JKSCI

한국컴퓨터정보학회논문지
Journal of The Korea Society of Computer and Information
Vol. 28 No. 9, pp. 47-54, September 2023
https://doi.org/10.9708/jksci.2023.28.09.047

Classification of Infant Crying Audio based on 3D Feature-Vector through Audio Data Augmentation

JeongHyeon Park*, JunHyeok Go*, SiUng Kim*, Nammee Moon*

*Student, Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea
*Student, Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea
*Student, Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea
*Professor, Dept. of Computer Science and Engineering, Hoseo University, Asan, Korea

[Abstract]

Infants utilize crying as a non-verbal means of communication [1]. However, deciphering infant cries presents challenges. Extensive research has been conducted to interpret infant cry audios [2,3]. This paper proposes the classification of infant cries using 3D feature vectors augmented with various audio data techniques. A total of 5 classes (belly pain, burping, discomfort, hungry, tired) are employed in the study dataset. The data is augmented using 5 techniques (Pitch, Tempo, Shift, Mixup-noise, CutMix). Tempo, Shift, and CutMix augmentation techniques demonstrated improved performance. Ultimately, applying effective data augmentation techniques simultaneously resulted in a 17.75% performance enhancement compared to models using single feature vectors and original data.

▸ **Key words**: 3D Feature Vector, Data Augmentation, Infant, MFCC, Nonverbal sound

[요 약]

영아는 비언어적 의사 소통 방식인 울음이라는 수단을 사용한다[1]. 하지만 영아의 울음소리를 파악하는 것에는 어려움이 따른다. 영아의 울음소리를 해석하기 위해 많은 연구가 진행되었다[2,3]. 이에 본 논문에서는 다양한 음성 데이터 증강을 통한 3D 특징 벡터를 이용한 영아의 울음소리 분류를 제안한다. 연구에서는 총 5개의 클래스 복통, 하품, 불편함, 배고픔, 피곤함(belly pain, burping, discomfort, hungry, tired)로 분류된 데이터 세트를 사용한다. 데이터들은 5가지 기법(Pitch, Tempo, Shift, Mixup-noise, CutMix)을 사용하여 증강한다. 증강 기법 중에서 Tempo, Shift, CutMix 기법을 적용하였을 때 성능의 향상을 보여주었다. 최종적으로 우수한 데이터 증강 기법들을 동시 적용한 결과 단일 특징 벡터와 오리지널 데이터를 사용한 모델보다 17.75%의 성능 향상을 도출하였다.

▸ **주제어**: 3D 특징 벡터, 데이터 증강, 영아, MFCC, 비언어적 소리

• First Author: JeongHyeon Park, Corresponding Author: Nammee Moon
*JeongHyeon Park (jh.park970609@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
*JunHyeok Go (junhyeok970306@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
*SiUng Kim (kimsiung990811@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
*Nammee Moon (nammee.moon@gmail.com), Dept. of Computer Science and Engineering, Hoseo University
• Received: 2023. 08. 25, Revised: 2023. 09. 14, Accepted: 2023. 09. 15.

http://www.ksci.re.kr pISSN:1598-849X | eISSN:2383-9945

I. Introduction

평범한 부모가 영아를 양육하기 위해서는 많은 노력과 시간이 필요한데, 이는 여러 가지 원인 중 영아가 일반적인 의사소통 방식이 아닌 비언어적 의사소통 방식인 울음이라는 수단을 쓰기 때문이다[1]. 영아는 울음을 통하여 본인의 모든 욕구를 표현하게 되는데, 이러한 욕구를 빠르게 해소해 주지 못한다면 아이에게는 코르티솔이 분비된다. 코르티솔이란 스트레스성 호르몬으로 성장하는 아이의 뇌에 치명적인 영향을 끼치게 되어 향후 아이가 성장하여 뇌 관련 질환이 발병할 확률이 높아지게 만든다[4]. 이에 따라, 영아의 욕구 파악을 위해 울음소리에 대한 연구가 활발히 진행되고 있다[5-7]. 또한 울음소리를 인공지능을 활용하여 분류하는 연구에서는 특징 벡터 사용이나 음성 데이터 증강 등 다양한 방법들이 사용되고 있다[2,3].

이전에 진행된 연구에 따르면 영아의 울음은 기본주파수, 소음대배음의 비율, 습관적 피치, 강도 4가지 음향학적 특성 평균의 차이가 나타나는 것으로 연구되었다[1]. 또한 Dunstan Baby Language에 따르면, 영아의 울음소리는 기본적으로 다섯 가지 특징을 가진 소리인 “Owh”,“Neh”, “Eairh”, “Heh”, “Eh”를 가지고 있다[8,9]. 각각의 소리는 영아가 원하는 욕구인 “피곤해”, “배고파”, “배아파”, “불편해”, “트림이 필요해”를 표현하는 방법이라고 한다. 이러한 특징들을 기반으로 한 소리를 이용해 영아의 상태를 분류하는 연구가 진행 중이며, 인공지능 기술을 활용하여 높은 성능을 나타내고 있다[2,3,9]. 최근 대부분의 연구에서는 음성 특징 벡터 중 Spectrogram, Mel Spectrogram, MFCC 중 1가지만을 사용하거나, 데이터 증강 기법을 Tempo, Pitch, Speed와 같은 단순한 기법을 사용하여 연구가 진행되었다.

