



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

석사학위논문

인간의 복합 감정을 예측할 수 있는
EfficientNet 아키텍처 기반 시스템 및 알고리즘 연구

계명대학교 대학원
의학과

김민영

지도교수 이종하

2022년 2월

인간의 복합 감정을 예측할 수 있는 EfficientNet 아키텍처 기반 시스템 및 알고리즘 연구

김민영
2022년
2월

인간의 복합 감정을 예측할 수 있는
EfficientNet 아키텍처 기반 시스템 및 알고리즘 연구

지도교수 이 중 하

이 논문을 석사학위 논문으로 제출함

2022년 2월

계명대학교 대학원
의 학 과

김 민 영

김민영의 석사학위 논문을 인준함

주 심 구 정 훈

부 심 이 종 하

부 심 김 선 칠

계 명 대 학 교 대 학 원

2 0 2 2 년 2 월

목 차

1. 서 론	1
2. 재료 및 방법	17
3. 성 적	31
4. 고 찰	36
5. 요 약	38
참고문헌	40
Abstract	44
국문초록	46

표 목 차

<표 1> 학습에 사용한 심박변이도 데이터	20
<표 2> 심박변이도 데이터 세부내용	21
<표 3> Confusion matrix 척도	32

그림 목차

<그림 1> 자율신경계의 길항작용	5
<그림 2> CNN의 시초인 LeNet의 기본구조	10
<그림 3> ImageNet 데이터 셋에서의 EfficeintNet 모델 성능	13
<그림 4> EfficientNet Compound Scaling 상수비	14
<그림 5> 기본 EfficientNet-B0 모델 구조	15
<그림 6> 학습에 사용된 표정 이미지 데이터	18
<그림 7> 생체신호 데이터 샘플링 과정 예시	23
<그림 8> 학습에 사용된 데이터	24
<그림 9> 표정 이미지 및 생체신호를 이용하여 감정 분석하기 위한 모델 비교	26
<그림 10> 실험 1의 학습 곡선	29
<그림 11> 실험 2의 학습 곡선	30
<그림 12> 모델이 분류한 감정 예측 결과	34
<그림 13> 실험에 따른 모델 정확도 비교 그래프	35

1. 서론

1.1 연구배경:

최근 인공지능의 기술 발달과 함께 감정 인식 분야가 중요한 연구 분야로 대두되고 있다. 감정이란 어떠한 현상이나 사건을 접했을 때, 마음에서 일어나는 느낌이나 기분을 뜻한다. 사람의 감정은 제품이나 생활환경 등에 응용되어 인간의 삶을 보다 쾌적하고 편리하게 만들기 위해 사용될 수 있는 지표이다. 감정은 인간에게 기본적으로 학습, 의사소통, 의사결정과 같은 인식과 일상적인 활동에 영향을 끼친다. 말이나 몸짓, 표정 그리고 다른 비언어적 단서를 통해 표현될 수 있고, 많은 생리학적 신호들 또한 인간의 감정 상태에 대한 정보를 전달한다(1,2).

감정은 매우 역동적이고 주관적이다. 제한된 언어만을 사용해 감정을 파악하는 것은 어렵기 때문에 감정이 실제로 어떻게 느껴지는지를 전달하는 것은 중요하다. Gallup 2017 Global Emotions Report에 따르면, 인간 행동의 70%는 추론이 아닌 감정에 기반을 두고 있다. 삶의 질과 의사결정 능력을 향상시킬 수 있는 감정 인식 시스템 개발이 중요하다는 통계가 존재한다.

앞선 연구들에서는 감정 인식을 위해 표정 이미지를 통해서 감정 상태를 파악했다. 얼굴 표정은 감정 인식을 위해 가장 널리 사용되는 양식이기 때문에 표정 이미지로 감정 상태를 파악하는 선행연구들이 많이 존재한다. 하지만 표정은 통제가 가능하며, 조작할 수 있으므로 진정으로 느껴지는 정서 상태를 표정으로 파악하기는 어렵다. 또한, 표정 이미지를 통한 선행 연구에서는 공포스러운 얼굴 이미지의 오인식 비율이 높았다고 발표했으며, 연구를 통해 슬픈 감정의 표정과 화난 감정의 표정을 구분 짓기 어려웠고 감정 판단이 제대로 되지 않는다는 것을 알 수 있었다. 이렇듯 표정 이미지만으로 감정을 파악하기에는 감정을 숨기거나 오인식할 경우가 있으

므로 부정확하다.

이러한 문제점을 해결하기 위해 최근 지속적인 측정이 가능하고 개인의 통제를 벗어난 생체신호를 이용한 감정 인식 연구가 활발히 진행 중이다. 인간의 감정 상태와 생체신호는 강한 상관관계가 있다는 많은 선행연구가 존재하며, 생체신호는 비교적 간단하게 획득할 수 있으므로 감정 인식을 위해 생체신호를 사용한다.

생체신호를 사용하여 감정 인식을 진행한 연구에서는 심박수, 피부전도도, 호흡, 근전도, 뇌파 등 다양한 생체신호를 사용하여 감정 인식에 사용하였다. 그중 한 연구에서는 인간의 감정에서 휴식과 활동 상태를 파악하기 위해 심박수와 피부전도도, 호흡의 생체신호가 효과적이라고 밝혔다. 스트레스 정도를 구별하기 위해서는 심박수와 호흡 신호를 사용한다고 밝혔으며, 결론적으로 감정 인식을 위한 생체신호로서 심박수와 호흡이 효과적이라고 말했다. 또 다른 연구에서는 생체신호로 뇌파를 사용하였는데 이는 뚜렷한 증가율을 보여주지 못했으며, 뇌파는 감정을 나타내는 주파수의 연구가 부족하다고 밝혔다. 또, 뇌파는 잡음이나 배경의 변화와 같은 제한을 많이 받아 감정 인식의 생체신호로써 적합하지 못하다는 연구 결과를 보였다(3,4).

이로써 본 연구에서는 선행연구들에서 공통적으로 효과적이라고 밝힌 심박변이도를 생체신호로 이용하여 감정 인식의 정확도를 높이고자 한다. 다른 생리학적 변수와는 달리 심박변이도는 교감신경계와 부교감신경계 활동에 대한 정보를 제공하여 감정 조절의 억제 및 흥분 과정에 대한 추론이 가능하다. 자율신경계는 생리적 및 심리적 스트레스에 따라 심박수 및 심박출량을 조절하는 데 중요한 역할을 하기 때문에 인위적이지 않고 자연스러운 감정 상태 획득이 가능하다. 이러한 이유로 인해 감정 인식을 위한 생체신호로 심박변이도가 가장 적합하다고 판단하였다(5,6).

또한, 생체신호와 표정 이미지를 결합하여 감정 인식을 진행한 몇몇 연구에서는 생체신호에 의한 감정 인식률과 표정 이미지에 의한 감정 인식률, 각각의 인식률 보다 두 가지를 함께 조합했을 때 가장 높은 인식률을 보였

다고 밝혔다. 하지만 이러한 생체신호와 표정 이미지 데이터를 같이 분석하여 감정 인식 분류에 사용한 연구는 거의 없을 뿐만 아니라 몇몇 연구들이 존재하지만 대부분 데이터에 기반한 feature extraction 과정이 필요한 등 분석 과정이 복잡하고 분류 모델의 효율성이 현저히 떨어진다(7).

따라서 본 연구에서는 딥러닝을 위한 신경망 아키텍처인 CNN(Convolution Neural Network)을 사용하여 생체신호와 표정 이미지를 함께 분석해 감정 상태를 예측하는 알고리즘을 개발하였다. 특히 본 연구에서는 CNN 기반의 높은 효율성을 보이는 EfficientNet-B0를 사용하였다. CNN은 이미지와 영상 처리에 유용하게 쓰이는 일반적인 구조를 제공하며, 대부분의 영상 인식에서 뛰어난 성능을 보이고 학습을 위한 파라미터 수가 적어 효율성이 뛰어나다(8,9,10). 따라서 본 논문에서는 생체신호와 표정 이미지 데이터를 함께 분석하여 감정을 예측하기 위해 CNN 기반의 모델을 이용한 방법을 제안한다.

1.2 배경 이론:

1.2.1 자율신경계:

자율신경계는 체성신경계와 달리 자신의 의지로 제어할 수 없는 말초신경계를 일컬어 말한다. 자율신경계는 신진대사인 소화, 호흡, 땀 같이 의식적으로 제어할 수 없는 기능에 대해 관여한다. 자율신경계는 부교감신경계(sympathetic nervous system)와 교감신경계(parasympathetic nervous system)라는 두 가지 신경계로 다시 나누어지며, 이들은 하나가 활발해지면 다른 하나는 침체되는 방식인 길항작용을 통해 제어된다.

