

AI 음향 데이터 기반 소규모 양계장 이상 상황 탐지 시스템

(Design of an AI-Based Acoustic Anomaly Detection System for Small-Scale Poultry Farms)

김희서*, 김민희**, 장윤서**, 황현빈**, 정경용***

(HeeSeo Kim, MinHee Kim, YunSeo Jang, HyunBin Hwang, KyoungYong Chung)

요약

닭은 고통, 스트레스, 질병 등 상태 변화에 따라 음향적으로 구별 가능한 발성을 산출하며, 이러한 발성 패턴은 외형적 변화에 선행하여 나타나는 것으로 보고되고 있다. 최근 딥러닝 기반 음향 분류를 활용한 가금류 건강 모니터링 연구가 활발히 이루어지고 있으나, 대부분의 선행 연구는 사전 정의된 클래스에 대한 지도 학습 기반 분류에 한정되어 있어, 비정상 데이터의 사전 확보가 어려운 양계 현장에서의 실질적 적용에는 제약이 따른다. 본 연구에서는 닭의 고통 울음, 호흡기 질환음, 스트레스성 발성 등을 분류 및 탐지하는 AI 기반 이상 음향 감지 시스템을 제안한다. CNN 기반 다중 분류 모델과 정상 데이터만으로 학습한 Autoencoder를 결합한 하이브리드 구조를 채택하여, 라벨링된 이상 데이터가 부족한 실제 농장 환경에서도 적용 가능한 경량 이상 탐지 시스템의 구현을 목표로 한다.

■ 중심어 : 양계장 이상 탐지 ; 닭 울음소리 분석 ; MFCC ; 하이브리드 딥러닝 ; 오토인코더

Abstract

Chickens produce acoustically distinguishable vocalizations corresponding to various physiological and psychological states, including pain, stress, and disease, and such vocal patterns have been reported to precede observable physical symptoms. Although deep learning-based acoustic classification for poultry health monitoring has received increasing research attention, most prior studies have been confined to supervised classification over predefined classes, which limits their practical applicability in commercial poultry environments where abnormal data are difficult to obtain in advance. This study proposes an AI-based abnormal vocalization detection system designed to classify and identify distress calls, respiratory illness sounds, and stress-induced vocalizations in chickens. The proposed system employs a hybrid architecture combining a CNN-based multi-class classification model with an Autoencoder trained exclusively on normal acoustic data, with the objective of implementing a lightweight anomaly detection system deployable in real farm settings where labeled abnormal samples are scarce.

■ keywords : Poultry Anomaly Detection ; Chicken Vocalization Analysis ; MFCC ; Hybrid Deep Learning ; Autoencoder

I. 서론

닭은 고통, 스트레스, 질병 등 상태 변화에 따라 음향적으로 구별 가능한 발성을 산출하며, 이

러한 발성 패턴은 외형 변화보다 먼저 나타나는 경우가 많다[1]. 특히 distress call은 사회적 고립 등 다양한 스트레스 상황에서 신뢰성 있게 유발되며, 닭은 맥락에 따라 20종 이상의 구별되는 발성을 보유하는 것으로 보고되고 있다[2]. 이를

* 준회원, 경기대학교 ICT융합학부 경영정보전공

** 준회원, 경기대학교 시컴퓨터공학부

*** 정회원, 경기대학교 시컴퓨터공학부

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 SW중심대학지원사업의 연구결과로 수행되었음(2021-0-01393)

접수일자 : 2026년 05월 14일

게재확정일 : 2026년 05월 27일

교신저자 : 정경용 e-mail : dragonhci@gmail.com

토대로 딥러닝 기반 음향 분류 모델을 활용한 가금류 건강 판별 연구가 활발히 진행되고 있으나 [3], 대부분 사전 정의된 클래스에 대한 지도 학습 분류에 초점을 두고 있어, 이상 데이터 확보가 어려운 양계 현장에서의 적용에는 구조적 한계가 존재한다.

이에 본 연구에서는 CNN 기반 다중 분류와 정상 데이터만으로 학습한 Autoencoder를 결합한 하이브리드 구조를 채택하여, 라벨이 부족한 실제 농장 환경에서도 적용 가능한 경량 이상 탐지 시스템을 제안한다.

