

# 앙상블 학습 알고리즘과 인공지능 표정 인식 기술을 활용한 사용자 감정 맞춤 힐링 서비스

양성연<sup>1</sup>, 홍다혜<sup>2</sup>, 문재현\*

<sup>1</sup>숙명여자대학교 컴퓨터공학과

<sup>2</sup>숙명여자대학교 소프트웨어융합과

\*한국기술거래사회

yeon1234551@gmail.com, stellaj0511@gmail.com, smjhoon@gmail.com

\*교신저자(Corresponding Author)

## Using Ensemble Learning Algorithm and AI Facial Expression Recognition, Healing Service Tailored to User's Emotion

<sup>1</sup>Yang seong-yeon, Dahye Hong<sup>2</sup>, Jaehyun Moon\*

Department of Computer Science, Sookmyung women University

<sup>2</sup>Department of Software and Communications Engineering, Sookmyung women University

\*Korea Technology Transfer Agents Association

### 요 약

The keyword 'healing' is essential to the competitive society and culture of Koreans. In addition, as the time at home increases due to COVID-19, the demand for indoor healing services has increased.

Therefore, this thesis analyzes the user's facial expression so that people can receive various 'customized' healing services indoors, and based on this, provides lighting, ASMR, video recommendation service, and facial expression recording service. The user's expression was analyzed by applying the ensemble algorithm to the expression prediction results of various CNN models after extracting only the face through object detection from the image taken by the user.

### 1. 서론

현대인들의 끊임없는 경쟁과 비교의 삶 속에서 '힐링'이라는 표현은 빼놓을 수 없는 키워드이다. 이렇기에 새로운 상품과 서비스 아이템을 추진할 때도 여지없이 "힐링" 개념이 들어간다. [1] 코로나19로 집에 있는 시간이 늘어나게 되면서 실내 힐링 서비스의 수요가 증가했다. 정보통신정책연구원(KISDI)의 "세대별 OTT 서비스 이용 현황"에 따르면 OTT 이용률은 지난 2019년 41.0%에서 2021년 81.7%까지 상승하였다. 이처럼 사람들은 집에서의 힐링 서비스를 필요로하므로 사람의 표정으로 감정을 분석하여 집에서 힐링 서비스를 제공 받을 수 있는 서비스를 제안한다.

### 2. 선행 연구

AI 스피커 누구(NUGU) 전용 명상 서비스 '누구 마음보기'는 총 41종의 명상 콘텐츠를 제공한다. 이용자가 AI 스피커에 대고 '아리아~ 마음보기에서 힐링 명상 들려줘'라고 말하면 '누구'가 카테고리별 또는 시간 별로 5-15분 분량의 명상 콘텐츠를 자동으로 추천한다. 하지만 이는 사용자가 직접 힐링 서

스를 찾아야 한다는 점을 극복하지 못했다. 따라서 본 논문의 서비스는 사용자의 표정만으로 그에 맞는 힐링 서비스를 제공한다는 점에서 차별화된 서비스라 할 수 있다. 또한 기록 서비스와 ASMR 서비스, 조명 서비스 등의 서비스를 제공함으로써 다양한 방법으로 사용자의 힐링을 돕는다.

현재 CNN의 발전에 따라 표정을 통해 기분을 분류하는 연구들이 활발히 이루어지고 있다. 하지만 대부분의 연구에서는 외국인들의 얼굴 이미지들이 활용되었다. 이에 해당 모형을 한국인들에게 적용 시 정확도가 떨어진다는 문제점이 있다. 이러한 문제점을 해결하기 위해 한국인들의 얼굴 표정을 학습시킨 모델을 탑재한 사용자 맞춤 힐링 서비스를 제안한다.

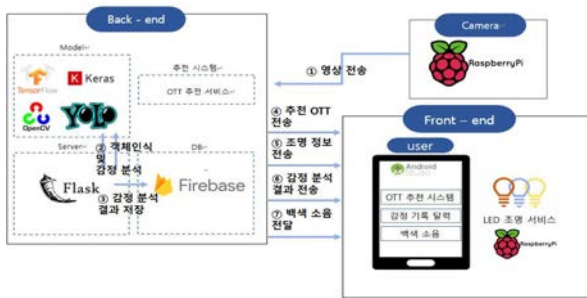
### 3 제안 시스템

#### 3.1 시스템 개요

본 시스템은 사용자의 표정을 분석하여 사용자의 기분에 따라 다양한 힐링 서비스를 제공한다. 사용자는 제공된 서비스를 모바일 애플리케이션을 통해 조절할 수 있으며 감정을 기록할 수 있다.