이에 본 연구에서는 영아 울음소리 분류 모델 성능 향상을 위하여 3DV-ADA (3D Vector - Audio Data Augmentation)를 제안한다. 기존 연구에서 1가지의 특징 벡터를 사용하는 방식이 아닌 3가지 특징 벡터 모두를 추출하여 3차원 특징 벡터로 구성해 학습에 사용한다. 또한, 기존 음성 연구에서 사용하던 데이터 증강 방법인 Tempo, Pitch, Speed 방법이 아닌 이미지 처리 분야에서 사용되는 데이터 증강법 중 MixUp과 CutMix를 음성처리 분야에 맞게 변환하여 적용하고자 한다. 최종적으로 영아 울음소리 분류를 위해 적절한 특징 벡터와 증강 방법을 탐색해 정확도를 향상하고자 한다.

II. Preliminaries

1. Related works

1.1 Nonverbal sound classification based on deep learning

이전에 비언어적 소리를 딥러닝으로 분류하는 연구들이 진행되었다. 조류의 울음소리를 분류하는 연구에서는 소리를 Spectrogram으로 변환하여 ResNet-34, ResNet-50, AlexNet 모델을 사용해 실험을 진행했다. 실험 진행 결과 Test에서 평균 93%의 성능 즉, 오차범위 10% 이내의 결과를 도출하였다[10]. 이때 사용한 모델 중 AlexNet을 제외한 ResNet-34, ResNet-50에서 높은 성능을 보여주는 것을 실험 결과로 도출하였다[10].

음악 장르 분류 연구에서는 소리를 MFCC와 OSC 두 단일 특징 벡터를 합쳐서 재생성한 복합 특징 벡터를 사용하여 SVM(Support Vector Machine) 기계학습 알고리즘과 SRC(Sparse Representation Classification)를 사용해 분류를 진행하였다[11].

본 연구에서는 비언어적 소리를 분류하는데 높은 성능을 보인 ResNet50 모델을 사용한다. 또한 단일 특징 벡터가 아닌 복합 특징 벡터를 사용하고자 한다.

1.2 Data augmentation

데이터 증강은 데이터의 수집이 어렵거나 데이터의 다양성을 증가시킬 수 있는 방법이다. 특히, 이미지 처리 분야에서 활발하게 연구 되어졌다. 이미지 처리 분야에서 다양한 방식으로 데이터 증강을 진행하는데 Fig.1의 예시처럼 Flip, Crop, Color Conversion 등 여러 방법의 데이터 증강 기법이 존재한다.

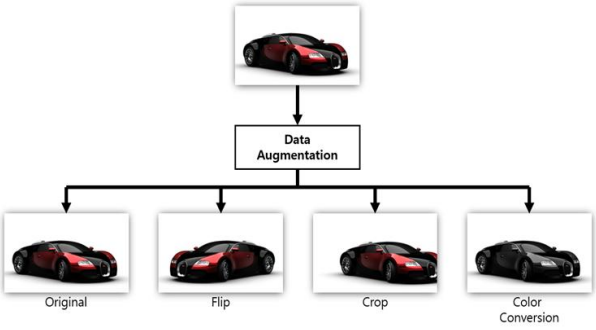

Fig. 1. System Architecture

이외에도 이미지 처리 분야에는 CutMix, MixUp 등의 여러 가지 방법을 이용한 데이터 증강 기법이 존재한다. MixUp은 Fig. 2 상단의 그림처럼 서로 다른 라벨값을 가

진 이미지 두 장을 선형 보간 방식으로 합치는 기법이다[12]. CutMix는 이미지의 부분 영역을 잘라내고, 비어있는 영역에 다른 라벨값을 가지는 이미지의 부분 영역을 합쳐 새로운 합성 이미지를 만드는 방법이다[13]. 이처럼 이미지 처리 분야에는 다양한 증강 기법들이 연구가 진행되었다. 마찬가지로 음성 처리 분야에서도 증강 기법을 활용한 여러 연구가 진행되어 왔다. 일반적으로 음성 데이터 증강은 Pitch, Tempo, Noise 등의 기법을 사용하여 데이터를 증강한다. 기존 선행 연구에서 연구된 음성 데이터 증강 기법 중 총 7가지 데이터 증강 방식 Pitch, Dynamic Value Change (DVC), Harmonic Percussive Source Separation (HPSS), Volume, Shift, Speed, Noise를 사용하여 비교해 본 결과 Pitch, DVC 증강 기법에서 성능 향상에 유의미한 결과를 보였다[14]. 또한, 다른 연구에서는 Speed와 Tempo 기반의 데이터 증강에서 가장 뛰어난 성능을 보여주었다[15].

