KIPS Trans. Comp. and Comm. Sys.
Vol.12, No.8 pp.235~242
ISSN: 2287-5891 (Print), ISSN 2734-049X (Online)

https://doi.org/10.3745/KTCCS.2023.12.8.235

Effect Analysis of Data Imbalance for Emotion Recognition Based on Deep Learning

Hajin Noh† · Yujin Lim††

ABSTRACT

In recent years, as online counseling for infants and adolescents has increased, CNN-based deep learning models are widely used as assistance tools for emotion recognition. However, since most emotion recognition models are trained on mainly adult data, there are performance restrictions to apply the model to infants and adolescents. In this paper, in order to analyze the performance constraints, the characteristics of facial expressions for emotional recognition of infants and adolescents compared to adults are analyzed through LIME method, one of the XAI techniques. In addition, the experiments are performed on the male and female groups to analyze the characteristics of gender-specific facial expressions. As a result, we describe age-specific and gender-specific experimental results based on the data distribution of the pre-training dataset of CNN models and highlight the importance of balanced learning data.

Keywords : Biased Data, XAI, LIME, Emotional Recognition, CNN

딥러닝기반 감정인식에서 데이터 불균형이 미치는 영향 분석

노 하 진† · 임 유 진††

요 약

최근 들어 영유아를 대상으로 한 비대면 상담이 증가함에 따라 감정인식 보조 도구로 CNN기반 딥러닝 모델을 많이 사용하고 있다. 하지만 대부분의 감정인식 모델은 성인 데이터 위주로 학습되어 있어 영유아 및 청소년을 대상으로 적용하기에는 성능상의 제약이 있다. 본 논문에서는 이러한 성능제약의 원인을 분석하기 위하여 XAI 기법 중 하나인 LIME 기법을 통해 성인 대비 영유아와 청소년의 감정인식을 위한 얼굴 표정의 특징을 분석한다. 뿐만 아니라 남녀 집단에도 동일한 실험을 수행함으로써 성별 간 얼굴 표정의 특징을 분석한다. 그 결과로 연령대별 실험 결과와 성별별 실험 결과를 CNN 모델의 사전 훈련 데이터셋의 데이터 분포를 바탕으로 설명하고 균형 있는 학습 데이터의 중요성을 강조한다.

키워드 : 데이터 불균형, XAI, LIME, 감정인식, CNN

1. 서 론

COVID-19 상황이 지속됨에 따라 대면 활동 기피 현상이 심화되었다. 이에 따라 의료 및 교육 상담 관련 업계에서도 비대면 상담 서비스 열풍이 불고 있다. 하지만 영유아를 대상으로 하는 행동 및 심리 문제 진단을 위한 비대면 상담에서는 영유아의 특성으로 인한 문제에 직면할 수 있다. 다시 말해서, 성인에 비해 언어 표현 및 사고 능력이 미숙한 영유아를 상담할 경우, 보다 정확한 영유아의 정서 파악을 위해 외부에서 관찰할 수 있는 비언어적 요소를 이용한 보조 도구 사용이 불가피하다. 따라서 영유아의 얼굴 표정을 인식하는 보조 도구로써 딥러닝 기술인 CNN(Convolutional Neural Network)을 활용하는 방식이 주목을 받고 있다.

그러나 이러한 CNN 모델을 훈련할 때 사용되는 얼굴기반 감정인식용 학습 데이터의 경우, 성인 데이터에 비하여 영유아 데이터가 상대적으로 그 양이 부족한 실정이며, 이로 인하여 훈련된 CNN 모델로 영유아의 감정을 인식할 시 성인 대상 감정인식에 비해 정답률이 상대적으로 낮다[1].

이렇듯 불균형한 학습 데이터는 모델의 성능 저하를 불러일으키기 때문에 훈련용 데이터는 각 조건마다 그 분포가 균일해야 한다. 그러나 사전 훈련된 모델을 사용할 때, 훈련 시 사용된 데이터에서 어느 특징이 불균형한지를 알기는 어렵다. 예를 들어, 남자 노인의 데이터셋으로 훈련한 감정인식 모델을 사용하여 여자 어린이 데이터 집단의 감정을 인식한

※ 이 성과는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2021R1F1A1047113).
※ 이 논문은 2022년 한국정보처리학회 ACK 2022의 우수논문으로 "영유아와 청소년의 얼굴표정기반 감정인식 성능분석"의 제목으로 발표된 논문을 확장한 것임.
† 준 회 원 : 숙명여자대학교 IT공학과 석사과정
†† 종신회원 : 숙명여자대학교 인공지능공학부 교수
Manuscript Received : December 19, 2022
Accepted : February 22, 2023
* Corresponding Author : Yujin Lim(yujin91@sookmyung.ac.kr)

경우, 산출되는 두 데이터 집단 간 정답률의 차이가 학습 데이터의 연령대 불균형 때문인지 또는 성별 불균형 때문인지는 훈련 데이터셋의 구성을 모르는 사용자는 알 수가 없다. 따라서, 이러한 인식 결과를 신뢰하는 것은 위험한 일이다. 이에 훈련 데이터셋을 알 수 없어도 학습 모델로부터 도출된 결과에 대한 근거를 규명할 수 있는 기술의 필요성이 많은 주목을 받고 있다.