### 3.2 시스템 구성도

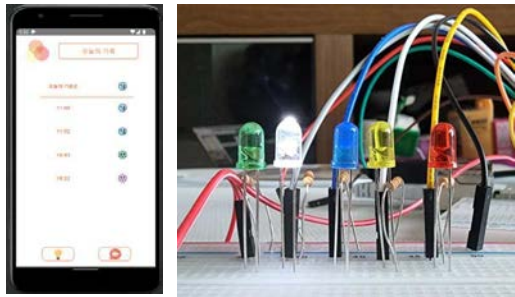
본 시스템은 딥러닝 서버, IoT기기, 앱(APP), 패널, 조명으로 구성되어 있다. [그림 1]는 전체 시스템의 구성도이다. IoT 기기인 라즈베리파이로 사용자는 원하는 시간에 사진을 촬영한다. openCV의 오브젝트 검출 알고리즘인 haar cascade 알고리즘이 촬영된 프레임에서 사용자의 얼굴을 파악하고 추출한다. 그 후, 프레임은 rescale 되어 촬영 시간 정보와 함께 파이어 베이스에 저장된다. 저장된 데이터는 실시간으로 딥러닝 서버로 전달되며 딥러닝 서버는 전달받은 프레임에서 표정 값을 도출한다.



[그림 1] 서비스 구성도

사용자에게 직접적인 서비스를 제공하는 앱(APP)은 TCP/IP 통신을 기반으로 라즈베리파이와 실시간으로 상호 작용하며 딥러닝 서버를 통해 도출된 표정 결과를 기반으로 사용자에게 개인화된 힐링 서비스를 제공한다. 사용자는 앱을 통해 조명 조절, ASMR 재생 등 제공받은 힐링 서비스를 제어할 수 있다.

### 3.3 기기조작



[그림 2] 애플리케이션을 통한 조명 조절

사용자는 앱(APP)에 회원 가입과 로그인을 하면 기록 서비스와 조명 서비스, 그리고 ASMR 서비스를 이용할 수 있다. 해당 앱(APP) 왼쪽하단의 조명 버튼을 클릭하면 [그림 2]의 오른쪽 사진과 같이 조명이 켜진다. 사용자가 IoT를 활용해 사진을 찍으면 해당 기록이 서버로 전송되고, 서버에서 해당 기록을 앱으로 가져온다. 기록을 확인한 후 조명서비스와 ASMR 서비스, 동영상 추천 서비스가 기록에 맞게 제공된다. 모든 서비스는 가장 최신의 기록에 맞게 제공된다.

## 4. 인공지능 알고리즘 선정 및 개선

### 4.1 표정 데이터 수집

본 서비스의 감정 분류에 사용할 이미지 처리 모델을 훈련 시키기 위해 학습(training) 데이터로 AI Hub에서 제공하는 “한국인 감정 인식을 위한 복합 영상” 개방 데이터를 사용하였다. Data set 다운로드 이후, 표정이 잘못 분류되거나 애매한 데이터를 제거하였다. Train, Validation, Test Data set은 각각 8:2:1 비율로 구성하였다. 본 서비스가 실내에서 개인의 표정에 따라 제공되는 맞춤형 서비스라는 점을 고려하여 표정은 ‘행복’, ‘무표정’, ‘화남’, ‘슬픔’, ‘놀람’으로 총 5가지로 구성하였으며, 수집 데이터의 수는 <표 1>과 같다.

	Happy (기쁨)	Neutral (무표정)	angry (화남)	Sad (슬픔)	Surprise (놀람)
Train Data	3,000	3,000	3,000	3,000	3,000
Validation Data	860	860	860	860	860
Test Data	430	430	430	430	430

<표 1> 표정별 Data-set 개수

Overfitting을 해결하고자 회전 범위, 수직/ 수평으로의 평행이동 범위, 수평 대칭, 확대를 설정해 데이터를 증식하였다.

### 4.2 표정 데이터 전처리

전처리 유무	x	o	x	o
	Val Acc[%]	Val Acc[%]	Val Loss	Val Loss
VGG-16	42.70	84.86	1.2437	0.4580
ResNet-50	44.45	67.12	1.3758	0.9354
DensNet-50	45.99	81.18	1.1331	0.5556
EfficientNet-b0	46.05	83.75	1.1287	0.4792
Inception-V3	43.98	86.10	1.6525	0.4094
Inception-V4	20.20	82.72	1.6365	0.4869
Inception-ResNet	25.34	85.45	1.6435	0.4389

<표 2> 전처리 전과 후

<표2>를 보면 사람의 얼굴 데이터를 전처리 없이 사용하였을 때 모든 종류의 모델에서 정확도가 낮고 loss도 크다. 이는 모델 훈련에 쓰일 이미지 데이터의 얼굴 위치, 각도가 다르거나 얼굴 외의 부분의 비율이 클 때 얼굴 특성 외의 다른 특성들이 훈련될 확률이 높아지기 때문이다. 따라서 모델의 정확도를 높이기 위해 프레임에서 동일한 형태의 정면 얼굴을 추출하는 데이터 전처리 과정을 진행하였다.