교감신경계는 활발한 활동을 통해 몸을 움직이거나 공포와 두려움 같은 상황에 처해 스트레스가 많아지면 활발해지게 된다. 교감신경계의 활성화로 스트레스 대처에 필요한 반응과 에너지 공급이 나타나고, 그에 따라 혈압과 심장박동수가 높아지고 동공이 확대되며 소름이 돋는 현상이 나타난다.

다. 반대로 편안한 상태가 되면 부교감 신경계가 활성화 된다. 활성화가 되면 혈압과 심장박동수가 낮아지게 되고 소화기관에 혈액이 많이 돌게 된다. 이후, 소화효소분비가 활발해지고 에너지를 확보하는 방향으로 온몸이 작동하게 된다.

구분		신경	
		교감신경	부교감신경
상태	심장박동	촉진	억제
	혈관	수축	확장
	혈압	상승	강하
	소화관 운동	억제	촉진
	소화액 분비	억제	촉진
	침분비	억제	촉진
	눈동자	확대	축소

그림 1. 자율신경계의 길항작용

1.2.2 심박수:

심박수는 1분 동안 심장의 박동수를 의미하고, 안정시 심박수와 최대 심박수로 구분을 한다. 안정시 심박수는 움직임이 거의 없는 상태에서 측정 한 것으로, 건강한 성인의 안정적인 심박수는 60~100회 정도이다. 심장은 심박수가 낮을수록 한 번에 더 많은 양의 혈액을 신체에 전달할 수 있음을 의미하며, 감정 상태나 나이, 복용하고 있는 약품 등 여러 가지 요소들이 심박수에 영향을 준다.

안정시의 심박수는 맥박을 확인함으로써 간단하게 측정할 수 있다. 목의 가쪽, 팔꿈치 안쪽, 손목, 사타구니, 발등 등 동맥이 피부와 가까운 곳에서 측정이 가능하다. 심실이 수축하는 기간을 수축기라 하며, 심실이 이완하는 기간을 확장기라 한다. 이러한 사건들을 통틀어 심장주기라 하며, 1분간에는 약 70주기가 반복된다. 심실 수축기가 시작되면 심방-심실 간의 판막이 닫히며 첫 번째 심음을 낸다. 수축기가 끝나면 심실 내의 압력이 떨어지고 심실-동맥 간의 판막이 닫히며 두 번째 심음을 낸다.

심장근 수축에 따르는 전기적인 변화를 피부에 부착한 전극들로 검출, 증폭, 기록한 것을 심전도, EKG 또는 ECG라고 하며, 파형 내의 여러 파들은 P, Q, R, S, T파 등으로 불린다.

심박수가 정상적이지 않다는 것은 심장에 문제가 있다고 판단할 수 있으며, 심장질환을 의심해볼 수 있다. 그 예로는 심장 혈관의 수축으로 충분한 혈액이 심장에 공급되지 않는 심근경색, 심박수 정상 수치가 지속적으로 벗어나고 박동이 심하게 불규칙적인 부정맥, 심장 기능에 이상이 생겨 혈액 수용 및 펌핑을 잘하지 못해 몸이 붓거나 숨이 차는 증상인 심부전 등이 있다.

1.2.3 심박변이도(Heart Rate Variability, HRV):

심박변이도는 심장 박동의 변이 정도를 의미한다. 즉, 하나의 심장 주기

에서부터 다음 심장 주기 사이의 미세한 변이를 뜻한다. 심박수는 동방결절의 고유 자발성에 자율신경계가 영향을 미쳐 결정되며 부교감신경과 교감신경 사이의 상호작용과 상관이 있다.

심박변이도의 감소는 끊임없이 변화하는 환경에 대해 체내 적응 능력의 감소를 말한다. 심전도를 통해 얻어낸 심박의 변화(HRV)는 주파수, 시간 두 가지 관점으로 분석하게 된다. 그 중 주파수 범위 분석을 통해 자율신경계의 균형 정도를 알 수 있다. 각기 다른 대역의 주파수 신호를 주파수 영역에 따라 저주파, 고주파 신호를 분석함으로써 교감, 부교감이 어느 정도의 영향을 미치는지 객관적으로 파악할 수 있다.

심박변이도 검사를 통해 임상 적용에 도움이 되는 경우는 여러 가지가 있다. 자율신경계의 균형도와 활성도를 확인할 수 있으며, 스트레스 관련 질환의 발병 위험도의 예측 및 평가가 가능하다. 신경계로는 긴장성 두통, 편두통, 뇌졸중 등을 예측할 수 있고, 심혈관계 질환으로 심근경색증, 부정맥, 협심증, 서맥, 빈맥 등을 예측할 수 있다. 정신계로 우울증, 불안, 불면, 공황장애와 당뇨, 갑상선질환, 비만 등을 알 수 있다. 소화기계로는 기능성 위장장애, 과민대장증후군, 식욕 저하 등과 피로, 기력저하에 대해 평가가 가능하다. 또한, 질병에 대한 저항능력의 평가와 약물, 치료 효과에 대한 확인과 추적 검사가 가능하다.

1.2.4 심전도(Electrocardiogram, ECG):

심전도란 심장박동과 관련된 전위를 신체 표면에서 도형으로 기록한 것으로, 표준 12 유도 심전도 외에 운동부하 심전도, 활동 중 심전도 등이 있다. 심전도는 간단하고 정확하며, 재현성이 있고 비교적 쉽게 반복하여 기록할 수 있으며, 검사 비용이 저렴한 비침습적 검사이다.

심전도는 심장 활동에 의해 국소적으로 발생하는 전기변화를 기록한 그림이다. 전기변화는 용적 도체를 이루는 조직에 전해져 체표에 미치기 때문에 체표에서도 기록할 수 있다. 사람은 보통 체표의 특정 부위에 전극을

붙여 전위를 유도해 기록한다. 양손에서 유도된 제 1 도출, 오른손-왼발에서 유도된 제 2 도출, 왼손-왼발에서 유도된 제 3 도출이 주된 도출법이지만, 단극유도 등의 새로운 도출방법도 시도되고 있다.

심근의 흥분은 정맥동에서 일어나고 심방, 심실 방향으로 나아가기 때문에 이 흥분을 임의의 두 점에서 전류계에 유도하면 심장의 활동전류가 그래프로 그려지게 된다. 심전도는 심방 수축상황(P)과 심실의 수축상황(QRST)이 그래프로 교시되어 있다. 협심증이나 심근경색 등의 관동맥 질환을 비롯하여 여러 가지 부정맥이나 전해질이상 등의 진단, 또는 수술 중의 심장 이상 유무의 조사, 확인 등 그 응용 면이 넓고, 심장질환의 진단학상 매우 중요하다.

1.2.5 딥러닝(Deep Learning):

먼저 머신러닝은 데이터를 구문 분석하고 해당 데이터를 활용해 학습한 후 정보를 바탕으로 결정 내리기 위하여 학습한 내용을 적용하는 알고리즘이다. 주어진 데이터로 기능을 수행하며, 시간이 경과 될수록 그 기능이 점차 향상된다. 딥러닝은 머신러닝의 하위 개념이다. 딥러닝 모델은 인간이 결론을 내리는 방식과 유사한 논리 구조를 이용하여 데이터를 계속해서 분석하도록 설계되었다. 이를 달성하기 위해 인공신경망이라는 계층화된 알고리즘 구조를 사용한다.

딥러닝은 알고리즘을 계층으로 구성하여 자체적으로 학습하고 지능적인 결정을 할 수 있는 인공신경망을 만든다. 또한, 우수한 적응력을 가지고 데이터가 추가될 때마다 계속해서 진화하며 동적인 예측 시스템을 만들 수 있어 널리 쓰인다.

딥러닝에는 다양한 종류의 심층 신경망 구조가 존재하며, 대표적으로 순환 신경망(Recurrent Neural Network, RNN), 합성곱 신경망(Convolution Neural Network, CNN) 등이 있다.

1.2.6 합성곱 신경망(Convolution Neural Network, CNN):

합성곱 신경망은 시각적 영상을 분석하는 데 사용되는 다층의 인공신경망의 한 종류이다. 딥러닝에서는 심층 신경망으로 분류되고, 시각적 영상 분석에 주로 적용된다.

기존 Multi-Layered Neural Network(ML-NN)는 단순한 구조인 경우에도 많은 양의 파라미터들을 요구하게 되고, 데이터에 변형이 조금이라도 생기는 경우 새로운 학습 데이터를 입력하지 않으면 좋은 결과를 얻기가 힘든 구조이다. 따라서 이런 기존의 ML-NN의 문제점을 해결하기 위해 CNN 알고리즘이 출현하게 되었다.

CNN은 정규화된 다층 퍼셉트론이다. 다층 퍼셉트론은 일반적으로 완전히 연결된 네트워크, 즉 한 계층의 각 뉴런이 다음 계층의 모든 뉴런에 연결되는 신경망 구조이다.