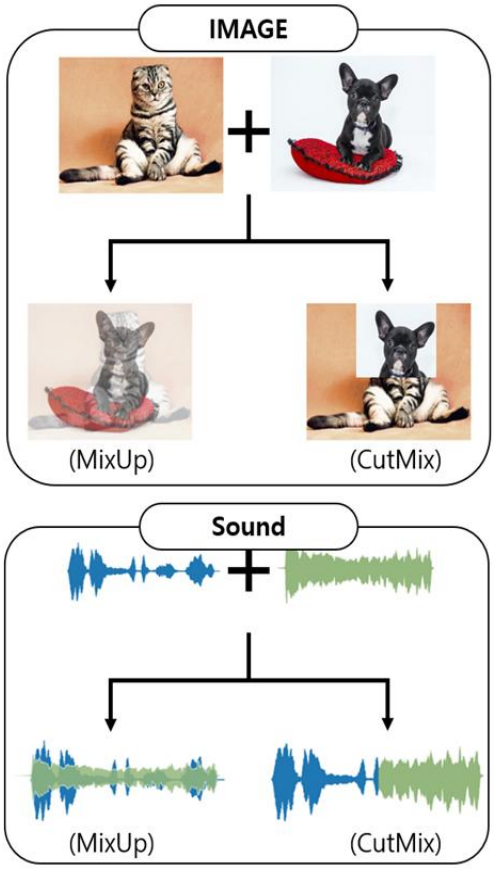

Fig. 2. System Architecture

본 연구에서는 기존 음성 데이터 증강 기법 중 Pitch와 Tempo, Shift를 사용한다. 또한 이미지 처리 분야에서 사용하는 방법 중 CutMix, MixUp 기법을 Fig. 2 하단의 그림과 같이 음성 처리 분야에 맞게 조정하여 데이터를 증강시키는 방법을 제안하고자 한다. MixUp기법은 이미지 처리 분야에서 사용하는 선형 보간 방식이 아닌 원본 음성에 다른 라벨값을 가진 음성 파일 하나의 볼륨을 낮추어 노이즈처럼 합성하는 새로운 MixUp기법을 사용한다. 또한 CutMix의 서로 다른 라벨값을 가지는 이미지를 합성하는 방식과는 같은 라벨값을 가지는 데이터를 둘로 나누어 앞뒤로 이어 붙이는 방식을 제시하고자 한다.

III. The Proposed Scheme

1. Crying classification through 3DV-ADA

1.1 3DV-ADA overview

울음소리 분류의 순서는 Fig. 3과 같다. 학습 데이터의 양을 늘리기 위해 기존 6초의 데이터를 1~3 초 단위로 잘라낸 후 데이터 전처리, 증강, 특징 벡터화를 거쳐 3D 벡터화를 시킨다. 이후에 음성 분류 모델에서 뛰어난 성능을 보여준 ResNet-50을 사용하여 모델 학습을 진행한다[11].

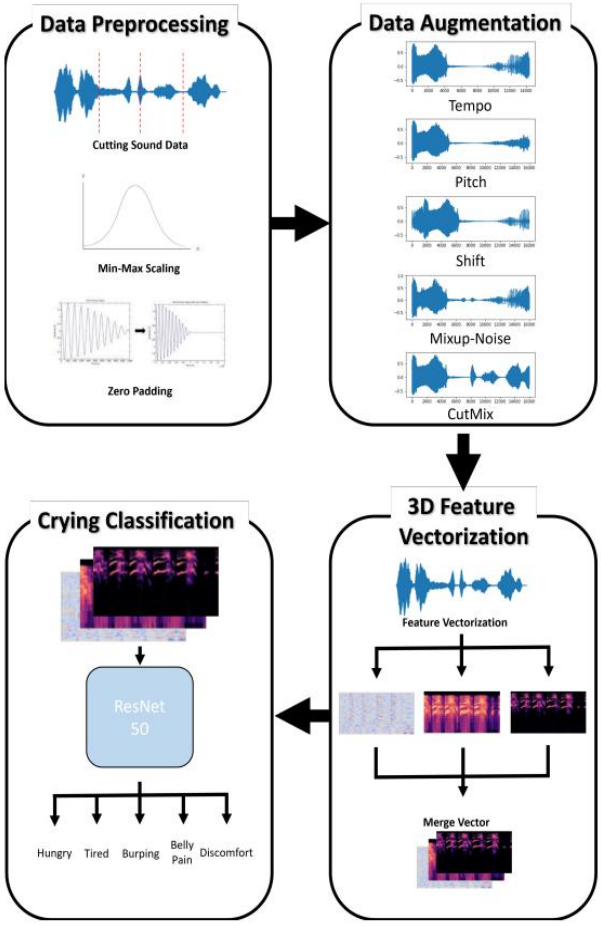

Fig. 3. 3DV-ADA Overview Diagram

1.2 Audio data augmentation

음성 데이터 증강에서는 총 5가지 증강 기법 Tempo, Pitch, Shift, Mixup-Noise, CutMix를 사용하여 음성 데이터 증강을 수행한다. 데이터의 증강 후 파형 예시는 Fig. 4와 같다.