영상이나 이미지 분석에 많이 사용되는 딥러닝 기법 중 하나인 CNN 모델은 블랙박스 모델(Black Box Model)로, 앞서 언급한 바와 같이 무수히 많은 은닉층의 데이터 처리 과정을 알기 어렵기 때문에 오직 모델의 최종 정확도만 가지고 신뢰성을 판단해야 한다는 한계가 있다. 이러한 한계점을 극복하기 위해 블랙박스 모델 해석을 도울 수 있는 설명 가능한 AI(eXplainable AI, XAI)가 최근 들어 새롭게 등장하였다. 이를 통해 인간이 딥러닝의 의사 결정 과정을 인식하고 그에 맞추어 훈련 데이터의 질을 개선할 수 있게 되었다.

본 연구에서는 영유아 및 청소년 감정인식을 위한 표정 데이터셋인 CAFE(Child Affective Facial Expression Set)(2-8세)[2, 3]와 NIMH Children's Emotional Face Picture Collection(NIMH-ChEFS)(10-17세)[4]를 대상으로 학습모델의 성능을 비교 분석한다. 또한 XAI 기법 중 블랙박스 모델과 유사한 기능을 하는 대리 분석 모델인 LIME(Local Interpretable Model-agnostic Explanation)을 사용하여 감정인식에 영향을 미치는 얼굴 특징을 시각화함으로써 연령대별 차이를 분석하고자 한다[5]. 더불어, 각 성별별 차이 또한 함께 살펴봄으로써 데이터 불균형의 문제점을 분석한다. 본 논문의 구성은 다음과 같다. 2장에서는 LIME 기법 설명과 관련 연구를 제시하며, 3장에서는 실험 시 사용된 데이터셋을 설명한 후, CNN을 사용하여 연령대별 인식률의 차이를 비교 및 분석하고, 그 원인을 파악하기 위해 LIME을 이용하여 시각화한다. 또한 시각화 결과를 기반으로 사전 훈련된 CNN 모델에 추가 훈련을 제안한다. 다음으로 각 데이터셋에서 남녀별 성능차이를 비교하고 실험 결과의 원인을 분석한다. 4장에서는 결론으로 마무리한다.

2. 실험 기법 및 관련 연구

2.1 LIME 기법

LIME은 블랙박스 모델의 비선형적인 함수의 일부분을 근사화하여 설명 가능한 모델을 제공한다. 설명성을 부여하기 위해 블랙박스 모델의 복잡성을 낮추어야 하며, 이를 위해 전체가 아닌 국소를 학습한다. 때문에 블랙박스 모델의 전체적인 판단 결과를 알 수는 없지만 특정 입력 데이터에 따라 특정 결괏값을 반환하는 근거를 알 수 있다. 근사화 모델로 쓰일 수 있는 대리 모델에는 선형 회귀, 의사결정나무 등이 있으며 텍스트, 테이블, 이미지 등 다양한 데이터 형식을 설명할 수 있다.

이미지 설명의 경우, 하나의 이미지를 같은 특성을 띠는 픽셀 그룹인 '슈퍼 픽셀'로 나눈다. 그 슈퍼 픽셀들의 모든 조합의 경우의 수만큼 이미지를 생성한 후, 생성된 이미지를 각각 입력으로 하여 나온 결괏값을 분석하여 어떤 조합이 가장 영향력이 높은지를 확인한다. 예를 들어, 'angry' 표정을 인식할 때 눈, 코, 입 3개의 슈퍼 픽셀로 나누었다고 가정한다. 눈과 코와 입이 모두 있는 사진, 눈, 코, 입 각자만 있는 사진, 눈과 코, 눈과 입, 코와 입 두 개만 있는 사진을 생성한다. 생성된 7장의 사진을 각각 입력으로 하였을 때 눈과 입이 있는 사진이 가장 정확도가 높다면 'angry'를 인식할 때에는 눈과 입이 가장 많은 영향을 미친다고 판단하는 것이다.

LIME은 다른 XAI 기법과는 달리 슈퍼 픽셀로 설명이 제공되므로 직관적이다. 또한, 필요한 연산 자원이 비교적 적다. 더불어, 블랙박스 모델의 종류와 무관하게 설명할 수 있어 적용 범위가 넓다. 경계가 분명한 슈퍼 픽셀로 나뉜다는 점에서 영향력이 있는 정확한 부분을 파악할 수 있고, 결과가 직관적이라는 점에서 사전 지식이 전무한 비전문가도 쉽게 알아볼 수 있기 때문에 LIME 기법을 실험에 사용하였다.