CNN을 사용한 영상 분류는 타 영상 분류 알고리즘에 비해 상대적으로 거의 전처리를 사용하지 않는다. 이는 신경망이 기존 알고리즘에서 자체적으로 직접 제작된 필터를 학습한다는 것을 뜻한다. 설계자가 기존 영상 분류 알고리즘에서 영상의 특징들을 미리 이해해 알고리즘을 만드는 과정이 없는 것이 CNN의 주요한 장점이다.

CNN은 이미지를 하나하나씩 처리하는 것이 아니라, 픽셀 간 유기적 관계를 한꺼번에 인지하는 인간의 대뇌 시각피질 구조를 모방한 방법을 사용한다. 입력 데이터 특성을 추출하는 합성곱층과 인자를 축소하고 과잉적합을 제어함으로써 학습 성능을 개선시키는 풀링층을 교대로 연결한 형태로서 작동되는 구조이다.

작동원리는 먼저, 입력 영상으로부터 Convolution을 통해 Feature Map을 만든다. 그 후, Sub-Sampling을 통하여 Feature Map의 크기를 줄이고, 여러 단계의 Convolution과 Sub-Sampling 단계를 거치며 Feature의 크기가 작아지게 된다. 이렇게 작아진 Feature가 전체를 대표할 수 있는 강력한 특징들만 남게 되므로 분류가 가능하다.

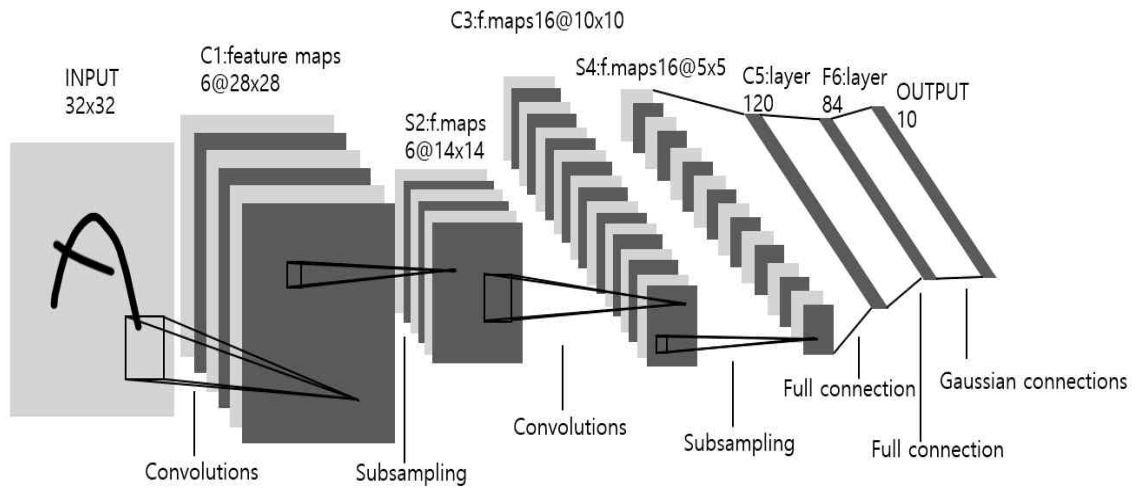


그림 2. CNN의 시초인 LeNet의 기본구조

1.3 선행연구:

1.3.1 EfficientNet-B0 :

본 연구에서는 표정 이미지와 생체신호 인식을 위해 딥러닝의 한 종류인 CNN을 사용하였다. 그 중, 본 연구에서는 CNN의 성능을 최소한의 파라미터로 최적의 성능을 내기 위해 연구한 모델인 EfficientNet을 사용하였다.

EfficientNet은 기존보다 훨씬 적은 수의 파라미터로 더욱 뛰어난 성능을 내서 Image Classification Task에 대해서 State-Of-The-Art(SOTA) 달성한 모델이다.

그림 3에서는 ImageNet 데이터 셋에서의 EfficientNet 모델 성능을 나타냈으며, 파라미터 수와 연산량에 따른 각 모델의 정확도를 나타낸 그래프이다. EfficientNet의 그래프를 보면 다른 Convolution Network와는 달리 적은 파라미터 수와 적은 연산량으로 훨씬 높은 정확도를 가지는 것을 확인할 수 있다. 즉, 모델의 구조가 작으며 계산 cost가 저렴하다는 장점이 있다.

모델의 정확도 높은 이미지 분류를 위해 기존 Convolution Network에서 Scaling up을 통한 성능개선은 계속 이어져 왔다. Scaling up 하는 방법은 세 가지가 있으며, 첫 번째로는 모델 신경망의 depth를 늘리는 것이다. Network가 깊어지게 되면 다른 task에 일반화하기 좋으며, 복잡한 특징을 추출할 수 있다. 두 번째로는 channel width를 늘리는 것이다. 더 넓은 channel은 더 세밀한 특징을 추출할 수 있고 train 하기가 쉽다. 세 번째로는 입력 이미지의 resolution을 올리는 것이다. 높은 해상도의 이미지를 input으로 사용할 때 모델은 더욱 세밀한 패턴을 학습할 수 있다.

하지만 기존 Scaling up 방법은 신경망의 depth를 늘리거나 channel width를 늘리거나, 입력 이미지의 resolution을 올리는 방식 중 각각 따로 1개씩만을 적용했다. EfficientNet은 이러한 기존 방식들에서 세 가지 방법에 대한 최적의 조합을 AutoML을 통해서 찾은 모델이며, 이 방법을

Compound Scaling이라고 한다. 이 세 가지는 일정한 관계가 있다는 것을 찾아내고, 이 관계를 간단한 상수비인 수식으로 만들었다. 즉, 고정된 비율로 세 가지 요소를 함께 조정하여 Scaling 하면 높은 정확도와 효율성을 가지는 것을 증명했으며, 그림 4를 통해 상수비를 확인할 수 있다.

EfficientNet의 기본 블록은 MBConv이며, EfficientNet-B0 모델을 기준으로 전체 Conv 블록은 18개이다. 양 끝의 두 블록은 Conv이고 나머지는 MBConv이다. 이처럼 본 논문에 사용한 EfficientNet-B0의 기본모델 구조를 그림 5에서 확인할 수 있다.

EfficientNet 모델은 크기에 따라 B0부터 B7의 순으로 나누어져 있으며, 본 논문에서는 이미지 데이터 인식을 위해 제일 가벼운 B0를 사용하여 표정 이미지와 생체신호 데이터를 학습시켰다. 효율성이 뛰어난 CNN 기반 데이터 분류 및 예측 모델을 통해 정확한 감정을 파악하는 도구로 활용하고자 한다.

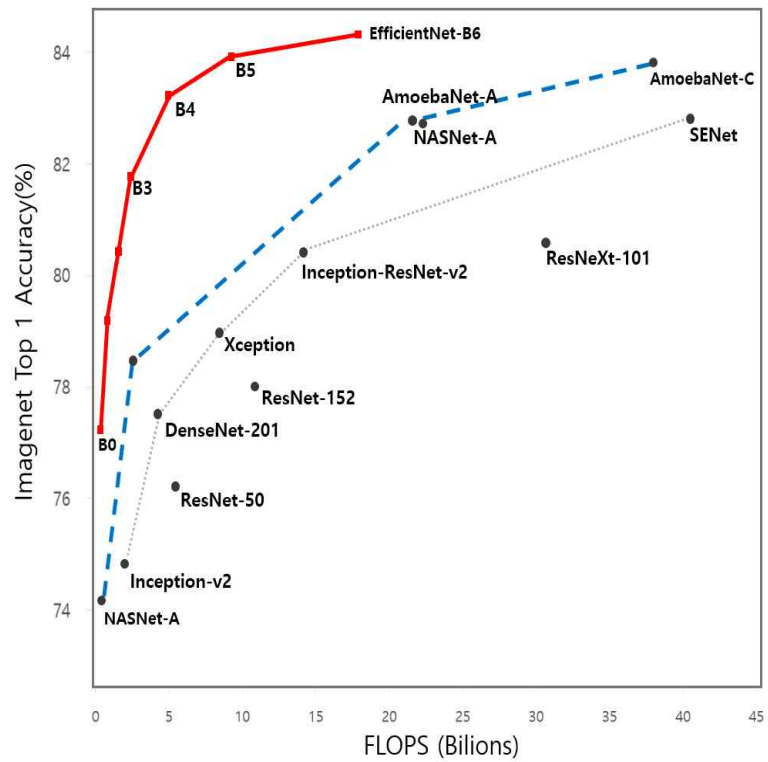
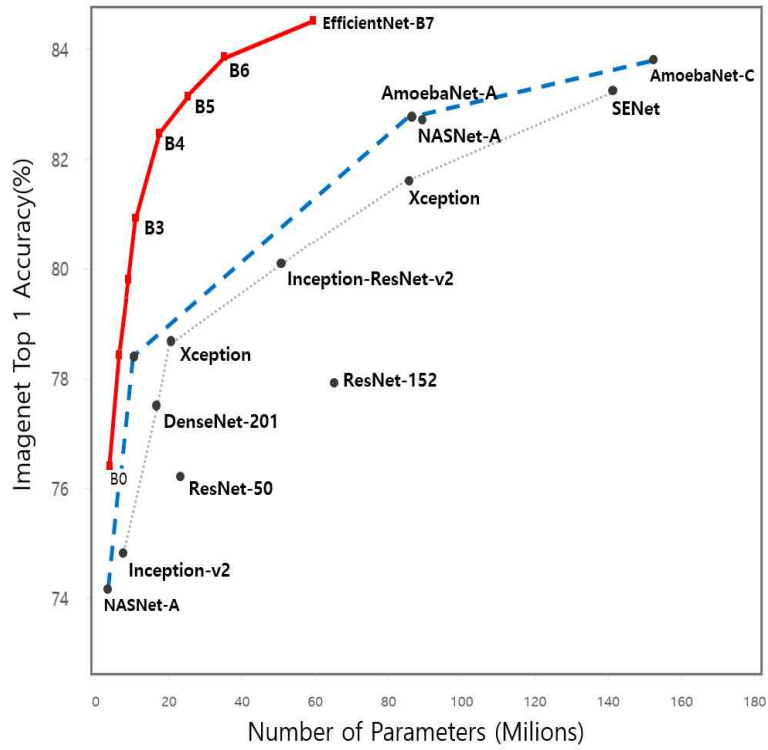