1) Tempo

Tempo란 음성 신호의 속도의 빠르기를 뜻한다. 속도를 조절하기 위해서는 음성 신호의 파형을 압축 또는 늘려야 하는데 이 과정에서 진동수에도 영향을 미치게 된다. 이는 곧 음의 높낮이 즉 Pitch가 변하게 된다.

본 연구에서는 Samplerate의 Resample를 사용하여 속도를 원본 대비 90%와 110%로 증가 또는 감소한 후 PSOLA 알고리즘을 사용하여 Pitch를 정상화한다[16].

2) Pitch

Pitch란 음성 신호의 속도에 영향을 끼치지 않고 음의 높낮이를 변경하는 것을 말한다.

본 연구에서는 파이썬 라이브러리인 librosa의 Pitch_Shift를 사용하여 Pitch를 조절을 진행한다. Pitch의 파라미터 값은 (-1, 1)로 설정한다.

3) Shift

Shift는 시계열 데이터 형태인 음성 데이터를 지정한 만큼 파형을 왼쪽 또는 오른쪽으로 이동시키는 방법으로 Numpy의 Roll 함수를 사용하여 구현한다. Shift는 왼쪽 오른쪽으로 원본 대비 10%와 -10%로 이동시킨다.

4) Mixup-Noise

음성 분야에서 사용되는 백색 소음을 합성하는 노이즈 증강이 아닌 다른 라벨값을 가지는 음성 파일을 10%에서 30% 사이로 볼륨을 조절하여 합성하는 방식으로 진행한다.

5) CutMix

기존 이미지 처리 분야에서 사용하는 CutMix 방식은 음성 분류에서 적합하지 않다고 판단하여 같은 라벨을 가진 음성 파일을 반으로 나눈 후 서로 이어 붙이는 방식을 채택하여 실험을 진행하였다. 이는 시계열 데이터 형식을 가진 동일한 라벨값의 모든 학습용 음성 파일을 반으로 나눠 랜덤 셔플을 진행한 뒤에 합치는 방식으로 진행한다.

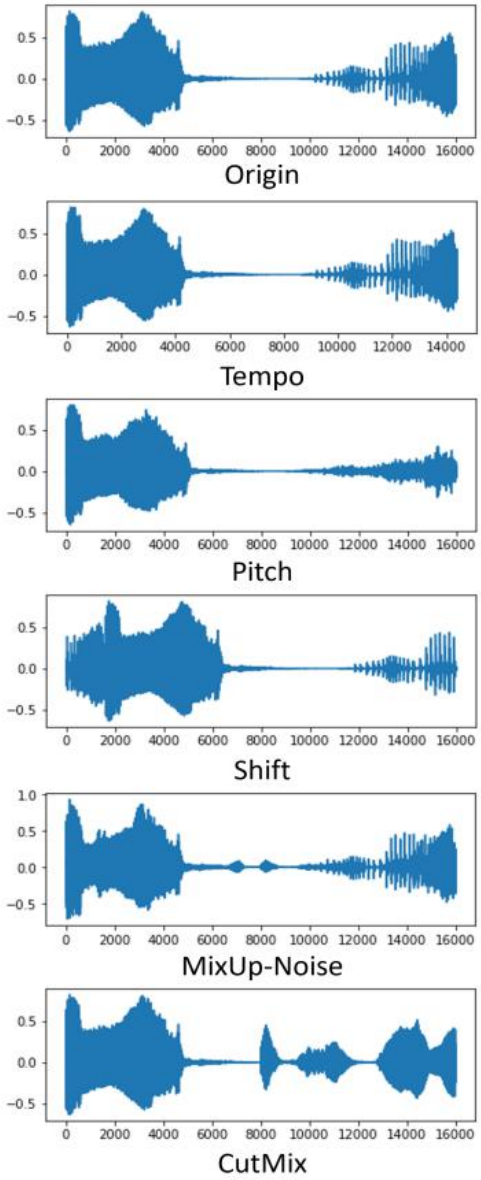

Fig. 4. Audio Augmentation Sample

1.3 Audio feature vectorization

음성 신호는 아날로그 데이터인 원시 데이터 형태를 가지고 있기 때문에 학습에 사용하기 위해서는 디지털 형태로 변환해주는 작업이 필요하다[17,18]. 이러한 작업을 특징 벡터화라고 하는데 대표적인 특징값 추출 방법에는 Spectrogram, MFCC (Mel-Frequency Cepstral Coefficient), Mel Spectrogram 등이 있다.