II. 관련 연구

닭 울음소리를 활용한 이상 상태 탐지 연구는 딥러닝 기반 음향 분석 기술의 발전과 함께 다양한 방향으로 시도되었다. Mao et al.[3]의 연구에서는 실제 상업 양계 농장의 음성 데이터를 수집하고 경량화된 합성곱 신경망 구조를 적용하여, Distress Call에 대해 95.07%의 탐지 정확도와 94.73%의 F1-Score를 달성하였다. 특히 기존 VGG11 대비 파라미터 수를 14분의 1 수준으로 줄이면서도 실제 계사 소음 환경 내 강건한 성능을 유지하였다는 점에서, 저사양 현장 장치에의 적용 가능성을 구체적으로 확인한 연구로 평가된다. 한편 SmartEars[4]는 기침 및 Rales를 포함한 호흡기 이상음을 스펙트로그램 이미지로 변환하여 분류하는 방식을 산란계 농장에 적용하였으며, 감염 개체에서 육안으로 증상을 확인하기 이전 단계에서 이미 음향 신호의 변화가 선행한다는 사실을 현장 실험을 통해 밝혔다. Neethirajan[5]은 스트레스 상황에서 발생하는 발성 빈도의 변화와 산란량 저하 사이의 연관성을 MFCC 특징 분석을 통해 실증하였고, 이상 발성 데이터 없이 정상 발성 데이터만으로 학습한 Autoencoder를 활용하여 재구성 오차 기반의 이상 탐지가 가능함을 제시하였다. Soster et al.[6] 또한 육계 군집 단위의 울음소리 패턴 변화를 추적하여 이상 행동을 자동으로 감지하는

시스템을 구현하였으며, 음향 모니터링이 개체 수준을 넘어 계사 전체의 상태 변화를 조기에 반영할 수 있음을 보였다.

그러나 위 연구들은 개별 음향 분류 모델의 성능 검증에 초점이 맞추어져 있으며, 이상 발생 데이터 확보가 현실적으로 어려운 환경을 전제로 지도 학습 기반 다중 분류와 비지도 학습 모델 기반 이상 탐지를 동시에 수행하는 하이브리드 구조에 대한 연구는 충분히 이루어지지 않았다. 본 연구는 이 지점을 출발점으로 삼아, CNN 분류기와 Autoencoder를 결합한 하이브리드 AI 모델을 설계함으로써 사전에 정의되지 않은 이상 징후에도 유연하게 대응 가능한 탐지 체계를 구현하고자 한다.

III. 시스템 및 연구 방법

1. 데이터 수집 및 전처리

본 연구는 distress, stress[7], respiratory 세 가지의 공개 음성 데이터셋[4][8][9][10]을 통합하여 학습 corpus를 구성한다. 모든 클립은 22,050 Hz, monaural, 3.0 s 고정 길이 WAV로 정규화하였다. 라벨은 distress, respiratory, stress 세 클래스에 대한 multi-label 형식으로 구성하였으며, 모델은 각 클래스의 sigmoid 확률을 출력한다. healthy 샘플[4][9][10]은 영벡터 [0, 0, 0]으로 라벨링 하여 별도의 negative class 없이 negative rejection을 동일 출력 공간에서 학습한다.

표 1. 클래스 구성 및 라벨 인코딩 (총 70,001 학습 샘플)

클래스	총 샘플	기존 샘플 수	원본 / 증강	라벨
distress	10,000	3,363	4,311 / 5,689	[1, 0, 0]
respiratory	10,001	1,597	3,577 / 6,424	[0, 1, 0]
stress	25,000	17	20,031 / 4,969	[0, 0, 1]
healthy	25,000	170	18,552 / 6,448	[0, 0, 0]

표 1의 noise[10][11][12] 836개는 학습 corpus와 분리된 train-time online augmentation 별도 공간으로 신호 $s(t)$ 에 SNR-controlled mix-in으로 혼합되어 생성된 샘플은 원본 신호 라벨을 계승한다.