2.2 관련 연구

감정인식 분야에 딥러닝을 적용한 연구는 꾸준히 많은 관심을 받고 있으며, 이에 대한 결과로 다양한 기법들이 제안되었다. [6]은 CK+ 데이터셋을 활용하여 DNN(Deep Neural Network)과 CNN 기법을 기반으로 감정인식 성능을 비교하였다. 원본 이미지에서 얼굴 감지 후, OpenCV 라이브러리로 데이터를 정규화하여 입력으로 사용하였다. 실험 결과, DNN보다 CNN 모델이 더 높은 정확도를 보여주며 감정인식 분야에서 본격적인 CNN 도입의 근거를 제공하였다. 이후 [7]은 얼굴 표정 데이터셋인 CK+, KDEF, MUG, RAFD로 CNN 모델인 VGG(Visual Geometry Group)를 활용하여 훈련 및 미세 조정하였다. 그 결과 CK+, RAFD, MUG 데이터셋의 테스트 정확도가 최신 감정인식 모델에 비해 가장 높은 성능을 보였다. [8]에서는 얼굴 일부분이 가려지거나 일정하지 않은 주변 환경에서 발생하는 문제를 해결하기 위한 CNN 모델을 제안하였다. 얼굴 표정 데이터셋 MultiPIE, MMI, CK+, DISFA, FERA, SFEW 및 FER2013을 사용한 결과 높은 정확도를 달성하였다. [9]는 합성곱 신경망과 잔차 블록을 포함하는 완전 심층 신경망을 제안하였다. CK+, JAFFE 데이터셋을 이용하였고 그 성능이 이전 모델보다 향상됨을 보여준다. 하지만, 미성년 얼굴 표정 데이터셋은 성인 데이터셋에 비해 접근성이 낮고 양적 한계가 있기 때문에 이러한 연구들은 성인 데이터에 적은 비율의 미성년 데이터가 혼합된 데이터를 이용하여 학습하거나 성인으로만 이루어진 데이터셋을 사용하고 있다. 따라서, 성능이 높은 감정인식 모델이라도 영유아 및 청소년 데이터를 인식할 시, 성인의 경우와 비슷한 수준의 정확도를 가진다고 하기 어렵다.

이러한 문제점을 해결하기 위한 연구가 수행되었으며 [1]에서는 2세 미만 영아의 대규모 데이터셋을 직접 구축하고 행복, 슬픔, 중립으로 감정 등급을 구성한 후 이에 적합한

CNN 모델 제안 및 학습을 통해 약 87.90%의 높은 인식률을 이끌어내었다. 하지만, 2세 미만의 연령대에서 표현할 수 있는 감정이 제한적이며 실질적으로 필요한 영유아 및 청소년기를 포함하지 못했다. [10]은 다양한 인종으로 구성된 6세에서 12세 사이의 12명의 비디오 얼굴 표정 데이터셋 'LIRIS-CSE'를 구축하였다. 또한, CNN 모델에 전이 학습을 사용하여 데이터셋에 적용한 결과 약 75%의 분류 정확도를 달성하였다. 하지만, 자연스러운 아이의 감정을 이끌어내는 과정에서 윤리적인 이유로 인해 '분노', '두려움', '슬픔' 등의 부정적인 감정 데이터는 소수라는 한계가 존재한다.

감정인식 분야에 XAI를 적용한 연구 중 하나로 [11]은 Grad-CAM을 이용하여 얼굴 표정을 인식한 후 실시간으로 감정을 표현하는 부분을 히트맵으로 표현하였다. [12]는 어린이 데이터셋인 LIRIS-CSE와 직접 제작한 인도, 방글라데시, 네팔 국적의 7-10세 어린이 데이터셋에 XAI를 적용하였다. XAI 기법은 Grad-CAM, Grad-CAM++, SoftGrad를 사용하였다. 이러한 연구들은 시각적으로 설명을 제공하였지만, Grad-CAM의 히트맵의 경계가 모호하고 추상적이기 때문에 설명력에 한계가 있을 수 있다. 경계가 분명한 슈퍼 픽셀을 나타내는 LIME을 CNN 모델에 연결하여 설명 가능한 하이브리드 프레임워크(HEF)를 제안한 연구도 존재한다[13]. 하지만, CNN 모델이 사용하는 학습 데이터의 분포에 따라 설명성이 달라진다는 한계가 있다.

데이터 불균형성을 설명하는 연구도 진행되었다. 얼굴 표정 데이터셋 중 하나인 BU-4DFE 훈련 데이터셋을 사용한 연구[14]는 성별 불균형의 영향을 알아보기 위한 실험을 수행하였다. 여성 데이터, 남성 데이터, 여성과 남성이 섞인 데이터를 사용하여 각각 훈련시킨 모델로 여성, 남성, 여성과 남성이 섞인 데이터를 인식한 결과에 LIME을 적용하였다. 실험 결과는 데이터 불균형을 줄이고 다양성을 갖춘 데이터를 구축해야 함을 보여주지만, 연령대의 불균형은 설명하지 않았다.

본 연구에서는 위 한계점들을 극복하기 위해 영유아와 청소년기 각각으로만 이루어진 데이터셋을 활용한다. 각 데이터셋은 최소 5가지 이상의 감정 클래스를 가진다. 성인 데이터셋에서 높은 정확도를 가진 CNN 모델이 영유아 데이터셋에서 낮은 정확도를 가지는 이유를 슈퍼 픽셀로 나누어 표현하는 LIME을 통해 설명한다. 또한, 동일한 데이터셋을 남녀로 나누어 동일한 실험을 수행한 결과도 포함하여 연령대별 정확도와 성별별 정확도가 차이나는 이유를 설명한다.