본 연구에서는 MFCC, Spectrogram, Mel Spectrogram 3가지의 특징 벡터를 사용하여 실험을 진행하였다. 음성 파일을 librosa 라이브러리를 사용하여 특징 벡터화를 진행하였다. 특징 벡터화에 특징값 파라미터를 (10, 80) 사이의 범위 내에서 진행하였으며 그 결과 (20, 30) 사이의 성능이 가장 좋은 결과값을 보여주었기

때문에, 본 연구에서는 특징값의 파라미터의 값을 20으로 설정하였다. 연구에 사용된 각 클래스 별 특징 벡터의 시각화 자료는 Fig. 5와 같다.

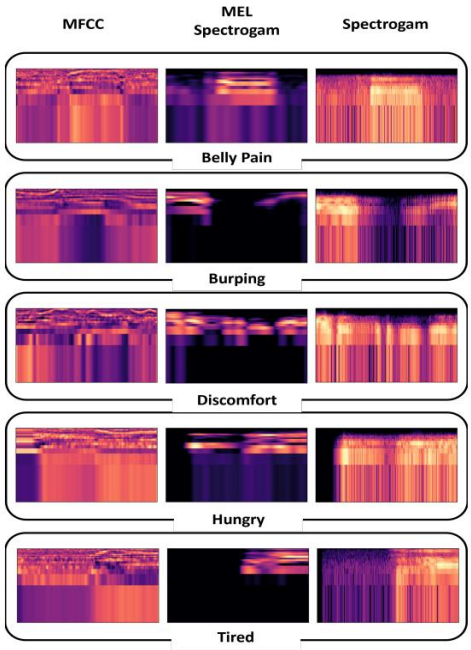

Fig. 5. Feature Vectorization Sample

1.4 3D feature vector

본 연구에서 데이터의 특징 개수를 20개로 설정하여 진행하였다. 하나의 특징 벡터는 20 * 600의 크기를 가지고 있으며 이러한 2차원 형태의 특징 벡터를 RGB 채널의 개념으로 각각의 채널을 담당하게 하여 3차원 특징 벡터로 구성하여 학습에 사용하였다. 이러한 3차원 특징 벡터의 모양은 (2* 20 *600) 또는 (3 * 20 * 600)으로 (3 * 20 * 600) 의 그림은 다음 Fig. 6과 같다.

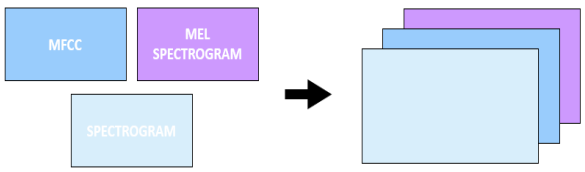

Fig. 6. 3D Feature Vector Example

2. Experiment

2.1 Audio dataset

학습과 평가에 사용되는 음성 데이터는 Donate a cry corpus, Dunstrun baby language를 사용하였다[9,19]. Donate a cry 데이터 세트는 'wav', 샘플레이트 8000hz, 6 초의 길이로 구성되어 있다. Dunstrun baby는 DVD내에서 영아의 울음소리를 추출하여 데이터 세트를 구성하였다. 데이터 세트의 클래스는 belly pain, burping, discomfort, hungry, tired 5종류로 분류되어 있다. 음성 데이터의 경우 어떠한 길이를 가졌을 때 최적의 성능을 보여주는지 찾아 내기 위해 데이터를 분할한다. 1초, 2초, 3초로 나눈 데이터와 기존 6초의 데이터를 가지고 3D 특징 벡터를 사용하여 학습을 진행한 결과는 Fig. 7과 같다. 오리지널 데이터인 6초의 경우 검증 세트의 F1-Score의 변동이 없는 모습을 보였다. 3초의 경우 모델의 성능이 낮은 모습을 보이고 그래프의 형태도 불안정한 모습을 보인다. 2초의 경우 검증 세트의 최고 F1-Score가 0.71로 가장 좋은 모습을 보여주지만, 수치의 변동이 커 안정성이 떨어지는 모습을 보여준다. 1초의 경우 준수한 Val F1-Score과 안정적으로 상승 폭을 그리는 그래프의 결과 보여주었다.