WAV의 모델 입력 텐서로 변환하는 과정은 다음의 7-stage pipeline으로 표준화하였다 (1)soundfile + librosa.resample fallback, (2)22,050 Hz monaural, (3)8th-order Butterworth low-pass filter ($f_c = 8$ kHz, sosfiltfilt zero-phase), (4) Mel-spectrogram ($n_{\text{mels}} = 128$, $n_{\text{fft}} = 2,048$, $\text{hop_length} = 512$), (5)시간축 128 frame center-crop / zero-pad, (6) min-max \rightarrow dataset statistics 표준화 ($\mu = 0.2599$, $\sigma = 0.2132$), (7) single-channel \rightarrow 3-channel replication으로 (3, 128, 128) tensor 산출 (ImageNet 호환). 8 kHz cutoff는 사전 ablation에서 composite +0.0384, macro F1 +0.0877 향상된 결과가 나타났다. 이로써 녹음 장비 의존적 high-frequency 잡음(mic hum, cable noise, quantization noise)을 차단하여 shortcut learning을 억제한 결과로 해석된다. 닭 발성 핵심 대역 300~4,000 Hz에 분포하므로 신호 손실은 무시 가능한 수준이다.

respiratory 원본은 장시간 연속 녹음으로 FFT 분석 결과 50 Hz 전원 주파수가 전체 에너지의 9.47%, 1,000 Hz 계열 tonal peak 61개가 광역 분포한 결과를 관찰하였다. 이는 호흡기 발성이 아닌 전기 잡음 패턴을 학습하는 shortcut 위험이 있다고 판단하였다. 이러한 문제를 방지하기 위해 사전 정제 모듈에서는 (1) 5th-order Butterworth high-pass ($f_c = 100$ Hz), (2) FFT 기반 tonal-peak 자동 탐지, (3) IIR notch filter ($Q = 50$)을 순차 적용하여 300~4,000 Hz 핵심 대역 에너지 비율을 32.6%에서 58.3%로 약 1.8배 향상시켰다.

본 데이터에서는 sklearn의 Group Shuffle Split을 70 : 15 : 15로 적용하되 group key에 (1) 증강 suffix 제거로 원본 식별자 추출, (2) cross-class

ID 충돌을 막는 클래스 prefix 부여, (3) PrS \leftrightarrow PoS 동일 개체(stress 처리 전후 동일 닭, Trt-W-D 패턴)를 의도적으로 같은 fold에 묶는 세 가지 보정을 주었다. 분할 후 누수 의심 group key 13개는 모두 PrS \leftrightarrow PoS 동일 개체로 확인되어 leakage = 0이 보장된다.

표 2. Group-aware Train / Val / Test 분할

분할	증강 후 샘플	기존 샘플 수	distress	resp.	stress	healthy
train	46,786	3,593	6,909	7,251	16,078	16,548
val	9,367	770	1,641	1,270	4,486	1,970
test	13,848	771	1,450	1,480	4,436	6,482

2. 데이터 증강 기법

표 3. 7-fold offline augmentation 설계

기법	도메인	파라미터	이론적 의미
Additive Gaussian noise (AWGN)	Waveform	$\epsilon \sim \mathcal{N}(0, \sigma^2)$, $\sigma \sim \mathcal{U}(0.002, 0.015)$	background noise 다양화
Circular time shift	Waveform	$\Delta t \sim \mathcal{U}(-0.5, 0.5)$ s, torch.roll	발성 onset invariance
Multiplicative gain	Waveform	$g \sim \mathcal{U}(0.5, 1.5)$	마이크 거리·개체 음량 변동
Pitch shift (PSOLA)	Waveform	± 3 semitone	F0 perturbation, 개체 차이 모사
Time-scale modification	Waveform	rate $\sim \mathcal{U}(0.8, 1.2)$, STFT phase vocoder	tempo perturbation, pitch-invariant
SpecAugment	Spectrogram	FreqMask 20, TimeMask 30	spectro-temporal masking
Convolutional reverb (RIR)	Waveform	$h(t) = \exp(-5t/r)$, $r \sim \mathcal{U}(0.2, 0.5)$	synthetic RIR 양계장 잔향 모사

양계장 이상 발성 corpus의 수집 한계와 도메인 특이성을 보완하기 위해 본 연구는 waveform·time-frequency 도메인의 7-fold offline augmentation과 train-time online noise injection을 상호 독립적으로 결합하였다. 그리고 증강 데이터는 동일 split에 배치하여 데이터 누수를 방지하였다.

$$s_{noise}(t) = s(t) + a \cdot n(t), a = \sqrt{\frac{P_s}{P_n \cdot 10^{\frac{SNR}{10}}}} \quad (1)$$

$$SNR \sim u(5, 20) \text{ dB}$$

주입 확률은 $p = 0.5$ 로 train split에 대해서만 적용하였다. MUSAN의 robustness와, Mendeley Poultry의 domain-specific adaptation을 동시에 사용한다. SNR $u(5, 20)$ dB 범위는 5 dB 근방의 강한 잡음에서 신호 보존(noise suppression)을, 20 dB 근방의 약한 잡음에서 미세 발생 단서 식별을 동시에 학습시킨다. 잡음을 라벨 클래스로 두지 않고 augmentation channel로 처리한 이유는 양계장 잡음이 끊임없이 변화하는 신호이기 때문이다. 단힌 라벨 집합으로 학습할 경우 학습 분포 외 잡음에 대한 일반화가 보장되지 않는다.