그림 3. ImageNet 데이터 셋에서의 EfficentNet 모델 성능

$$\begin{aligned} \text{depth} &: d = \alpha^\Phi \\ \text{width} &: W = \beta^\Phi \\ \text{resolution} &: R = \gamma^\Phi \\ \text{s.t. } &\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2 \\ &\alpha \geq 1, \beta \geq 1, \gamma \geq 1 \end{aligned}$$

그림 4. EfficientNet Compound Scaling 상수비

Stage i	Operator \hat{F}_i	Resolution $\hat{H}_i \times \hat{W}_i$	#Channels \hat{C}_i	#Layer \hat{L}_i
1	Conv3×3	224×224	32	1
2	MBCConv1,k3×3	112×112	16	1
3	MBCConv6,k3×3	112×112	24	2
4	MBCConv6,k5×5	56×56	40	2
5	MBCConv6,k3×3	28×28	80	3
6	MBCConv6,k5×5	14×14	112	3
7	MBCConv6,k5×5	14×14	192	4
8	MBCConv6,k3×3	7×7	320	1
9	Conv1×1 & Pooling & FC	7×7	1280	1

그림 5. 기본 EfficientNet-B0 모델 구조

1.4. 연구목적:

감정은 목소리, 행동, 얼굴 표정, 생체신호 등과 같은 다양한 방법으로 표현된다. 실제로 얼굴 표정을 통한 감정 인식 연구는 지속적으로 진행되었으나, 표정만으로 감정을 인식할 경우 모든 감정이 표정으로 나타나지 않기 때문에 측정이 어렵다. 또한, 표정을 숨긴다거나 거짓 표정으로 인해 정확한 감정을 파악하는 데 어려움이 있다(11,12).

그에 반해 생체신호는 자율신경계에 의해 통제되므로 거짓 없이 얻을 수 있다. 또한, 생체신호를 이용한 감정 인식은 주변 환경에 민감하지 않다는 장점을 가진다. 생체신호 처리 기술은 건강검진 분야, 각종 재활 분야 등의 의료 분야에도 응용될 수 있는 등 향후 다양한 분야에서 상용성 및 활용성에 대한 무한한 잠재력 또한 지니고 있으므로 해당 기술개발에 대한 중요성이 대두되고 있다(13,14). 따라서 생체신호를 이용한 감정 인식의 연구가 활발히 진행 중이며, 높은 감정 인식의 정확도를 가진다는 연구 결과가 존재한다.

여러 생체신호 중 심박변이도는 특히 감정 응답에 대한 객관적인 척도로 유용하게 사용되어 진다는 선행연구들이 존재한다. 다른 생리학적 변수와는 달리 심박변이도는 교감신경계와 부교감신경계 활동에 대한 정보를 제공하여 감정 조절의 억제 및 흥분 과정에 대한 추론이 가능하다. 이렇듯 사람의 감정은 자율신경계의 조절과 깊은 연관이 있다. 다른 요인에 의해 통제되는 것이 아닌 자율신경계의 조절에 의해 변화하는 심박변이도를 사용하여 감정을 파악하는 것이 정확도 높은 감정 파악이 가능하다(15).

따라서 본 연구의 목적은 사람의 표정으로부터 얻을 수 있는 감정 상태와 더불어 생체신호를 통해 알 수 있는 감정을 더 해 감정 인식의 정확도를 높이고 그 진실성을 파악하는 것이다. 감정의 Classification과 Prediction을 위해 적은 파라미터 수로 가볍고 높은 성능을 내는 모델인 EfficientNet을 사용해 표정 이미지와 심박변이도를 동시에 분석하여 감정 예측의 정확성을 높이고자 한다(16,17,18,19,20).

2. 재료 및 방법

2.1 재료:

2.1.1 실험환경:

본 연구는 RAM 32.0GB, Intel Core i7-8700, NVIDIA GeForce RTX 2070 환경에서 진행했다. PyTorch와 OpenCV 등의 라이브러리를 사용하여 Python으로 구현하였다.

2.1.2 표정 이미지 데이터:

이미지 데이터는 48×48 크기의 Gray Scale 해상도를 갖는 오픈 데이터인 FER2013 이미지를 사용했다. 본 실험에서는 총 7개 클래스 중 4개의 클래스인 angry, happy, neutral, sad를 사용하여 레이블을 변경하였다. 실제 실험에 사용한 표정 이미지 데이터의 예시를 그림 6과 같이 표현하였다.



그림 6. 학습에 사용된 표정 이미지 데이터

2.1.2. 생체신호 데이터:

실험에 사용된 생체신호 데이터는 감정에 따라 변화하는 심박변이도의 6가지 인덱스값을 표기한 기존 논문을 참고하여 각 클래스에 맞는 데이터를 발생시켰다. 기존 논문의 표 1 데이터처럼 감정 클래스에 따라 추출된 데이터가 확보되어 있으므로, FER2013 이미지 데이터의 표정 이미지는 표 1의 감정 클래스 데이터 범위 안에 포함된다고 가정하였다.

표 2에서는 실험에 사용된 심박변이도 데이터의 세부 내용을 확인할 수 있으며, 총 6가지의 Index에 대한 설명이 첨부되어 있다.

표 1. 학습에 사용한 심박변이도 데이터

Index		Emotions					
		Neutral	Fear	Sadness	Happiness	Anger	Disgust
SDNN	mean±SD	50±18	53±16	47±16	55±19	51±18	51±17
	reference order	2	5	1	6	3	4
RMSSD	mean±SD	41±19	43±18	40±18	43±18	44±21	44±18
	reference order	2	4	1	3	6	5
LF	mean±SD	0.54±0.17	0.55±0.15	0.54±0.16	0.66±0.15	0.53±0.18	0.54±0.17
	reference order	4	2	2	6	1	4
HF	mean±SD	0.46±0.17	0.45±0.15	0.46±0.16	0.34±0.15	0.47±0.18	0.46±0.17
	reference order	3	2	5	1	6	4
LF/HF	mean±SD	1.6±1.2	1.6±1.0	1.5±1.0	2.9±3.6	1.5±1.1	1.5±1.1
	reference order	5	4	1	6	2	3
SampEn	mean±SD	1.78±0.25	1.85±0.24	1.90±0.26	1.90±0.28	1.97±0.24	1.91±0.24
	reference order	1	2	4	3	5	6

