단 저주파 대역에 강한 붉은색 활성화가 형성되어 정상 상태 예측 시 회전축의 기저 진동 대역 정보가 상대적으로 크게 기여하고 있음을 보여준다. 반면, 4.4절에서 기술한 바와 같이 구름 요소의 자전·공전 및 미끄러짐 현상으로 인해 가변 속도 환경에서 주파수 smearing과 왜곡이 심하게 발생하는 롤러 결함([Figure 7(b)])의 Grad-CAM에서는 좌측 AE 스펙트로그램 하단의 고주파 영역과 우측 진동 스펙트로그램 중·하단의 공진 영역 양쪽 모두에서 높은 활성화 응답이 함께 관측된다. 비록 대표 샘플에 대한 사후 기울기 기반 시각화 결과만으로 신경망 내부의 작동 메커니즘을 확정할 수는 없으나, 이러한 공동 활성화 양상은 비선형적 주파수 왜곡 환경에서 제안 모델이 단일 텐서 내에 병합된 AE와 진동 스펙트로그램의 상호 보완적인 시간-주파수 단서를 함께 참조하고 있음을 뒷받침하는 보조적 근거를 제공한다.

V. Conclusion

본 연구에서는 연산 자원이 제한된 산업용 엣지 디바이스 환경에서도 가변 속도로 구동되는 회전 기계의 베어링 고장을 정확하게 진단할 수 있는 경량 Multi-sensor Input-level fusion 프레임워크를 제안하였다. 기존의 딥러닝 기반 고장 진단 연구들이 안고 있던 단일 센서의 정보적 한계, Feature-level fusion의 연산량 과부하, 그리고 무작위 데이터 분할이 야기하는 시간적 데이터 누수 문제를 파일 독립적 분할 기법과 공간 최적화 기법을 통해 해결하고자 하였다. 진동과 AE 센서를 단일 텐서로 융합한 제안 기법은 모든 단일 센서 모델 대조군(최고 91.17 ± 2.05%) 대비 일관되게 우수한 평균 성능을 보이며 98.77 ± 0.13%의 도메인 일반화 정확도를 달성하였다. 이는 두 센서의 상호 보완적 특징 결합을 통해 미지의 가변 속도 노이즈를 극복하였음 보여준다. 제안 기법에서 모델 최적화가 고도화되는 후반부에서는 원본 주파수 척도를 보존하고 여백을 무소음으로 격리한 Zero-padding 기법이 불필요한 특징 오염을 줄이며 반복 실험 전반에서 가장 높은 평균 정확도와 낮은 표준편차로(98.77 ± 0.13%) 안정적인 진단 성능을 도출함을 확인하였다. 롤러 결함과 같이 미끄러짐으로 인해 주파수 왜곡이 극심한 클래스에 대해서도 높은 정밀도로 분류해 내었으며, t-SNE 시각화 분석을 통해 제안 모델이 미학습 속도 조건에서도 각 결함 클래스의 특징이 뚜렷한 군집으로 분리되는 경향을 확인하였다.

본 연구는 단일 실험 셋업 및 고정된 센서 구성에서 수집된 실험실 데이터와 단일 크기의 결함(12 mm)만을 기반으로 평가되어, 다양한 설비와 강한 배경 노이즈가 공존하는 실제 산업 현장에 즉각 일반화하기에는 한계가 있다. FLOPs 분석 및 단일 보드 엣지 디바이스(Raspberry Pi 5) 상에서의 신경망 순전파 지연 시간 측정을 통해 모델 차원의 연산 경량성은 확인하였으나, 해당 측정에는 STFT 변환 및 스펙트로그램 생성 등의 전처리 소요 시간이 포함되지 않았으며 실제 산업 설비에 탑재된 현장 엣지 환경에서의 연속 구동 검증은 이루어지지 않았다. 따라서 향후 연구에서는 다양한 결함 크기와 복합 결함 상태, 가변 부하 및 가우시안 노이즈가 포함된 현장 밀착형 시나리오와 다양한 센서 구성을 대상으로 일반화 성능을 검증하고, 신호 수집 및 시간-주파수 전처리부터 신경망 추론까지 통합된 종단간 현장 엣지 파이프라인에서의 실시간 처리 가능성을 실증할 예정이다.

ACKNOWLEDGEMENT

This research was funded by a National Research Foundation of Korea (NRF) grant through the Korean Government [Ministry of Science and ICT (MSIT)] under Grant RS-2024-00453676.

REFERENCES

[1] R. B. Randall and J. Antoni, "Rolling element bearing diagnostics—A tutorial," Mechanical Systems and Signal Processing, Vol. 25, No. 2, pp. 485-520, Feb. 2011. DOI: 10.1016/j.ymssp.2010.07.017

[2] G. Wu, T. Yan, G. Yang, H. Chai, and C. Cao, "A Review on Rolling Bearing Fault Signal Detection Methods Based on Different Sensors," Sensors, Vol. 22, No. 21, art. No. 8330, Nov. 2022. DOI: 10.3390/s22218330

[3] R. Zhao, R. Yan, Z. Chen, K. Mao, P. Wang, and R. X. Gao, "Deep learning and its applications to machine health monitoring,"

www.kci.go.kr

Mechanical Systems and Signal Processing, Vol. 115, pp. 213-237, Jan. 2019. DOI: 10.1016/j.ymssp.2018.05.050

[4] W. A. Smith and R. B. Randall, "Rolling Element Bearing Diagnostics Using the Case Western Reserve University Data: A Benchmark Study," Mechanical Systems and Signal Processing, Vols. 64-65, pp. 100-131, Dec. 2015. DOI: 10.1016/j.ymssp.2015.04.021