3. 연령대 및 성별기반 성능 분석

3.1 데이터셋

본 실험에서는 다음과 같은 두 가지 데이터셋을 사용하였다. 먼저, CAFE(Child Affective Facial Expression Set) 데이터셋은 약 2-8세 아동의 얼굴 표정 데이터셋이다[2, 3]. 감정인식 학습을 위해 사용될 수 있으며 분노(angry), 혐오감(disgust), 두려움(fear), 행복(happy), 중립(neutral), 슬픔(sad), 놀라움(surprise)의 7가지 클래스로 분류된다(Fig. 1). 이 외에도 입의 개폐 여부, 성별, 인종 등 다양한 조건의 데이터가 존재한다.

Fig. 1. Example of CAFE Dataset

Fig. 2. Example of NIMH Dataset

다음으로, NIMH-ChEFS(NIMH Children's Emotional Face Picture Collection)는 약 10-17세 청소년의 얼굴 표정 데이터셋이다[4]. CAFE와 동일하게 감정인식 학습 시 활용될 수 있으며 분노(angry), 두려움(afraid), 행복(happy), 중립(neutral), 슬픔(sad)의 5가지 클래스로 분류된다(Fig. 2). 화면을 직접 응시하는 데이터(direct)와 다른 곳을 응시하는 데이터(averted)가 존재한다.

3.2 연령대 기반 감정인식 성능 분석 결과

CAFE 데이터셋 669장 중 얼굴이 정상적으로 인식되지 않은 혐오 감정 클래스에 속한 1장을 제외하여 668장을 실험에 활용하였다. NIMH-ChEFS 데이터셋은 모든 데이터인 534장을 사용하였다. 각 데이터셋의 남녀 감정 클래스별 데이터 개수는 Table 1과 Table 2와 같다.

감정인식 실험에는 VGGFace2 데이터셋으로 사전 훈련된 CNN 모델인 MobileNet을 사용하였다. 이 모델을 AffectNet 데이터셋 인식에 적용한 결과 약 64.71%의 정확도를 나타내었다[15].

CAFE와 NIMH 데이터셋을 활용하여 감정인식 실험을 시행한 결과는 Fig. 3과 같다. NIMH 데이터셋에 존재하지 않는 disgust, surprise 감정 클래스는 비교에서 제외하였다.

CAFE 데이터셋의 각 클래스 정확도는 angry(0.553), disgust(0.818), fear(0.667), happy(0.902), neutral(0.452), sad(0.491), surprise(0.772)이며, 전체 정확도는 약 0.671이다.

Table 1. Number of Data by Gender Class(CAFE)

	angry	disgust	fear	happy
Male	35	34	22	41
Female	77	75	62	82
total	112	109	84	123
	neutral	sad	surprise	total
Male	44	19	17	212
Female	82	38	40	456
total	126	57	57	668

Table 2. Number of Data by Gender Class(NIMH)

	angry	fear	happy	neutral	sad	total
Male	38	40	38	39	38	193
Female	66	67	70	72	66	341
total	104	107	108	111	104	534

반면, NIMH 데이터셋의 각 클래스 정확도는 angry(0.827), fear(0.888), happy(1), neutral(0.793), sad(0.798)이며, 전체 정확도는 약 0.861이다.

happy 클래스를 제외한 나머지 감정에서 두 데이터셋의 정확도의 차이가 큼과 동시에 영유아 데이터셋(CAFE)의 전체 정확도가 상대적으로 낮은 것을 알 수 있다. 이는 MobileNet 모델이 성인 데이터를 위주로 구성된 학습 데이터를 사용하였으므로, 비교적 성인의 특성과 비슷한 청소년기의 얼굴 표정(NIMH)을 더 잘 인식한 것으로 분석된다.

이러한 감정인식 성능차이의 근거를 밝혀내기 위해 정답으로 처리된 데이터에 LIME을 적용하였다. 해당 감정인식의 근거가 되는 얼굴 표정의 일부분을 색으로 표현하였으며, 원본 데이터에는 빨간색으로, 단순화 맵에는 노란색으로 표현하였다. LIME을 통해 생성된 이미지를 감정 클래스별로 10장씩 겹쳐 색이 진하게 표현된 공통적인 부분을 파악하였다(Fig. 4). 단, happy 클래스는 각 데이터셋의 정답률 차이가 미미하였으므로 제외하였다.