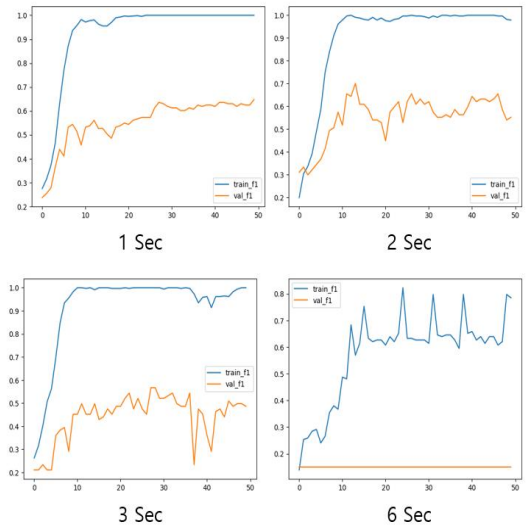

Fig. 7. Performance Evaluation Graphs

결론적으로, 본 연구에서는 준수한 성능과 안정적인 수치를 보여준 1초 단위로 음성 파일을 나누어 실험을 진행한다. 모든 음성 데이터를 1초 단위로 나눈 데이터의 수는 다음 Table 1과 같다.

Table 1. Experimental Data Set

Class	Train	Test
belly pain	124	14
burping	119	17
discomfort	165	23
hungry	206	24
tired	232	38

2.2 Experiment environment & evaluation methods

실험 환경은 Table 2와 같다. 성능 측정의 지표는 Precision, Recall, F1-Score를 사용한다. Precision은 모델이 양성으로 예측한 샘플 중에서 실제로 양성인 샘플의 비율을 계산한다. Recall은 실제로 양성인 샘플 중에서 모델이 정확하게 양성으로 예측한 샘플의 비율을 계산한다. F1-Score는 식(1)과 같이 Precision과 Recall의 조화 평균을 계산한 값이다.

Table 2. Experimental Environment

CPU	Intel i9-9900k
RAM	64G
VGA	GeForce Titan RTX 2way
CUDA	11.7
Pytorch	1.13.0
MODEL	ResNet-50 (Pretrained)

$$F1-Score = 2 \cdot \frac{precision \cdot recall}{precision + recall} \quad (1)$$

2.3 Evaluation of 3D feature vectors

3D 특징 벡터 성능 실험은 2차원의 특징 벡터만을 사용했을 때와 2차원 특징 벡터로 묶어 사용했을 때의 성능을 측정하였다.

단일 특징 벡터를 사용하였을 때의 가장 높은 성능을 보여준 모델은 Spectrogram 사용 모델로 테스트 스코어 기준으로 58.11%의 성능을 보여주었다. 하지만 MFCC 사용 모델은 33.62%로 가장 낮은 성능을 보여주었다.

각각의 2차원 특징 벡터를 합친 3차원 특징 벡터들의 학습 결과에서 단일 특징 벡터에서 낮은 점수를 기록한 MFCC와 다른 특징 벡터들을 묶어 학습에 사용하였을 때 수치가 단일 Spectrogram 모델보다 낮은 점수를 기록하였다. 하지만 Table 3의 6번 결과는 MEL-Spectrogram과 Spectrogram으로 합쳐 만든 모델의 경우는 기존 단일 특징 벡터 모델 중 가장 높은 점수보다 검증 세트 기준 4.14% 성능이 향상되었다.

Table 3. Performance Evaluation Results by Feature Vector

No.	Feature Vector	Val F1-Score	Test F1-Score
1	MFCC	37.57	33.62
2	MEL-Spectrogram	60.11	54.31
3	Spectrogram	61.64	58.11
4	MFCC + MEL-Spectrogram	58.38	54.31
5	MFCC + Spectrogram	54.33	56.03
6	**MEL-Spectrogram + Spectrogram**	**65.78**	**62.20**
7	MFCC + MEL-Spectrogram + Spectrogram	53.57	55.17

2.4 Evaluation of data augmentation techniques

데이터들을 각각 증강 기법에 적용하였을 때의 성능을 측정하였다. 학습 방법은 3D 특징 벡터 성능 중 가장 성능이 좋게 평가된 MEL-Spectrogram과 Spectrogram 특징 벡터를 사용하여 실험을 진행하였으며, 증강 방법을 Original 데이터에 적용하였을 때 성능을 측정하였다.

Table 4. Performance assessments and results for data augmentation methods

No.	Augmentation	Val F1-Score	Test F1-Score
1	Original	65.78	62.20
2	Pitch	68.78	60.34
3	**Tempo**	**70.52**	**67.24**
4	**Shift**	**72.14**	**66.47**
5	**CutMix**	**69.94**	**64.93**
6	MixUp-Noise	67.05	62.93

기존 Original 데이터만을 사용하였을 때보다 Tempo, Shift, CutMix의 증강 기법에서 테스트 F1-Score 기준 각각 5.04%, 4.27%, 2.73% 성능 향상을 보여주었다. 또한 MixUp의 경우 미미한 성능 향상을 보여주었다.