[5] L. Wen, X. Li, L. Gao, and Y. Zhang, "A New Convolutional Neural Network-Based Data-Driven Fault Diagnosis Method," IEEE Transactions on Industrial Electronics, Vol. 65, No. 7, pp. 5990-5998, Jul. 2018. DOI: 10.1109/TIE.2017.2774777

[6] W. Zhang, G. Peng, C. Li, Y. Chen, and Z. Zhang, "A New Deep Learning Model for Fault Diagnosis with Good Anti-Noise and Domain Adaptation Ability on Raw Vibration Signals," Sensors, Vol. 17, No. 2, art. No. 425, Feb. 2017. DOI: 10.3390/s17020425

[7] R. S. Gunerkar and A. K. Jalan, "Classification of Ball Bearing Faults Using Vibro-Acoustic Sensor Data Fusion," Experimental Techniques, Vol. 43, No. 6, pp. 635-643, Dec. 2019. DOI: 10.1007/s40799-019-00324-0

[8] P. Ong, J. Y. Cheah, C. K. Sia, et al., "Intelligent fault diagnosis of bearings using multi-sensor spectrogram fusion and machine learning models," Iran Journal of Computer Science, Vol. 8, pp. 2295-2305, 2025. DOI: 10.1007/s42044-025-00316-x

[9] A. Howard et al., "Searching for MobileNetV3," 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pp. 1314-1324, Seoul, South Korea, Oct. 2019. DOI: 10.1109/ICCV.2019.00140

[10] H.-A.-H. Nguyen and C. H. Kim, "A Lightweight Heterogeneous Neural Network Architecture for Multiscale Feature Extraction in Bearing Fault Diagnosis," IEEE Access, Vol. 14, pp. 77717-77735, 2026. DOI: 10.1109/ACCESS.2026.3694523

[11] J. P. Vieira, V. A. Bauler, R. K. Rosa, and D. Silva, "Towards a more realistic evaluation of machine learning models for bearing fault diagnosis," Mechanical Systems and Signal Processing, Vol. 258, art. No. 114640, 2026. DOI: 10.1016/j.ymssp.2026.114640

[12] Y. Sun, Y. Qin, W. Chen, W. Zhao, and H. Sun, "Leakage-Resistant Multi-Sensor Bearing Fault Diagnosis via Adaptive Time-Frequency Graph Learning and Sensor Reliability-Aware Fusion," Sensors, Vol. 26, No. 8, art. No. 2484, Apr. 2026. DOI: 10.3390/s26082484

[13] L. Wheat, M. V. Mohrenschildt, S. Habibi, and D. Al-Ani, "Impact of Data Leakage in Vibration Signals Used for Bearing Fault Diagnosis," IEEE Access, Vol. 12, pp. 169879-169895, Nov. 2024. DOI: 10.1109/ACCESS.2024.3497716

[14] H. Wan, X. Gu, S. Yang, and Y. Fu, "A Sound and Vibration Fusion Method for Fault Diagnosis of Rolling Bearings under Speed-Varying Conditions," Sensors, Vol. 23, No. 6, art. No. 3130, Mar. 2023. DOI: 10.3390/s23063130

[15] M. T. Pham, J.-M. Kim, and C. H. Kim, "Accurate Bearing Fault Diagnosis under Variable Shaft Speed using Convolutional Neural Networks and Vibration Spectrogram," Applied Sciences, Vol. 10, No. 18, art. No. 6385, Sep. 2020. DOI: 10.3390/app10186385

[16] D. Verstraete, A. Ferrada, E. L. Droguett, V. Meruane, and M. Modarres, "Deep Learning Enabled Fault Diagnosis Using Time-Frequency Image Analysis of Rolling Element Bearings," Shock and Vibration, Vol. 2017, art. No. 5067651, Aug. 2017. DOI: 10.1155/2017/5067651

[17] M. Kisantal, Z. Wojna, J. Murawski, J. Naruniec, and K. Cho, "Augmentation for small object detection," 9th International Conference on Advances in Computing and Information Technology (ACITY 2019), pp. 119-133, Sydney, Australia, Dec. 2019. DOI: 10.5121/csit.2019.91713

[18] G. Ghiasi et al., "Simple Copy-Paste is a Strong Data Augmentation Method for Instance Segmentation," 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2917-2927, Nashville, TN, USA, Jun. 2021. DOI: 10.1109/CVPR46437.2021.00294

[19] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," Journal of Big Data, Vol. 6, No. 1, art. No. 60, Jul. 2019. DOI: 10.1186/s40537-019-0197-0

[20] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," Journal of Machine Learning Research, Vol. 9, No. 86, pp. 2579-2605, Nov. 2008.

Authors



Yoonsik Cho received the B.S. degree in Information and Communications through the Academic Credit Bank System at Soongsil University, Korea, in 2025. Yoonsik Cho joined the Department of Computer Science and Engineering at Soongsil University, Seoul, Korea, in 2025, where he is currently pursuing the M.S. degree. His research interests include Fault Diagnosis and Deep Learning.



Cheol Hong Kim received the B.S. and M.S. degrees in Computer Engineering from Seoul National University, Korea, in 1998 and 2000, respectively and the Ph.D. degree in Electrical and Computer Engineering from the same university, in 2006. Dr. Kim was a Senior Engineer at Samsung Electronics, Korea, from December 2005 to January 2007. and a Professor at Chonnam National University, Korea, from 2007 to 2020. He is currently a Professor in the School of Computer Science and Engineering at Soongsil University, Korea. His research interests include Computer Systems, Embedded Systems, Deep Learning, Fault Diagnosis, and Low-power Systems.

www.kci.go.kr