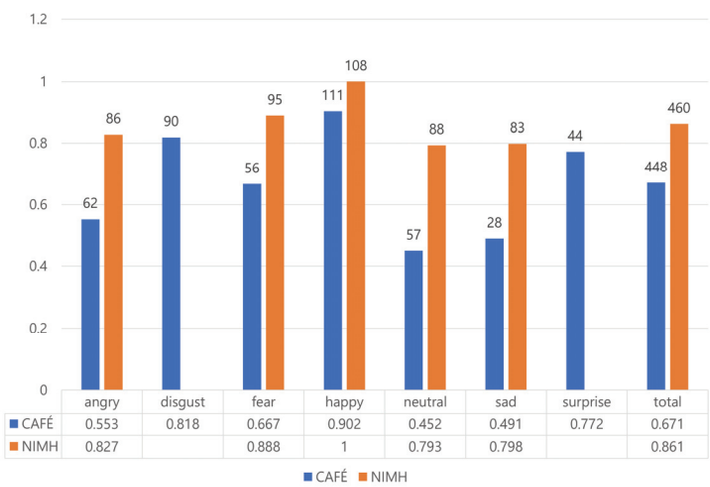

Fig. 3. Number of Correct Answers and Answer Rate by Using CNN Model with CAFE and NIMH Dataset

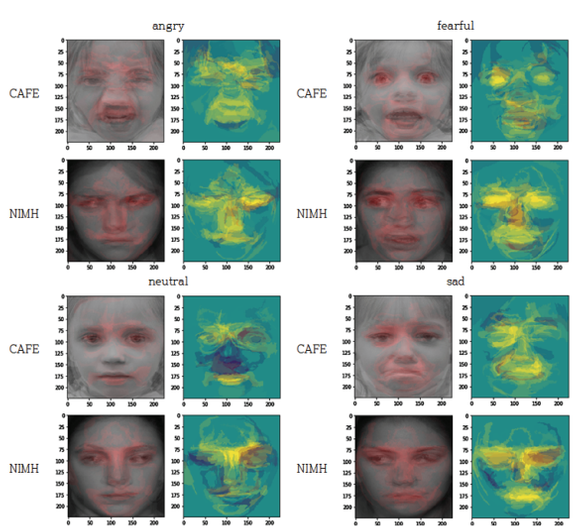

Fig. 4. Age-specific Experimental Results Using LIME based on CAFE and NIMH Datasets

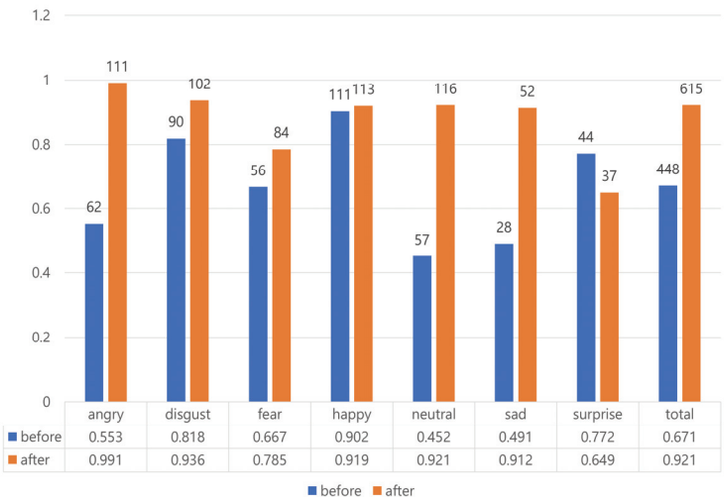

Fig. 5. Number of Correct Answers and Answer Rate Before and After Additional Training(CAFE)

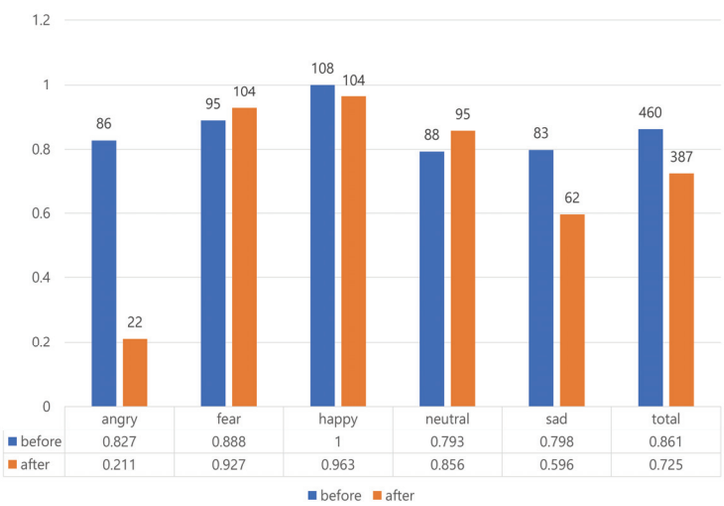

Fig. 6. Number of Correct Answers and Answer Rate Before and After Additional Training(NIMH)

영유아 집단 대상인 CAFE 데이터셋의 시각화에서 감정 클래스에 영향을 미친 부분이 통일되어 있지 않고 각각 얼굴 전체적으로 분산되어 있으며 특히 콧대 부분은 영향이 적다는 것을 알 수 있다. 반면 청소년 집단 대상인 NIMH 데이터셋의 시각화에서는 눈과 눈썹, 미간, 콧대 부분에 집중되어 있다. neutral을 제외한 나머지 클래스에서는 입 부분까지 영향을 미친다. 이러한 분석 결과는 사용한 모델에 영유아의 얼굴 표정 특징이 성인에 비해 적게 훈련되어있음을 보여준다.

본 논문에서는 영유아 감정인식의 정답률 향상을 위해 실험에 사용한 CNN 모델을 대상으로 영유아의 특징을 학습할 수 있도록 CAFE 데이터를 추가로 훈련하였다. 실험 전후 정답률을 비교한 결과는 Fig. 5와 Fig. 6으로 나타내었다.