3. 모델 설계

본 모델은 supervised CNN classifier와 unsupervised contrastive autoencoder(AE)를 결합한 hybrid architecture이다. 두 모듈은 동일한 mel-spectrogram 입력을 공유하지만 학습 목표·손실·하이퍼 파라미터가 분리되어 있어 한 경로의 재학습이 다른 경로에 영향을 주지 않는다. closed-set classifier와 open-set anomaly detector를 상보적으로 결합한 학술적 표준 패턴이다.

가. CNN classifier

Backbone은 ImageNet pretrained EfficientNet-B0 (≈ 4.0 M parameters)이며, classifier head를 Dropout($p = 0.2$) \rightarrow Linear(1280 \rightarrow 3) \rightarrow sigmoid로 교체하였다. multi-label sigmoid 출력은 all-zero 상태로 negative rejection을 자연스럽게 표현하며, 입력은 (3, 128, 128) channel-replicated mel-spectrogram이다.

전이 학습 안정성을 위해 2-stage fine-tuning을 채택한다. Stage 1은 backbone freeze + head only (LR = $1e-3$, 20 epoch), Stage 2는 unfreeze

+ 전체 미세조정 (LR = $1e-4$, max 100 epoch, EarlyStopping patience = 15, monitor = composite). 두 stage 공통으로 AdamW (weight_decay = $1e-4$), cosine annealing + warmup 5 epoch, gradient clipping (max_norm = 1.0)을 적용하였다. 본 학습에서 Stage 2 ep1이 composite 0.9547로 best 선정되고 ep16에서 EarlyStopping이 작동되었다.

손실 함수는 BCEWithLogitsLoss에 세 가지 보정을 결합한다. (1) label smoothing 0.05로 sigmoid over-confidence 완화하고 (2) negative sample(label = [0, 0, 0])에 $10\times$ weight 부여로 false positive 강제 제한하고 (3) per-class pos_weight를 비대칭으로 적용하였다. B2_v2 설정은 distress = 1.0, respiratory = 0.7, stress = 2.0이다. WeightedRandomSampler로 stress sample을 $1.5\times$ oversampling 하여 mini-batch 내 stress 비율을 $0.344 \rightarrow 0.440$ 으로 높였다.

$$L = \text{mean}(w_{neg(i)} \cdot \text{BCE}(\sigma(z_i; pos_weight), y_i^{smooth})) \quad (2)$$

$$w_{neg} = \begin{cases} 10, & \text{if } y_i = [0, 0, 0] \\ 1, & \text{otherwise} \end{cases}$$

학습 안정성과 best-checkpoint 선정 noise를 줄이기 위해 모든 학습 가능 파라미터의 Exponential Moving Average (EMA, decay = 0.999) shadow weight를 별도로 유지하며, validation-checkpoint 평가 시점에 EMA를 임시 swap-in한다. best ckpt 선정과 EarlyStopping 기준은 과적합 문제로 macro F1 단독이 아닌 다 음 합성 지표를 사용한다.

$$\text{composite} = 0.3 \cdot \text{macro}_{F1} + 0.7 \cdot \text{negative}_{rejection \neq rate} \quad (3)$$

나. Contrastive Autoencoder

Supervised classifier는 사전에 정의되지 않은 anomaly(out-of-distribution; OOD) 대응이 제한적이다. AE는 healthy 분포 중심의

reconstruction-based unsupervised anomaly detection으로 closed-set classifier를 보완하였다. CNN은 ImageNet 호환을 위해 (3, 128, 128)을 사용하지만 AE는 1ch 형태 (1, 128, 128)을 그대로 사용한다. 4-stage convolutional encoder + symmetric ConvTranspose decoder의 fully convolutional 구조이며, 인코더 채널 1 → 32 → 64 → 128 → 256 ($k = 3, s = 2$), latent shape (256, 8, 8)으로 약 8× spatial compression이고 디코더는 ConvTranspose2d ($k = 4, s = 2$)를 통해 대칭 업샘플링을 수행한다. 모든 단계 BatchNorm + LeakyReLU(0.1)을 적용하였고 총 파라미터는 약 1.4 M이다.