2.5 Experiment result

이전 실험에서 가장 우수한 성능을 보인 3D 특징 벡터(MEL-Spectrogram, Spectrogram)와 증강기법(Tempo, Shift, CutMix)을 사용하여 최종 모델 학습을 수행한 결과이다. Table 5는 최종 베스트 모델의 테스트 결과 값 표이다. 테스트 F1-Score는 75.86%로 원본 데이터를 Spectrogram으로 변환하여 구성한 모델의 테스트 성능보다 17.75% 향상된 것을 보여주었다.

Table 5. Final Experimental results

Class	Precision	Recall	F1-Score
belly pain	71.35	71.69	71.31
burping	83.52	59.91	69.05
discomfort	77.33	74.83	76.11
hungry	72.68	88.55	79.34
tired	77.58	79.16	78.87

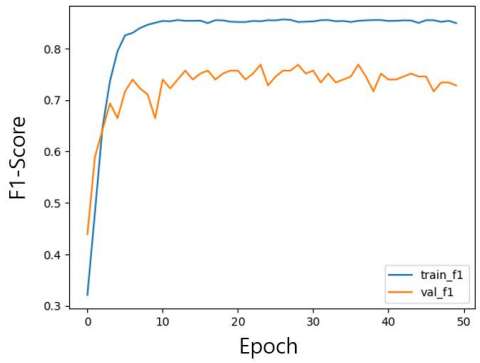

Fig. 8. Performance Evaluation Graph

또한 Fig. 8과 같이 학습이 진행될 때의 val_f1 스코어의 그래프 형태를 보면 안정적인 그래프를 확인할 수 있다.

IV. Conclusions

본 연구에서는 5가지 클래스(belly pain, burping, discomfort, hungry, tired)로 구성된 영아의 울음소리를 분류하기 위해 음성을 Spectrogram, Mel Spectrogram, MFCC으로 특징 벡터화한 후 2차원 특징 벡터를 3차원 특징 벡터로 합쳤다. 또한 Pitch, Tempo, Shift, CutMix, Mixup-Noise 데이터 증강 기법을 통하여 음성 데이터를 증강하였다. 이러한 음성 데이터들을 CNN기반의 ResNet50 모델을 활용하여 영아의 울음소리를 분류하였다. 울음소리 분류 모델의 성능 향상을 위해 특징 벡터 개수와 종류, 데이터 증강 기법들이 영향을 주는지를 실험하였다. 실험 결과 Spectrogram, Mel Spectrogram으로 구성된 3차원 특징 벡터가 2차원 특징 벡터보다 좋은 성능을 보여주었다. 데이터 증강 기법은 Pitch, MixUp-Noise를 제외한 Tempo, Shift, CutMix 기법을 적용하였을 때 성능 향상을 보여주었다. 최종적으로 우수한 특징 벡터들과 데이터 증강 기법들을 적용한 결과 성능 향상을 보였다.

본 연구는 영아를 보육함에 있어서 양육자의 부담을 덜어주고 영아의 스트레스를 줄여 뇌 관련 질환에 걸릴 확률을 낮춰줄 수 있다. 또한 비언어적 소리 분류 분야에서 여러 데이터 증강 기법을 사용하는 데에 있어서 도움을 줄 것으로 기대된다.

ACKNOWLEDGEMENT

This research was supported by the MIST(Ministry of Science and ICT), Korea, under the National Program for Excellence in SW supervised by the IITP(Institute for Information & communications Technology Promotion)" (2019-0-01834)

REFERENCES

[1] H. R Jang. “Acoustic characteristic of crying infants related to communicating intent” M.S. dissertation, Yonsei University, Korea, 2012. DOI:10.1111/j.1467-8624.1997.tb01947.x

[2] Lichuan Liu, Wei Li, Xianwen Wu and Benjamin X. Zhou, “Infant Cry Language Analysis and Recognition: An Experimental Approach”, IEEE/CAA Journal of Automatica Sinica 6.3, pp.778-788, 2019 DOI:10.1109/JAS.2019.1911435

[3] Chunyan Ji, Thosini Bamunu Mudiyanselage, Yutong Gao and Yi Pan, “A review of infant cry analysis and classification” EURASIP Journal on Audio, Speech, and Music Processing, pp.1-17. 2021.1 DOI:https://doi.org/10.1186 /s13636-021-00197-5

[4] K. J. Park, J. Y. Choi, Y. H .Kwon and J. H. Kim“Children's Cortisol Patterning at ChildCare Centers”, Korean Journal of Child Studies 28.6 pp.201-215, 2007

[5] P. S. Zeskind and B. M. Lester, "Acoustic features and auditory perceptions of the cries of newborns with prenatal and perinatal complications", Child Dev., Vol. 49, No. 3, pp. 580-589, Sep. 1978.