추가 훈련 전 CAFE 전체 정답률은 0.671에서 추가 훈련 후 0.921로 증가하였으며 angry 클래스는 0.553에서 0.991, neutral 클래스는 0.452에서 0.921, sad 클래스는 0.491에서 0.912로 대폭 증가하였다. 반면 NIMH 데이터셋의 fear 클래스는 0.888에서 0.927, neutral 클래스는 0.793에서 0.856으로 소폭 증가하였다. 하지만 angry 클래스는 0.827에서 0.211, happy 클래스는 1에서 0.963, sad 클래스는 0.798에서 0.596으로 오히려 감소하였다. 특히 angry와 sad 클래스의 정답률은 대폭 감소하였는데, 이는 CAFE 데이터셋에서 정답률이 크게 증가한 클래스이다. 추가 훈련 전에는 성인 특징을 위주로 인식했기 때문에 NIMH 데이터셋의 정답률이 높았다면 추가 훈련 후에는 영유아의 특징 비율이 더 커져 CAFE 데이터셋에서의 정답률은 높아지고 NIMH 데이터셋의 정답률이 낮아지는 것이다.

3.3 성별기반 감정인식 성능 분석 결과

원래의 MobileNet CNN 모델로 CAFE와 NIMH 데이터셋을 사용하여 인식한 성별별 감정인식 결과는 Table 3과 Table 4와 같다. NIMH 데이터셋에는 disgust, surprise 클래스가 존재하지 않으므로 제외하였다.

Table 3. Number of Correct Answers by Gender(CAFE)

	angry	disgust	fear	happy
Male	20(0.571)	32(0.941)	11(0.5)	38(0.927)
Female	42(0.545)	58(0.773)	45(0.726)	73(0.890)
	neutral	sad	surprise	total
Male	22(0.5)	8(0.421)	13(0.765)	144(0.679)
Female	35(0.427)	20(0.526)	31(0.775)	304(0.667)

Table 4. Number of Correct Answers by Gender(NIMH)

	angry	fear	happy	neutral	sad	total
Male	30(0.789)	34(0.85)	38(1)	24(0.615)	31(0.816)	157(0.813)
Female	56(0.848)	61(0.910)	70(1)	64(0.889)	52(0.788)	303(0.888)

CAFE 데이터셋의 정확도는 angry 클래스에서 0.571(남)과 0.545(여), disgust 클래스에서 0.941(남)과 0.773(여), fear 클래스에서 0.5(남)와 0.726(여), happy 클래스에서 0.927(남)과 0.890(여), neutral 클래스에서 0.5(남)와 0.427(여), sad 클래스에서 0.421(남)과 0.526(여), surprise 클래스에서 0.765(남)과 0.775(여)이며 전체 정확도는 0.679(남)와 0.667(여)를 보였다. fear 클래스에서 0.226의 정확도 차이를 제외하고 나머지 클래스에서는 남녀의 정답률이 0.2 이하로 크게 차이 나지 않는 것을 알 수 있다.

NIMH 데이터셋의 정확도는 angry 클래스에서 0.789(남)와 0.848(여), fear 클래스에서 0.85(남)와 0.910(여), happy 클래스에서 1(남)과 1(여), neutral 클래스에서 0.615(남)와 0.889(여), sad 클래스에서 0.816(남)과 0.788(여)이며 전체 정확도는 0.813(남)과 0.888(여)를 보였다. NIMH 데이터셋 역시 neutral 클래스에서 0.274의 정확도 차이를 제외하고는 다른 클래스에서 큰 차이가 없다.

성별에 따라 얼굴 표정에서 차이가 발생하는 부분을 확인하기 위해 정답으로 분류된 데이터에 LIME을 적용하였다. 연령대별 시각화에서와 동일하게 해당 감정인식에 영향을 미친 부분을 빨간색, 노란색으로 나타내었고 생성된 결과 이미지를 클래스별로 10장씩 겹쳐 공통적인 부분을 파악하였다(Fig. 7과 8). 단, CAFE 데이터셋에서 sad 클래스에 속하는 남성 데이터는 정답이 8장뿐이었으므로 8장으로 구성되었다. 남녀 모두의 정답률이 매우 높았던 happy 클래스는 제외하였다.

각 데이터셋에서 남녀의 뚜렷한 차이점은 나타나지 않았다. CAFE 데이터셋의 neutral 클래스에서 남녀 모두 미간을 주로 인식하였으며 나머지 클래스에서도 남녀 모두 눈, 미간, 입 등 얼굴의 전체적인 부분이 영향을 미친 것으로 나타났다. NIMH 데이터셋에서도 모든 클래스를 관찰하였을 때, 감정에 영향을 미친 남녀의 얼굴 부분이 거의 동일한 것을 알 수 있다.