표 2. 심박변이도 데이터 세부내용

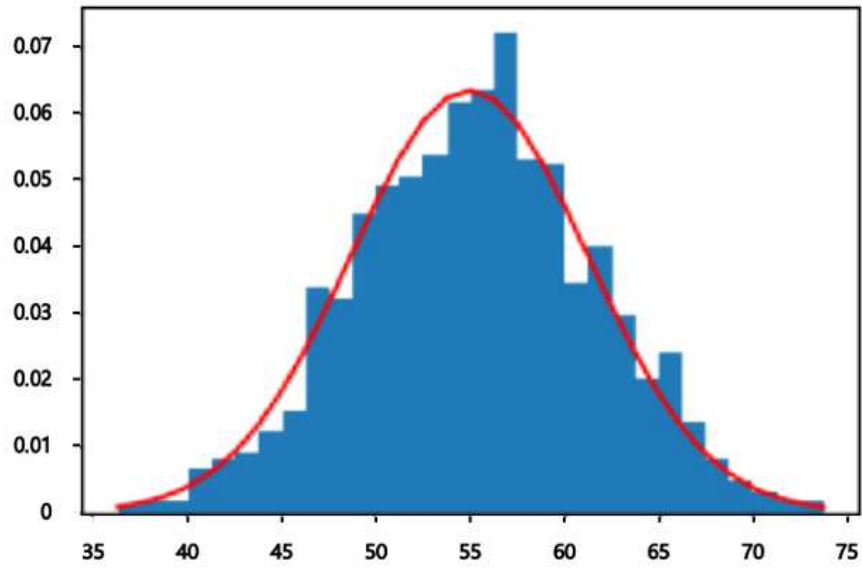
Index	
SDNN	RR 간격의 표준편차
RMSSD	RR 간격의 평균 제곱 차이
LF	주파수 영역에서의 정규화된 저주파 전력
HF	주파수 영역에서의 정규화된 고주파 전력
LF/HF	저주파 전력 / 고주파 전력
SampEn	심박수 변동성 시계열의 복잡성 분석

2.2 연구 방법:

2.2.1 데이터 전처리 과정:

표정 이미지 데이터의 경우, 오픈 데이터인 Kaggle의 데이터 셋 FER2013을 사용했다. 총 20,000장의 표정 이미지 중, 학습 데이터로 15,000장, 검증 데이터로 2,500장, 테스트 데이터로 2,500장을 사용했으며, shuffle을 True 값으로 주어 랜덤으로 이미지 데이터를 나눴다. 클래스로는 angry를 0, happy를 1, sad를 2, neutral을 3으로 지정했다.

생체신호 데이터의 경우, 기존 논문의 데이터인 감정에 따른 6개의 HRV (심박변이도) 인덱스가 정규분포를 따른다는 가정을 하고, 평균과 표준편차에 따라 정규분포를 통해 데이터를 발생시켰다. 한 장의 표정 이미지에 6개의 HRV 인덱스마다 1,000개의 샘플링을 수행하고 랜덤으로 값을 하나씩 추출하였다. 이렇게 이미지마다 총 6개의 HRV 인덱스값을 추출했다. 예를 들어, Happy 클래스의 SDNN 인덱스의 경우 평균값이 55이고 표준편차가 19이므로, 36~74의 숫자 범위에서의 정규분포가 생성되며, 정규분포 그래프는 그림 7과 같다. 총 1,000개의 샘플 값을 발생시켜 그 중 랜덤으로 하나의 값을 가져와 클래스와 맞는 표정 이미지와 매치시켰다. 같은 방법으로 하나의 이미지당 6가지 인덱스에 해당하는 샘플을 획득하였다. 실험에 사용한 데이터의 예시는 그림 8과 같으며, 클래스와 표정 이미지 데이터값, 생체신호의 6가지 Index 값이 숫자로 표현되어 있다.



Happy의 SDNN Index 정규분포 그래프

그림 7. 생체신호 데이터 샘플링 과정 예시

클래스 표정 이미지 데이터값		생체신호 데이터값						
label	imgs	series	SDNN	RMSSD	LF	HF	LF/HF	SampEn
0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121 119 115 110 98 91	(48.97022787838137,	36.36963951594076,	0.6389083305379127,	0.4403952909763408,	2.0256037109318363,	2.0415246422762445)	
0	151 150 147 155 148 133 111 140 170 174 182 154 153 164 173 178	(60.63363742835894,	44.00047763899982,	0.35806394906621086,	0.4411471539393018,	1.3823961472953346,	2.168548632866893)	
2	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 19 43 52 13 26 40 59	(46.016420543540356,	36.95893468152582,	0.7704403069604165,	0.3858336044421048,	1.4754987031112978,	2.034209275580203)	
3	4 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84 115 127 137 142 151	(50.6219647112044,	40.189283107970624,	0.5259779586703368,	0.435345787377264,	1.0716848101600434,	1.8224188577973108)	
2	20 17 19 21 25 38 42 42 46 54 56 62 63 66 82 108 118 130 139 134	(48.06444528155421,	51.016269650217644,	0.6629062078872417,	0.44118850207072596,	1.481530651801824,	1.7511972760912495)	
1	77 78 79 79 78 75 60 55 47 48 58 73 77 79 57 50 37 44 56 70 80 82	(58.866848364971915,	39.088982903431294,	0.7400529923379657,	0.3397194435381675,	6.031096833782262,	2.0030068565310977)	
1	85 84 90 121 101 102 133 153 153 169 177 189 195 199 205 207 209	(46.699291360131745,	42.346898467220754,	0.6841000622615773,	0.3602375583939133,	3.6862323769740644,	1.86605264448246)	
0	30 24 21 23 25 25 49 67 84 103 120 125 130 139 140 139 148 171 11	(46.20970930973775,	49.98232893911779,	0.5390628349356142,	0.39979225697896403,	0.8614606286476859,	1.9021716618911977)	
3	39 75 78 58 58 45 49 48 103 156 81 45 41 38 49 56 60 49 32 31 28	(55.308314370723835,	45.44279543478799,	0.523703110560022,	0.40856036311536764,	1.4862272333840223,	1.675977944844429)	

그림 8. 학습에 사용된 데이터

2.2.2 Image Classification 알고리즘 EfficientNet 모델 설계:

딥러닝 CNN 기반인 EfficientNet 알고리즘은 Image Classification Task에 대해서 기존보다 훨씬 적은 파라미터 수로 더욱 좋은 성능을 내서 State-Of-The-Art(SOTA)를 달성한 모델이다. 간단하면서도 높은 효율을 보이는 새로운 Scaling 방법을 제안함으로써 빠른 inference와 높은 정확도를 유지한다.

본 논문에서는 이러한 특징을 가진 EfficientNet-B0를 사용하여 확보한 데이터를 분류해 정확한 감정 상태를 예측하는 모델을 구성하였다. 본 논문에서 제시한 표정 이미지와 생체신호인 HRV Index를 기반으로 사람의 감정을 인식하기 위한 전체 시스템 흐름도는 그림 9와 같다.

모델은 두 가지의 실험을 위해 설계되었는데 실험 1은 FER2013 이미지 데이터만을 이용하여 EfficientNet-B0 모델을 학습시키고 테스트를 진행했다. EfficientNet-B0는 $1280 \times 7 \times 7$ feature를 출력하므로 Dense layer를 이용하여 256 feature로 인코딩한다. 최종적으로는 1개의 감정 예측값을 출력하게 된다. 실험 2에서는 생성된 HRV 인덱스를 개별 Dense layer를 거쳐 64 feature로 인코딩한다. 최종적으로 FER2013 이미지 데이터 256 feature와 concat 되어 감정을 예측하는 데 사용되었다.

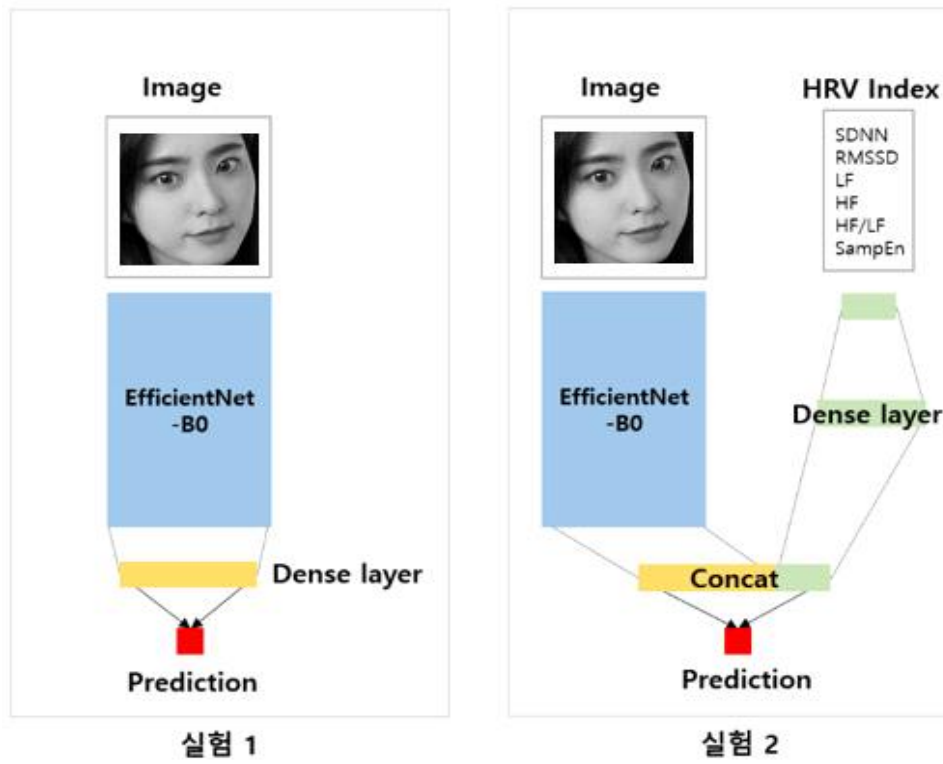


그림 9. 표정 이미지 및 생체신호를 이용하여 감정 분석하기 위한 모델 비교

2.2.3. 데이터 학습 및 모델 검증:

Training 과정에서 데이터는 총 15,000장의 표정 이미지 데이터와 심박변이도 데이터를 사용했고, 검증 데이터로는 2,500장, 테스트 데이터로는 2,500장을 사용했다. 학습 과정은 매 Epoch마다 검증 데이터를 사용하여 모델의 성능을 평가하고 가장 높은 정확도를 달성할 때마다 모델의 하이퍼 파라미터 값을 저장하였다.