얼굴 검출에는 Haar Cascades 알고리즘을 사용하였다. [그림 3]은 해당 알고리즘으로 얼굴을 검출하는 과정을 보여준다. 위 알고리즘을 사용하여 얼굴에 대한 위치를 Rect(x,y,w,h) 형태로 출력 받아 얼굴 영역을 검출하였고 이후, 데이터 사이즈를 동일하게 변경해 주었다. [3] 데이터 전처리 이후 모든

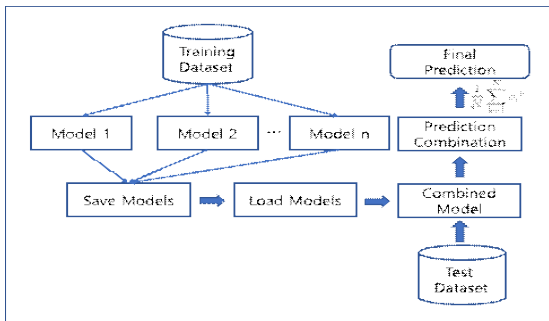
종류의 모델이 정확도는 높아지고 loss는 낮아졌다.



[그림 3] 얼굴 검출 과정(저자 얼굴 검출 샘플)

### 4.3 사용자 표정(감정) 알고리즘

감정 분류를 위해 CNN 알고리즘 중 VGG-16, EfficientNet-B0, DenseNet-50, InceptionNet-V3, InceptionNet-V4, InceptionNet v4+res 모델을 학습하였다. 각각의 모델들의 예측값은 <표3>과 같다. 학습 시 총 Epoch은 모두 50으로 설정하였으며, Optimizer는 Adam으로 진행하였다.



[그림 4] Ensemble Methods

앙상블 방법은 여러 모델들이 예측한 결과인 예측값을 사용하여 더 높은 정확도를 도출해내는 방법으로, 본 논문에서는 여러 앙상블 방법 중 [그림 4]와 같은 Soft voting 방식을 사용하였다.[4] Soft voting 방식은 weak learner들의 예측 확률값의 평균을 사용하는 방법으로 하나의 모델을 사용하였을 때보다 높은 확률로 예측 가능하다. <표3>를 보면 각각의 모델들의 정확도는 약 63%에서 82%까지 그 편차가 심하고 높은 정확도를 보이지는 않지만 앙상블 기법 중 Average Rule을 활용하여 여러 모델의 예측값을 평균한 결과 가장 높은 정확도가 89%로 1개의 모델의 예측값보다 훨씬 높은 예측값을 보였다. <표 3>은 4개 또는 5개의 모델의 예측 값을 평균한 결과 정확도가 가장 높은 4개의 조합을 나타낸 것으로 서비스에는 정확도가 가장 높은 EfficientNet-b0,

VGG-16, Inception-V3, Inception-ResNet 앙상블 모델(89.44%)을 사용 하였다.

## 5. 결론

### 5.1 감성적 인공지능 힐링 서비스 제공

본 서비스는 객체 탐지 알고리즘과 표정 인식 알고리즘을 활용해 사용자의 감정에 맞는 힐링 서비스를 제공하도록 구현하였다.

### 5.2 학습 모델의 정확도 및 인식을 제고

다양한 CNN 모델들을 비교하고 앙상블 기법을 활용하여 82% 정확도에서 90% 정확도까지 높여 해당 알고리즘을 사용하여 표정 인식률을 높였다. 이러한 학습 결과를 활용하여 사용자에게 다양한 힐링 서비스를 제안하는 시스템을 구현하였다.

### 5.3 맥내 감성 서비스 구현

맥내 조명 서비스를 통해 집안의 분위기를 변화시킴으로써 힐링할 수 있도록 하였고, 기록 서비스를 통해 자신의 표정변화를 보며 하루의 마음을 정리하는 시간을 가져 힐링할 수 있다. 또한 ASMR 서비스를 이용함으로써 정서적 안정을 가져올 수 있어 지친 현대인의 마음을 위로하는 힐링 서비스가 될 것이다.

## 참고문헌

- [1] 이승겸, "코로나관련 집콕으로 '집 꾸미기' 물품 수입액 역대 최대" 일간NTN 국제신문, 2022.4.28
- [2] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol. 1. Ieee, Kauai, HI, USA, 2001, p9
- [3] Sung-Wook Park, Jong-Chan Kim and Do-Yeon Kim, "A Study on Classification Performance Analysis of Convolutional Neural Network using Ensemble Learning Algorithm", Journal of Korea Multimedia Society Vol. 22, No. 6, June 2019(pp. 665-675)

본 논문은 과학기술정보통신부 정보통신창의인재양성사업의 지원을 통해 수행한 ICT멘토링 프로젝트 결과물입니다.

<표 3> Top 5 accuracy according to ensemble generation methods

	VGG-16	EfficientNet-b0	DenseNet-50	Inception-V3	ResNet-50
Test Acc [%]	82.56	81.02	79.3	66.37	63.58
	EfficientNet-b0	EfficientNet-b0	DenseNet-50	DenseNet-50	EfficientNet-b0
	VGG-16	VGG-16	EfficientNet-b0	EfficientNet-b0	VGG-16
	Inception-V3	Inception-V3	Inception-V3	Inception-V3	ResNet-50
	Inception-ResNet	Inception-V4	Inception-V4	Inception-ResNet	Inception-V3
Test Acc [%]	89.44	89.30	88.51	88.19	88.09