Healthy MSE만으로 학습할 경우 healthy·disease 모두 동일 도메인(담 발성)에서 mel 분포가 유사하므로 AE가 healthy 분포가 아닌 담 발성 일반을 학습하는 문제가 발생한다. 이를 차단하기 위해 본 연구는 healthy reconstruction MSE를 최소화하되 disease reconstruction MSE가 사전 정의된 margin 이하로 내려가지 않도록 hinge penalty를 부여하는 contrastive 비대칭 손실을 채택하였다.

$$L_{AE} = \frac{1}{N_h} \sum_{x \in H} \|x - \hat{x}\|^2 + 0.3 \cdot \max(0, 0.05 - \frac{1}{N_d} \sum_{x \in D} \|x - \hat{x}\|^2) \quad (4)$$

30 epoch 학습 결과 best val healthy MSE = 0.00677, test 단계에서 disease 평균 reconstruction error는 0.1278로 healthy 평균 0.0117의 약 11×로 분리되었다.

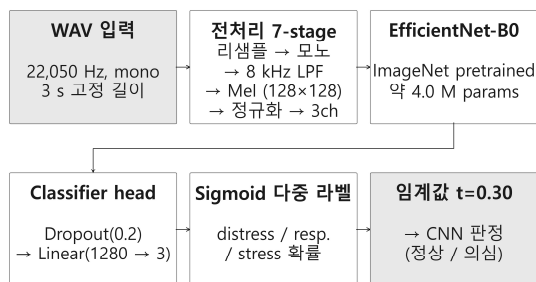


그림 1. 이상 탐지 시스템의 지도 학습 구성도

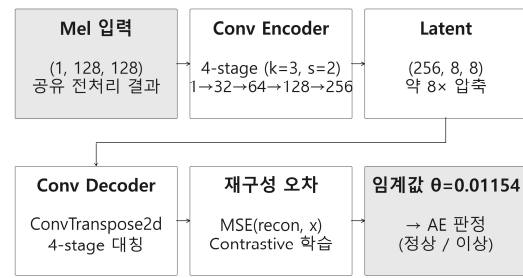


그림 2. 이상 탐지 시스템의 비지도 학습 구성도

해당 모델은 담 음성의 이상 징후 탐지를 위해 CNN 분류기와 Autoencoder를 결합한 하이브리드 모델을 설계하였다. 입력 데이터는 3초 길이의 WAV 파일이며, 22,050 Hz로 샘플링하였다. 입력 음성은 Butterworth 8 kHz 저역통과필터를 통해 고주파 잡음을 제거한 뒤 Mel-Spectrogram으로 변환하였고, 128×128 크기로 정규화하여 3×128×128 형태의 모델 입력으로 구성하였다.

CNN 분류기에는 ImageNet으로 사전 학습된 EfficientNet-B0를 사용하였다. 분류기 말단에는 Dropout과 Linear Layer를 추가하였으며, Sigmoid 함수를 통해 distress, respiratory, stress의 세 가지 클래스별 이상 확률을 산출하였다. 산출된 확률은 임계값 $t=0.30$ 과 비교하여 정상 또는 의심 여부를 판정한다. 비지도 학습 기반 Autoencoder는 정상 음성의 재구성 특성을 학습하도록 설계하였다. 입력 Mel-Spectrogram을 재구성한 뒤 원본과의 평균제곱오차를 계산하고, 이를 이상 점수로 사용하였다. 재구성 오차가 임계값 $\theta=0.01154$ 를 초과하면 이상 음성으로 판단한다.

최종 판정은 CNN 분류 결과와 Autoencoder의 이상 점수를 함께 고려하여 수행된다. 두 모델이 모두 이상으로 판단한 경우 신뢰도 높은 이상으로 분류하고, 한쪽 모델만 이상을 탐지한 경우에는 재확인이 필요한 음성으로 판단한다. 이를 통해 제안 모델은 특정 이상 클래스 분류와 정상 패턴 이탈 탐지를 동시에 수행할 수 있다.