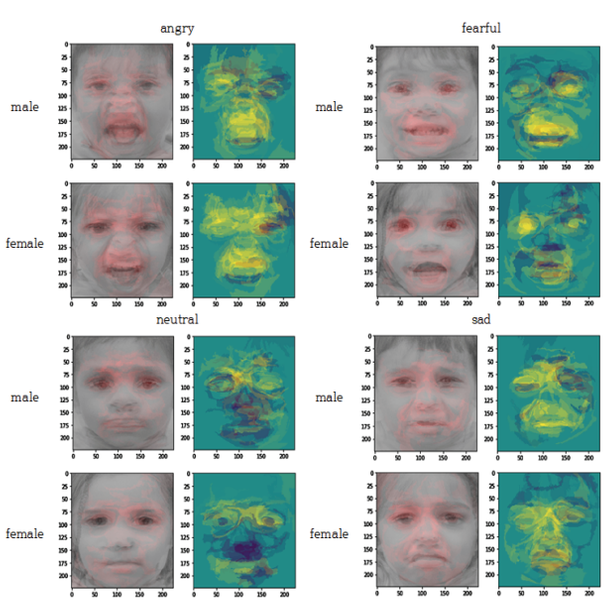

Fig. 7. Gender-specific Experimental Results Using LIME based on CAFE Dataset

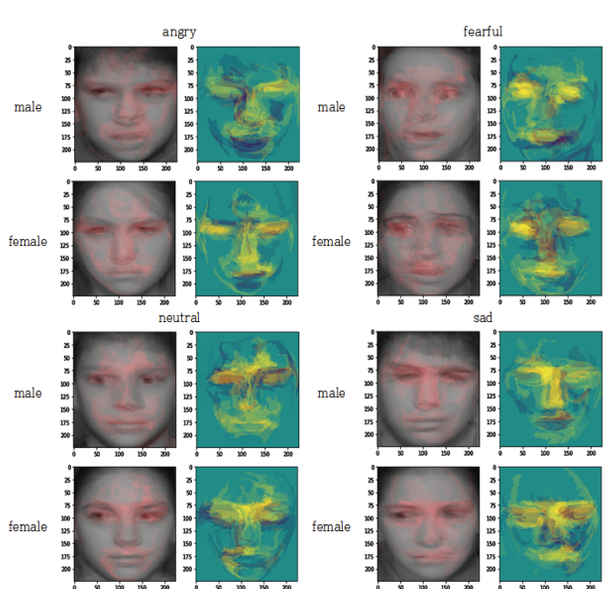

Fig. 8. Gender-specific Experimental Results Using LIME based on NIMH Dataset

3.4 연령대와 성별기반 실험결과 분석

CAFE와 NIMH 데이터셋을 비교한 실험으로 연령대별 성능 차이가 존재하며 이는 학습된 영유아의 특징이 성인에 비해 부족하기 때문임을 보였다. CAFE와 NIMH 데이터셋 내에서 각각 남녀를 비교한 실험에서는 성별에 따른 성능 차이가 거의 존재하지 않았으며 실제로 영향을 미치는 부분도 거의 동일함을 확인했다. 실험에 사용된 CNN 모델을 사전 훈련한 VGGFace2 데이터셋의 데이터 분포로 두 결과를 해석할 수 있다.

VGGFace2 데이터셋은 자세, 나이가 다른 다양한 조건에서의 얼굴 데이터를 포함한다. 하지만, 0-12세의 데이터는 전체 데이터의 약 0.01%이며 13-17세의 데이터는 약 0.02%로 두 집합을 합한 0-17세의 데이터는 전체의 0.1%도 되지 않는다[16]. 13-17세 집단의 훈련 데이터도 충분하지는 않지만 2차 성징이 드러나는 청소년은 성인과 비슷한 특징을 가지는 것으로 간주한다면 영유아 집단은 데이터 불균형으로 인해 차이가 발생하는 것으로 생각할 수 있다. 따라서, 청소년 데이터는 비교적 잘 인식한 반면 훈련이 거의 되지 않고 성인의 특징과 다른 특징을 가지는 영유아 감정인식 성능은 저조할 수밖에 없다는 결론을 도출할 수 있다.

연령대 분포와는 달리 성별 분포로는 약 59.3%가 남성 데이터, 나머지 30.7%가 여성 데이터로 이루어져 있어 두 집단 간의 균형이 양호한 것으로 보인다[16]. 비슷한 비율의 훈련 데이터를 사용함으로써 남녀의 특징이 균형있게 학습되어 성별 간 얼굴 표정 특징 차이가 거의 존재하지 않는 것으로 설명할 수 있다.

4. 결 론

본 연구에서는 CNN 학습 모델을 사용하여 영유아와 청소년의 감정인식 정확도를 측정하고 그 결과를 LIME으로 시각화하여 정확도 차이의 원인을 분석하였다. 감정 분석 시 영유아는 얼굴 전체가 판단 결과에 영향을 미치는 반면, 청소년은 눈과 미간, 코로 이어지는 특정한 부분이 영향을 미치는 것으로 나타났다. 이러한 분석 결과를 기반으로 CNN 모델에 영유아의 특징이 잘 훈련되어 있지 않은 것을 확인하고 영유아 데이터셋으로 추가 훈련을 진행하여 정확도를 개선하였다. 추가로, 각 데이터셋 내에서 남녀의 정확도를 각각 측정하고 LIME을 통해 시각화하였으나 특별히 구분되는 특징은 없음을 확인하였다. 이는 CNN의 사전 훈련 데이터셋인 VGGFace2 데이터셋의 성별 분포는 비슷한 반면 연령대 분포에서 영유아의 데이터가 매우 적었기 때문임을 확인했다.