데이터의 input size는 224이며, 이는 원본 이미지 48×48 크기의 이미지를 모델의 입력으로 넣기 위해 224×224 로 resize 시켰다. 학습 파라미터의 Epoch는 40, Batch size는 16으로 지정했으며, Learning rate는 0.0001로 설정하였다.

그림 10은 실험 1의 학습 및 검증 데이터 셋에서 훈련의 성능을 체크한 학습 곡선이다. 학습 곡선을 체크하는 이유는 현재 모델이 underfit 되고 있는지, overfit 되고 있는지의 문제를 파악할 수 있기 때문이다. 이러한 문제점을 체크함으로써 데이터를 추가해야 하는지, 모델의 크기를 늘려야 하는지 등 모델의 성능을 높이기 위한 방안을 모색할 수 있다. Loss(손실)는 실제 정답과 모델이 예측한 값 사이의 차이이며, 손실이 클수록 데이터의 추론에 대한 오류가 커지므로 Loss 값은 작을수록 훈련이 잘 진행되고 있다는 뜻이다. 정확도는 말 그대로 모델이 예측한 값과 정답이 얼마나 정확한지에 대한 값으로 학습 곡선에서 Loss 값은 계속 떨어지고 Accuracy는 높아지는 것이 훈련이 잘되고 있음을 나타낸다.

학습 곡선에서는 Epoch에 따른 Loss와 Accuracy 값을 확인할 수 있다. 그림 10의 Loss 그래프를 보면 validation에서 학습이 진행될수록 값이 증가하고 Accuracy 그래프에서는 정확도가 더는 증가하지 않고 일정하게 머무르는 것을 알 수 있다. 이를 통해 학습이 진행되면서 오버피팅이 발생했다는 것을 의미하며, 표정 이미지만으로 감정을 예측했을 때의 정확도는 높지 않다는 것을 알 수 있다.

그림 11은 실험 2의 학습 및 검증 그래프이며, 마찬가지로 Epoch에 따른

Loss와 Accuracy 값 확인이 가능하다. 실험 1에서 생체신호 데이터만 추가했을 뿐인데 Loss 그래프의 Validation에서 Loss 값이 점점 줄어드는 것을 볼 수 있으며, Accuracy 값도 점차 증가해 실험 1보다 정확도가 높다는 것을 알 수 있다.

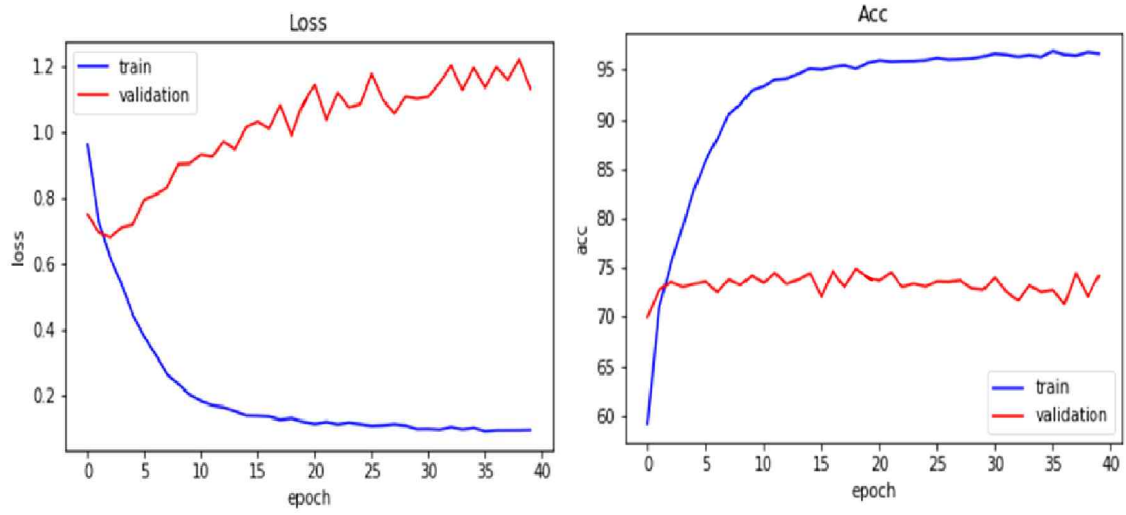


그림 10. 실험 1의 학습 곡선

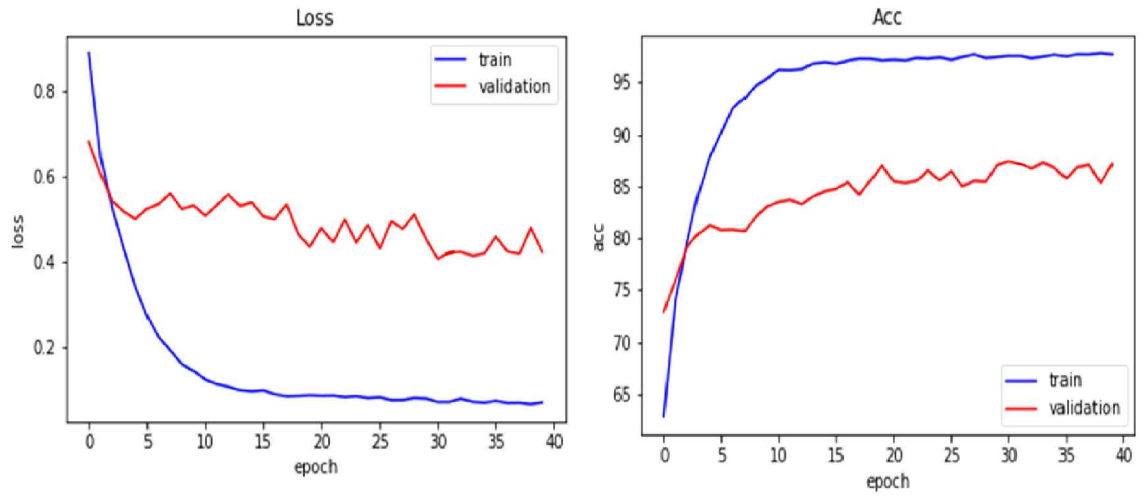


그림 11. 실험 2의 학습 곡선

3. 성 적

3.1. 연구결과:

3.1.1. 성능 지표:

본 연구에서는 훈련된 모델을 평가하기 위해 이미지 Classification and Prediction 모델의 평가 방법인 Confusion matrix의 Accuracy를 사용하였다. Confusion matrix는 의도에 맞게 주어진 데이터를 잘 분류해내기 위해 모델이 얼마나 정밀한지, 얼마나 실용적인지, 얼마나 정확한지 평가하는 방법이다.

관심 범주를 정확하게 분류한 값을 True Positives(TP)라고 하며, 관심 범주가 아닌 것으로 잘못 분류한 것을 False Negatives(FN), 관심 범주라고 잘못 분류한 것을 False Positives(FP), 관심 범주가 아닌 것을 정확하게 분류한 것을 True Negatives(TN)라고 하는데 이것을 조합해 정확도, 정밀도, 재현도, 정밀도와 재현도를 함께 평가하는 F1-Score의 척도를 사용해 모델을 평가한다. 그중 본 연구에서 사용한 Accuracy는 전체 샘플 중 올바르게 예측한 샘플 수의 비율을 뜻하며, 계산식은 표 3과 같다.