[6] T. Murry and P. Amundson, "Acoustical characteristic of infant cries: fundamental frequency", Child Lang., Vol. 4, No. 3, pp. 321- 328, Oct. 1977.

[7] WhyCry Technology, [Internet], http://www.why-cry.com, Apr. 05, 2019.

[8] Priscilla Dunstan, “dunstan baby language” [Internet], https://dunstan-babies.com/

[9] C . A. Bratan, M. Gheorghe, I. Ispas, E. Franti, M. Dascalu, S. M. Stoicescu, I. Rosca, F. Gherghiceanu, D. Dumitrache and L. Nastase, “Dunstan Baby Language Classification with CNN”, 2021 International Conference on Speech Technology and Human-Computer Dialogue (SpeD). IEEE, 2021. DOI: 10.1109/SpeD53181.2021.9587374

[10] M. J. Kang, Y. S. Kim, H. Y. Shin, and J. W Park, “Development of the Deep Learning System for Bird Classification Using Birdsong“, Journal of Knowledge Information Technology and Systems(JKITS), Vol. 15, No. 2, pp.195-203, April 2020 DOI : 10.34163/jkits.2020.15.2.005

[11] S. C. Lim, S. J. Jang S. P. Lee and M. Y. kim, "Multiple octave-band based genre classification algorithm for music recommendation", Journal of the Korea Institute of Information and Communication Engineering VOL 15 NO. 07 pp.1487-1494 2011. 07 DOI:https://doi.org/10.6109/jkiice.2011.15.7.1487

[12] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz"mixup: Beyond Empirical Risk Minimization", arXiv:1710.09412v2 [cs.LG] 27 Apr 2018 DOI:https://doi.org/10.48550/arXiv.1710.09412

[13] S. D. Yun, D. Y. Han, S. J. Oh, S. h. Chun, J. S. Choe and Y. J Yoo, "CutMix: Regularization Strategy to Train Strong Classifiers with Localizable Features", arXiv:1905.04899v2 [cs.CV] 7 Aug 2019 DOI:https://doi.org/10.48550/arXiv.1905.04899

[14] H. J. Choia and I.Y. Kwak, "Data augmentation in voice spoofing problem" The Korean Journal of Applied Statistics Vol. 34, No. 3, pp.449-460, 2021. DOI: https://doi.org/10.5351/KJAS.2021.34.3.435

[15] S. G. Lee, S. M, Lee,"Data Augmentation for DNN-based Speech Enhancement", Journal of Korea Multimedia Society Vol. 22, No. 7, pp. 749-758, July 2019 DOI:https://doi.org/10.9717/kmms.2019.22.7.74

[16] YUAN YUAN Wang and SHUN Yang. Speech synthesis based on PSOLA algorithm and modified pitch parameters. In: International Conference on Computational Problem-Solving. IEEE, p. 296-299, 2010

[17] J.D. Lim, S.W. Han, B.C. Choi, B.H. Chung "The Technology of the Audio Feature Extraction for Classifying Contents", Electronics and Telecommunications Trends 24.6, ,pp.121-132, 2009 DOI:10.22648/ETRI.2009.J.240613

[18] J. H. Park and, N. M Moon, "Design and Implementation of Attention Depression Detection Model Based on Multimodal Analysis". Sustainability, 14(6), 3569. DOI:https://doi.org/10.3390/su14063569

[19] Donate a cry-corpus, https://github.com/gveres/donateacry-corpus

Authors

JeongHyeon Park received the B.S. degrees in Computer Engineering from Hoseo University, Korea, in 2023. Park joined the Department of computer science at Hoseo University, asan, korea in 2023.

He is currently a Master in the Department of computer science, Hoseo Graduate School Asan, Korea, in 2023. He is interested in Data analysis and BigData

JunHyeok Go received the B.S. degrees in Computer Science and Engineering from Hoseo University, in 2023. B.S. Kim joined the Department of computer science at Hoseo University, asan, korea in 2023.

He is currently a Master in the Department of computer science, Hoseo university He is interested in Computer vision, object detection, image generation.

SiUng Kim received the B.S. degrees in Computer Science and Engineering from Hoseo University, in 2023. B.S. Kim joined the Department of computer science at Hoseo University, asan, korea in 2023.

He is currently a Master in the Department of computer science, Hoseo university He is interested in Computer vision, object detection, image generation.

Nammee Moon received B.S., M.S., and Ph.D. degrees in School of Computer Science and Engineering from Ewha Womans University in 1985, 1987 and 1998, respectively.

She served as an assistant professor at Ewha Womans University from 1999 to 2003. From 2003 to 2008, she is a professor of Department Digital Media, Graduate School of Seoul Venture Information. Since 2008, she is currently a professor in the Department of Computer Science and Engineering, Hoseo University. She is current research interests include Social Learning, HCI and User Centric Data, Big-data Processing and Analysis.