이렇듯 학습 데이터 불균형 문제는 감정인식 성능에 큰 영향을 미친다. 따라서, 정확한 감정인식을 위해 연령, 성별, 인종 등 뚜렷한 특징을 가진 그룹을 세분화하고 그룹 간 균형을 고려하여 데이터셋을 구축해야 한다. 또한, 필요 시 인식에 중요한 영향을 미치는 요인을 분석하고 상대적으로 부족한 그룹의 데이터를 추가로 훈련하는 방법을 사용하여 인식률을 향상시킬 수도 있다.

본 실험을 통해 비대면 상담에서 전문가가 보다 정확하고 신속한 진단을 내리도록 근거를 제공해줄 수 있을 것으로 기대된다. 또한, 균형 잡힌 양질의 데이터를 구축하여 성능을 더욱 향상시킨다면 비대면 상황에서뿐만 아니라 가정 및 학교에서도 전문가의 도움 없이 간이 자가 진단에 활용할 수 있을 것이다.

References

[1] Q. Lin, R. He, and P. Jiang, "Feature guided CNN for Baby's facial expression recognition," *Hindawi Complexity*, Vol. 2020, pp.1-10, 2020.

[2] V. LoBue and C. Thrasher, "The Child Affective Facial Expression (CAFE) set: Validity and reliability from untrained adults," *Frontiers in Psychology*, Vol.5, pp.1-8, 2015.

[3] The Child Affective Facial Expression (CAFE) Set. Databrary, http://doi.org/10.17910/B7301K.

[4] H. L. Egger, D. S. Pine, E. Nelson, E. Leibenluft, M. Ernst, K. E. Towbin, and A. Angold, "The NIMH Child Emotional Faces Picture Set (NIMH-ChEFS): A new set of children's facial emotion stimuli," *International Journal of Methods in Psychiatric Research*, Vol.20, No.3, pp.145-156, 2011.

[5] H. Noh and Y. Lim, Proceedings of the Annual Conference of Korea Information Processing Society Conference (KIPS) 2022, Vol.29, No.2, pp.700-702, 2022.

[6] H. Jung et al., "Development of deep learning-based facial expression recognition system," *2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV)*, Mokpo, Korea (South), pp.1-4, 2015.

[7] A. Fathallah, L. Abdi, and A. Douik, "Facial expression recognition via deep learning," *2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA)*, Hammamet, Tunisia, pp.745-750, 2017.

[8] A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, Lake Placid, NY, USA, pp. 1-10, 2016.

[9] D. K. Jain, P. Shamsolmoali, and P. Sehdev, "Extended deep neural network for facial emotion recognition," *Pattern Recognition Letters*, Vol.120, pp.69-74, 2019.

[10] R. A. Khan, A. Crenn, A. Meyer, and S. Bouakaz, "A novel database of children's spontaneous facial expressions (LIRIS-CSE)," *Image and Vision Computing*, Vol.83-84, pp. 61-69, 2019.

[11] T. A. Araf, A. Siddika, S. Karimi, and M. G. R. Alam, "Real-time face emotion recognition and visualization using Grad-CAM," *International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies*, Bhilai, India, pp.21-22, 2022.
[12] M. Rathod et al., "Kids' Emotion Recognition Using Various Deep-Learning Models with Explainable AI," *Sensors (Basel)*, Vol.22, No.20, pp.8066, 2022.
[13] M. Deramgozin, S. Jovanovic, H. Rabah, and N. Ramzan, "A hybrid explainable ai framework applied to global and local facial expression recognition," *IEEE International Conference on Imaging Systems and Techniques*, Kaohsiung, Taiwan, pp.24-26, 2021.
[14] C. Manresa-Yee and S. Ramis, "Assessing gender bias in predictive algorithms using eXplainable AI", *XXI International Conference on Human Computer Interaction*, Malaga Spain, pp.22-24, 2021.
[15] HSEmotion (High-Speed face Emotion recognition) library [Internet], https://github.com/HSE-asavchenko/face-emotion-recognition.
[16] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, Xi'an, China, pp.67-74, 2018.

노 하 진

https://orcid.org/0000-0003-3065-9407
e-mail : hajins@sookmyung.ac.kr
2023년 숙명여자대학교 IT공학과(학사)
2023년 ~ 현 재 숙명여자대학교 IT공학과 석사과정
관심분야 : 딥러닝, 강화학습

임 유 진

https://orcid.org/0000-0002-3076-8040
e-mail : yujin91@sookmyung.ac.kr
2000년 숙명여자대학교 전산학과(박사)
2013년 일본 Tohoku University, Department of Information Sciences(박사)
2022년 ~ 2015년 수원대학교 정보미디어학과 부교수
2016년 ~ 현 재 숙명여자대학교 인공지능공학부 교수
관심분야 : 지능형 시스템, IoT, Edge Computing