표 3. Confusion matrix 척도

민감도 (Sensitivity/ Recall)	$(TP / TP + FN)$ 실제 Positive인 것 중 Positive로 예측한 것의 비율
정밀도 (Precision)	$(TP / TP + FP)$ Positive로 예측한 것 중 실제 Positive인 비율
재현율 (Recall)	$(TP / TP + FN)$ 실제 Positive인 것 중 Positive로 예측한 것의 비율
정확도 (Accuracy)	$(TP + TN / TP + FN + FP + TN)$ 전체 데이터 중에 맞게 예측한 것의 비율
F1-Score	$(2 * Recall * Precision) / (Recall + Precision)$ 정밀도와 재현율의 조화평균 점수

3.1.2. 모델 성능 평가:

모델 평가에 사용된 데이터는 총 2,500개의 표정 이미지와 심박변이도 생체신호 데이터를 통해 감정 상태를 올바르게 분류하도록 하였으며 테스트 이미지에 적용한 결과를 그림 12와 같이 출력하여 분류 상태를 직접 볼 수 있었다.

Accuracy는 실험 1에서 74%, 실험 2에서는 88.2%로 표정 이미지와 생체신호 데이터를 함께 분석했을 때의 정확도가 훨씬 더 높다는 것을 그림 13에서 알 수 있다. 현재로서는 Accuracy만 가지고 모델의 정확도를 판단했지만, 추후 다양한 모델 평가 지표를 통해 모델의 분류 성능을 판단할 수 있을 것이다.

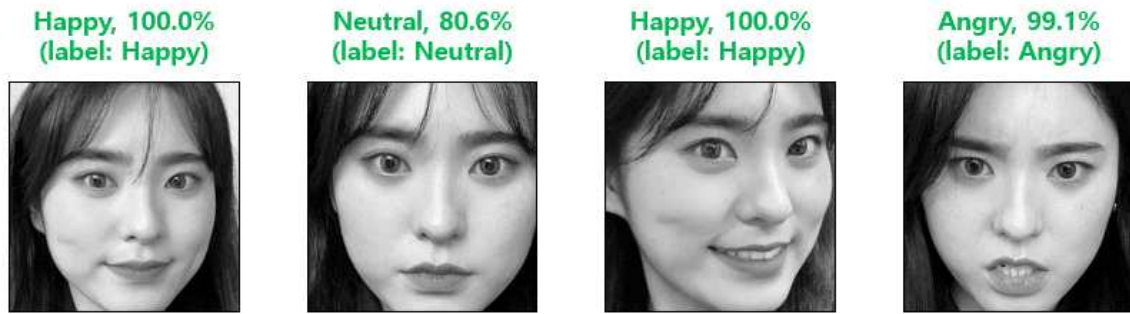


그림 12. 모델이 분류한 감정 예측 결과

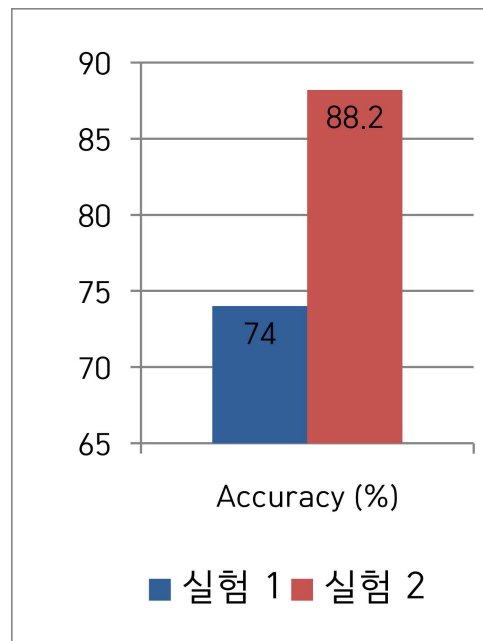


그림 13. 실험에 따른 모델 정확도 비교 그래프

4. 고찰

본 논문에서는 표정 이미지와 생체신호를 결합한 데이터를 기반으로 딥러닝 모델 EfficientNet을 설계하여 인간의 감정 예측 정확도를 높일 수 있는 방안을 제시하였다. 하지만 실제 인간의 표정 이미지나 생체신호를 사용하기 위해서는 IRB 인증 과정이 필요하며, 그 과정이 까다롭고 시간이 상당히 소요되기 때문에 본 연구에서는 오픈 데이터를 이용해 인위적으로 발생시켜 사용했다. 생체신호 데이터를 발생시켜 표정 이미지의 클래스 안에 포함된다고 가정한 실험이기 때문에 다른 데이터를 사용했을 때는 결과값이 달라질 가능성도 존재한다. 따라서 추후의 연구에서는 피실험자를 모집하여 실제 데이터를 이용해 확장된 연구를 진행하려 한다. 높은 데이터 차원에서 연구를 진행하기 위해서는 CNN 모델의 성능을 높이는 다양한 테크닉들을 활용하여 정확성과 확장성을 키워야 할 것이다.

본 논문에서는 총 20,000장의 이미지를 사용했는데, 감정 예측의 정확도 향상을 위해 딥러닝 알고리즘의 성능을 높여 추가적인 데이터를 확보하여 모델을 학습시키려 한다. 또, 현재 Angry, Happy, Sad, Neutral 4가지 클래스를 사용하여 감정을 예측했지만, 이 외에도 Fear, Surprise 등 다양한 감정들이 존재하므로 여러 클래스를 사용하여 연구의 확장성을 키울 수 있을 것이다.

본 논문의 궁극적 목표는 실제 사람의 표정 데이터와 생체신호 데이터를 카메라 영상을 통해 실시간으로 측정해 감정을 예측하는 것이다. 이러한 목표의 가능성을 확보하기 위해 현재 진행한 연구는 두 가지의 데이터를 같이 사용하는 것이 감정 예측 정확도가 더 높다는 것을 사전연구로써 증명한 기초연구단계라고 할 수 있다. 따라서 두 데이터를 함께 사용하는 것이 감정 예측에 더 정확하다는 것을 파악했기 때문에 추후 확장된 데이터 차원인 영상 데이터에 활용하여 실시간으로 감정을 예측할 수 있을 것이다. 또한, 가볍고 높은 성능의 CNN 모델을 사용함으로써 더 높은 차원의

데이터를 사용할 수 있으므로 IoT를 결합한 인식 시스템 개발에도 유용할 것으로 보인다. 또한, 감정 인식 연구는 스마트 홈이나 자율주행 시스템, 의료 분야 등 다양한 분야에서 사용될 수 있으므로, 연구의 필요성이 더욱 많을 것으로 보여진다.

5. 요약

감정 인식을 위해 가장 널리 사용되는 지표는 얼굴 표정이지만, 통제와 조작이 가능하다. 이러한 문제점 때문에 최근 개인의 통제를 벗어난 생체 신호를 이용한 연구가 활발히 진행 중이다. 하지만 현재까지 생체신호와 표정을 다차원 분석한 경우는 거의 없다.

본 논문은 이러한 문제점을 해결하기 위해 생체신호와 얼굴 표정 이미지를 함께 분석해 대상자의 감정 상태 정확도를 높일 수 있는 CNN 기반의 뉴럴 네트워크를 설계한 시스템을 제안했다. 최소한의 파라미터로 최대한의 성능을 낼 수 있는 EfficientNet-B0 아키텍처를 사용했으며, 이는 Single board computer에도 적용이 가능할 정도로 가볍고 성능이 우수하다.

모델을 학습시키기 위해 표정 이미지 20,000장을 확보하였으며, 이중 15,000장은 Train data, 2,500장은 Validation data, 2,500장은 Test data로 사용했다. 또한, 데이터 전처리 과정으로 생체신호 데이터를 각 클래스에 맞게 정규분포를 통해 데이터를 발생시켰으며, 1,000개의 샘플 데이터 중 랜덤으로 하나의 값을 가져와 해당하는 표정 이미지와 매치시켰다.

학습된 결과로 테스트 데이터 셋에서 생체신호를 추가한 실험 2의 정확도가 88.2%로 표정 이미지만을 가지고 학습한 모델의 정확도인 74%보다 높은 수준의 결과를 얻을 수 있었다. 따라서 두 가지의 데이터를 함께 사용하는 것이 감정 예측의 정확도를 높일 수 있다는 것을 파악했고, 본 연구에서 사용한 최적의 파라미터를 가지는 알고리즘이 높은 성능을 가지는 것을 검증하였다.

본 연구에 사용한 EfficientNet 아키텍처는 최소한의 파라미터로 높은 정확도와 성능을 낼 수 있는 가벼운 모델이므로, 추후 영상 데이터에도 적용하여 실시간으로 영상과 생체신호를 측정할 감정 예측의 가능성도 보여진다.

또한, 감정 인식 연구는 거짓말 탐지거나 의료환경 분야에 이용되어 연구의 필요성과 가능성이 더욱 클 것으로 예상된다. 모델이 가볍기 때문에 단

일보드 컴퓨터에서도 구동이 가능하며, IoT를 결합한 감정 인식 시스템 개발에도 유용할 것으로 예상된다.