4. 예측 방법 및 이상 탐지 기준

운영 단일 threshold t 는 validation split에서 composite metric을 최대화하는 값을 grid search ($t \in [0.05, 0.60]$, step 0.01)로 탐색하여 결정하였다. B2_v2의 최적 $t^* = 0.30$ 이며, 이를 test에 적용하였다. 클래스별 PR-curve의 F1-max threshold는 distress = 0.32, respiratory = 0.22, stress = 0.08로 산출되었으나 stress의 0.08은 권장 범위 [0.20, 0.35]에서 크게 이탈하여 group 다양성 부족이라는 데이터 차원 한계의 신호로 해석되며, 운영은 단일 global threshold로 통일하였다.

AE anomaly threshold θ 는 validation healthy 샘플(N = 1,970)의 reconstruction error 분포의 95-percentile로 결정하였다. 결정값 $\theta = 0.01154$ 를 고정값으로 사용한다.

표 4. CNN × AE 이중 경로 의사결정 매트릭스

CNN	AE	최종 판정	신뢰도	권장 조치
all < t	err < θ	정상	high	정기 모니터링
$\geq 1 \geq t$	err $\geq \theta$	이상-의심 클래스	high	즉시 경보, IoT trigger
$\geq 1 \geq t$	err < θ	이상 가능성	mid	추가 관찰, 경보 hold
all < t	err $\geq \theta$	OOD / 미지 패턴	mid	신중 패턴 검토

이중 경로 의사결정의 학습 가치는 다음 세 가지이다. (1) 두 독립 학습 경로를 통한 high confidence 판정 강화, (2) CNN positive \wedge AE normal 불일치를 false positive 의심으로 분류한 negative gating 보강, (3) CNN negative \wedge AE anomaly를 학습 분포 외 신호로 포착하는 OOD 감지이다.

Calibration 검증으로 validation logits에 LBFGS-fit한 temperature scaling 결과 $T = 1.046$ 으로 거의 1에 수렴해 EMA + label smoothing의 자동 보정 효과가 확인된다. 따라서 raw sigmoid 확률에 단일 임계값 $t = 0.30$ 과

AE 임계값 $\theta = 0.01154$ 를 적용한 표 4의 4-quadrant decision rule을 그대로 사용한다.

IV. 성능 평가 및 결과

본 연구는 그룹 단위로 사전 봉인된 held-out test split에서 글로벌 임계값 0.34를 적용하여 최종 성능을 평가하였다. 평가 지표는 양계장 운영 환경의 특성을 반영하여 Macro F1과 건강 반응 (Healthy) 거절률을 0.3:0.7로 결합한 composite metric을 1등급 지표로 채택하였다. 또한 통계적 신뢰성 확보를 위해 bootstrap(1,000회 반복) 기반 95% 신뢰구간을 함께 보고하였다. 잘못된 양성 판정이 폐사 위험률을 높이는 양계장 특성을 고려해 건강 반응 거절률에 높은 가중치를 부여하는 것은 운영 환경의 비대칭 손실 구조를 직접 반영한 설계 결정이다.

표 5. CNN 분류기 종합 성능 지표

지표	값	95% 신뢰구간
Composite	0.9023	-
Macro F1	0.7737	[0.7677, 0.7797]
건강 반응 거절률	0.9574	[0.9526, 0.9623]

표 6. 클래스별 분류 성능

클래스	F1	정밀도	재현율
고통 반응	0.9924	0.9938	0.9910
호흡기 질환	0.9267	0.9184	0.9351
스트레스 반응	0.4021	0.8621	0.2622

표 5, 6에서 확인할 수 있듯이 고통 반응과 호흡기 질환 클래스는 전반적으로 높은 성능을 달성하였다. 반면 스트레스 반응 클래스는 F1 0.4021로 나머지 두 클래스 대비 낮은 수치를 기록하였는데, 이는 학습 데이터의 그룹 다양성이 전체 17개 세션에 불과한 구조적 한계에서 기인하며, 클래스별 최적 임계값 분석에서도 스트레스 반응의 최적값이 0.087로 정상 범위

(0.20~0.35)를 크게 벗어나 새로운 세션 패턴에 대한 일반화 한계를 명확히 드러낸다.

Autoencoder 보조 모듈은 건강 반응 샘플 중심 학습에 더해 질환(Disease) 샘플에 대해 의도적으로 높은 재구성 오차를 유도하는 대조 손실을 도입하였다. 그 결과 질환 샘플 평균 재구성 오차가 건강 반응 평균의 약 11배(0.1278 vs 0.0117)로 명확히 분리되었으며, 단일 임계값(validation healthy 95 percentile = 0.01154) 기준으로 고통 반응 99.8%, 호흡기 질환 96.6%, 스트레스 반응 35.0%의 이상 감지율을 달성하였다. 건강 반응 정상 분류율은 84.1%로 유지되어 오탐 억제와 이상 감지 사이의 trade-off가 매우 유리하였다. CNN 분류기와 완전히 독립적으로 동작하는 구조 덕분에 두 모듈의 판정 일치 여부를 기반으로 신뢰도를 3단계(양쪽 동의 / 한쪽 동의 / Autoencoder 단독 이상)로 구분하여 실제 운영에서의 의사결정 보조 기능을 제공한다.

학습 과정에서 composite metric의 유효성도 실증적으로 검증되었다. Stage 2 학습 중 Macro F1은 지속 상승하였으나 건강 반응 거절률이 급락하는 과적합 패턴이 관찰되었다. Composite metric이 이를 조기에 감지하여 EarlyStopping이 16 epoch에서 정상 작동하였으며, 이는 양계장 운영 환경에 특화된 평가 기준으로서 본 시스템의 신뢰성 확보에 실질적으로 기여하였다.

V. 결론

본 연구는 양계장 환경에서 수집된 닭 음성 데이터를 기반으로 고통 반응, 호흡기 질환, 스트레스 반응을 동시에 판별하는 다중 라벨 분류 시스템을 설계하고 평가하였다. EfficientNet-B0를 백본으로 하는 CNN 분류기와 Contrastive Autoencoder를 결합한 지도·비지도 하이브리드 구조를 통해 오탐 억제와 이상 감지 신뢰도를 동시에 확보하였으며, Tkinter 기반 GUI를 통해 운영자가 모델 선택과 임계값 조정을 직접 수행할 수 있는 실용적 추론 환경을 구현하였다.

고통 반응 및 호흡기 질환 클래스에서 CNN 분류기는 높은 정밀도와 재현율을 달성하였으며, composite metric 설계가 과적합 함정을 효과적으로 차단함을 실험적으로 입증하였다. 또한 Butterworth 8 kHz 저역통과 필터 적용이 composite 지표를 3.84%p 향상시킴을 ablation 실험으로 확인하였고, 그룹 단위 데이터 분할과 클래스별 prefix 부여를 통해 녹음 장비별 단축 학습과 데이터 누수를 완전히 차단하였다. 다만 스트레스 반응 클래스의 성능 한계는 본 연구의 명확한 제약으로 남는다. 전체 17개 세션에 집중된 그룹 다양성 부족은 클래스 가중치 조정이나 오버샘플링과 같은 학습 기법만으로 극복하기 어려운 데이터 차원의 본질적 제약이며, 근본적 해결을 위해서는 다양한 농장 및 개체 환경에서 수집된 추가 데이터 확보가 선행되어야 한다.

향후에는 AST-Small과의 앙상블 실험, Grad-CAM 기반 오분류 시각화 분석, 실제 양계장 현장 데이터를 활용한 외부 일반화 검증을 통해 시스템의 실용성을 더욱 높이고, IoT 환경 데이터와 결합하여 양계장의 이상 상황을 조기에 파악하고 폐사율을 감소시킬 계획이다.

REFERENCES

- [1] G. Manteuffel, B. Puppe, P. C. Schön, "Vocalization of farm animals as a measure of welfare," *Applied Animal Behaviour Science*, Vol. 88, No. 1-2, pp. 163-182, Sep. 2004.
- [2] K. A. Herborn, A. G. McElligott, M. A. Mitchell, V. Sandilands, B. Bradshaw, L. Asher, "Spectral entropy of early-life distress calls as an iceberg indicator of chicken welfare," *Journal of The Royal Society Interface*, Vol. 17, No. 167, pp. 1-10, Jun. 2020.
- [3] A. Mao et al., "Automated identification of chicken distress vocalizations using deep learning models," *Journal of the Royal Society Interface*, vol. 19, no. 191, 20210921, 2022.
- [4] H. Qiao et al., "SmartEars: A practical framework for poultry respiratory monitoring via spectrogram-based audio classification and AI-assisted labeling," *Computers and Electronics in Agriculture*, vol. 241, 111241, 2026.

- [5] R. Srinivasagan, M. S. El Sayed, M. I. Al-Rasheed, and A. S. Alzahrani, "Edge intelligence for poultry welfare: Utilizing tiny machine learning neural network processors for vocalization analysis," *PLOS ONE*, vol. 20, no. 1, e0316920, Jan. 2025.
- [6] P. de Carvalho Soster et al., "Automated detection of broiler vocalizations: A machine learning approach for broiler chicken vocalization monitoring," *Poultry Science*, vol. 104, no. 5, 104962, May 2025.
- [7] Suresh Neethirajan, Vocalization Patterns in Laying Hens: An Analysis of Stress-Induced Audio Responses (2023), <https://www.biorxiv.org/content/10.1101/2023.12.26.573338v2.full>, (accessed Apr., 2, 2026).
- [8] Axiu Mao, Claire S. E. Giraudet, Kai Liu, Inês De Almeida Nolasco, Zhiqin Xie, Zhixun Xie, Yue Gao, James Theobald, Devaki Bhatta, Rebecca Stewart, Alan G. McElligott, Automated identification of chicken distress vocalisations using deep learning models (2022), https://figshare.com/articles/dataset/Automated_identification_of_chicken_distress_vocalisations_using_deep_learning_models/20049722, (accessed Apr., 20, 2026).
- [9] Neethirajan, Suresh, Vocalization Patterns in Laying Hens - An Analysis of Stress-Induced Audio Responses (2023), <https://zenodo.org/records/10433023>, (accessed Apr., 11, 2026).
- [10] Halleluwah Aworinde, Segun Adebayo, Akinwale Akinwunmi, Olufemi Alabi, Adebamiji Ayandiji, Olaide Oke, Abel Oyebamiji, Adetoye Adeyemo, Aderonke Sakpere, Kizito Echetama, Poultry Vocalization Signal Dataset for Early Disease Detection (2023), <https://data.mendeley.com/datasets/zp4nf2dxbh/1>, (accessed Apr., 20, 2026).
- [11] David Snyder, Guoguo Chen, Daniel Povey, MUSAN: A Music, Speech, and Noise Corpus (2015), <https://arxiv.org/abs/1510.08484>, (accessed Apr., 25, 2026).
- [12] David Snyder, Guoguo Chen, Daniel Povey, MUSAN: A Music, Speech, and Noise Corpus (2015), <https://www.openslr.org/17/>, (accessed Apr., 25, 2026).

 저자 소개



김희서(준회원)

2022년 경기대학교 소프트웨어경영대학 ICT융합학부 경영정보전공 입학, AI컴퓨터공학부 복수전공

<주관심분야 : 정보시스템, 데이터베이스, 데이터 마이닝, 인공지능 등 >



김민희 (준회원)

2024년 경기대학교 AI컴퓨터공학부 컴퓨터공학과 입학

<주관심분야 : 데이터 마이닝, 인공지능, AI 응용 시스템, 운영체제 >



장윤서 (준회원)

2024년 경기대학교 컴퓨터공학과 AI컴퓨터공학부 입학

<주관심분야 : AI 응용 시스템, 딥러닝 모델링, IoT, 임베디드 시스템 >



황현빈 (준회원)

2022년 경기대학교 ICT융합학부 산업시스템공학과 입학

2023년 경기대학교 AI컴퓨터공학부 컴퓨터공학과 전과

현재 경기대학교 AI컴퓨터공학부 컴퓨터공학과 재학 중

<주관심분야 : 모바일 앱 개발, 데이터 마이닝, IoT 연동 애플리케이션, 웹/앱 기반 데이터 시각화 >



정경용 (정회원)

2000년 2월 : 인하대학교 전자계산공학과(공학사)

2002년 2월 : 인하대학교 전자계산공학과(공학석사)

2005년 8월 : 인하대학교 컴퓨터정보공학부(공학박사)

2006년 3월 : 상지대학교 컴퓨터정보공학부 교수

2017년 3월: 경기대학교 컴퓨터공학부 교수
<주관심분야 : 데이터 마이닝, 헬스케어, 빅데이터, HCI, 지능시스템, 인공지능, 추천 시스템 >