참 고 문 헌

1. Cha, Won-Young, Shin, Dong-Kyoo, Shin, Dong-Il. (2017). Analysis and Comparison of The Emotion Recognition by Multiple Bio-Signal. Proceedings of the Korean Information Science Society Conference, 1598-5164(pISSN), 104-105
2. C. Chang, J. Tsai, C. Wang and P. Chung, "Emotion recognition with consideration of facial expression and physiological signals," 2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology, 2009, pp. 278-283
2. J. Kortelainen et al., "Multimodal emotion recognition by combining physiological signals and facial expressions: A preliminary study," 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, San Diego, CA, 2012, pp. 5238-5241
3. C. Godin, F. Prost-Boucle, A. Campagne, S. Charbonnier, S. Bonnet and A.Vidal. "Selection of the Most Relevant Physiological Features for Classifying Emotion" In Proceedings of the 2nd International Conference on Physiological Computing Systems (PhyCS-2015), pages 17-25. ISBN: 978-989-758-085-7
4. Dzedzickis, Andrius, Artūras Kaklauskas, and Vytautas Bucinskas. 2020. "Human Emotion Recognition: Review of Sensors and Methods" Sensors 20, no. 3: 592.

5. Daniel S. Quintana, Adam J. Guastella, Tim Outhred, Ian B. Hickie, Andrew H. Kemp, Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition, *International Journal of Psychophysiology*, Volume 86, Issue 2, 2012, Pages 168-172, ISSN 0167-8760
6. C. Jing, G. Liu and M. Hao, "The Research on Emotion Recognition from ECG Signal," 2009 International Conference on Information Technology and Computer Science, 2009, pp. 497-500
7. G. Du, S. Long and H. Yuan, "Non-Contact Emotion Recognition Combining Heart Rate and Facial Expression for Interactive Gaming Environments," in *IEEE Access*, vol. 8, pp. 11896-11906
8. A. Kumar, J. Kim, D. Lyndon, M. Fulham and D. Feng, "An Ensemble of Fine-Tuned Convolutional Neural Networks for Medical Image Classification," in *IEEE Journal of Biomedical and Health Informatics*, vol. 21, no. 1, pp. 31-40, Jan. 2017
9. T. Guo, J. Dong, H. Li and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), 2017, pp. 721-724
10. K. K. Pal and K. S. Sudeep, "Preprocessing for image classification by convolutional neural networks," 2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), 2016, pp. 1778-1781

11. Spiros V. Ioannou, Amaryllis T. Raouzaïou, Vasilis A. Tzouvaras, Theofilos P. Mailis, Kostas C. Karpouzis, Stefanos D. Kollias, Emotion recognition through facial expression analysis based on a neurofuzzy network, *Neural Networks*, Volume 18, Issue 4, 2005, Pages 423-435, ISSN 0893-6080
12. Y. -D. Zhang et al., "Facial Emotion Recognition Based on Biorthogonal Wavelet Entropy, Fuzzy Support Vector Machine, and Stratified Cross Validation," in *IEEE Access*, vol. 4, pp. 8375-8385, 2016
13. S. Wioleta, "Using physiological signals for emotion recognition," 2013 6th International Conference on Human System Interactions (HSI), 2013, pp. 556-561
14. J. Shin, J. Maeng and D. Kim, "Inner Emotion Recognition Using Multi Bio-Signals," 2018 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia), 2018, pp. 206-212
15. A. S. Ravindran, S. Nakagome, D. S. Wickramasuriya, J. L. Contreras-Vidal and R. T. Faghih, "Emotion Recognition by Point Process Characterization of Heartbeat Dynamics," 2019 IEEE Healthcare Innovations and Point of Care Technologies, (HI-POCT), 2019, pp. 13-16
16. Mingxing Tan, Quoc V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks, *Machine Learning (cs.LG); Computer Vision and Pattern Recognition (cs.CV); Machine Learning*

(stat.ML), International Conference on Machine Learning, 2019

17. Diah Anggraeni Pitaloka, Ajeng Wulandari, T. Basaruddin, Dewi Yanti Liliana, Enhancing CNN with Preprocessing Stage in Automatic Emotion Recognition, *Procedia Computer Science*, Volume 116, 2017, Pages 523-529, ISSN 1877-0509

18. X. Li, D. Song, P. Zhang, G. Yu, Y. Hou and B. Hu, "Emotion recognition from multi-channel EEG data through Convolutional Recurrent Neural Network," 2016 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2016, pp. 352-359

19. H. Zhang, A. Jolfaei and M. Alazab, "A Face Emotion Recognition Method Using Convolutional Neural Network and Image Edge Computing," in *IEEE Access*, vol. 7, pp. 159081-159089

20. J. X. Chen, P. W. Zhang, Z. J. Mao, Y. F. Huang, D. M. Jiang and Y. N. Zhang, "Accurate EEG-Based Emotion Recognition on Combined Features Using Deep Convolutional Neural Networks," in *IEEE Access*, vol. 7, pp. 44317-44328, 2019

EfficientNet Architecture-Based Systems and Algorithms for Predicting Emotions in Humans

Kim, Min Young

Department of Bio Medical Engineering

School of Medicine

Keimyung University

(Supervised by Professor Lee, Jong-Ha)

(Abstract)

Recently, with the development of artificial intelligence technology, the field of emotion recognition has emerged as an important research field. Emotion refers to the feeling or feeling that arises in response to a certain phenomenon or event. The most widely used indicator for emotion recognition is facial expressions. However, since facial expressions can be controlled and manipulated, it is difficult to grasp the emotions that are truly felt only by facial expressions.

In order to solve this problem, recently, research on emotion recognition using independent biosignals that can be continuously measured and that is out of personal control is being actively conducted. There are many preceding studies that there is a strong correlation between human emotions and biosignals, and emotion recognition through biosignals is controlled by the autonomic nervous system, so it can be obtained without falsehood. However, so far, few studies have used biosignals and facial expression templates for multidimensional analysis to recognize the subject's emotions.

Therefore, in this paper, an algorithm that can predict the accuracy of the subject's emotional state by designing a super dimension-based convolution neural network and analyzing bio-signals and facial expression templates together was proposed. Also, in this study, a network was designed and verified using the Convolution Neural Network based EfficientNet-B0 architecture, which can produce maximum performance with minimum parameters.

Through the experimental results, it was verified that it has high accuracy when predicting the emotional state by combining bio-signals rather than predicting emotions only with the subject's facial expression image, and that it has high performance as an optimal parameter. This study is seen as a starting point for emotional intelligence research that can be combined with artificial intelligence in the future, and it is expected to maintain high performance even in the extended data dimension in the future.

인간의 복합감정을 예측할 수 있는 EfficientNet 아키텍처 기반 시스템 및 알고리즘 연구

김 민 영

계 명 대 학 교 대 학 원
의학과 의용공학 전공

(지도교수 이 중 하)

(초록)

최근, 인공지능의 기술 발달과 함께 감정 인식 분야가 중요한 연구 분야로 대두되고 있다. 감정이란 어떤 현상이나 일에 대하여 일어나는 마음이나 느끼는 기분을 뜻한다. 감정 인식을 위해 가장 널리 사용되는 지표는 얼굴 표정이다. 하지만 표정은 통제할 수 있으며, 조작이 가능하므로 진정으로 느껴지는 정서를 표정만으로 파악하기는 어렵다. 이러한 문제점을 해결하기 위해 최근 들어 지속적인 측정이 가능하고 개인의 통제를 벗어난 독립적인 바이오 신호를 이용한 감정 인식 연구가 활발히 진행 중이다. 사람의 감정과 바이오 신호는 강한 상관관계가 있다는 많은 선행연구가 존재하며, 바이오 신호를 통한 감정 인식은 자율신경계에 의해 통제되므로 거짓 없이 얻을 수 있다. 하지만 현재까지 바이오 신호와 얼굴 표정 템플릿을 다차원 분석하여 대상자의 감정 인식에 사용한 연구는 거의 없었다.

따라서 본 논문에서는 슈퍼디멘션 기반 컨볼루션 뉴럴 네트워크를 설계하여 바이오 신호와 얼굴 표정 템플릿을 함께 분석하여 대상자의 감정 상태 정확도를 예측할 수 있는 알고리즘을 제안하였다. 또한, 본 연구에서는 최소한의 파라미터로 최대한의 성능을 낼 수 있는 Convolution Neural Network based EfficientNet-B0 아키텍처를 사용하여 네트워크를 설계하였고 이를 검증하였다.

실험 결과를 통해 대상자의 얼굴 표정 이미지만으로 감정을 예측하는 것보다 바이오 신호를 결합하여 감정 상태를 예측했을 때 높은 정확도를 가지며, 최적의 파라미터로 높은 성능을 가지는 것을 검증하였다. 본 연구는 앞으로 인공지능과 결

합될 수 있는 감정지능 연구의 출발점으로 보여지며, 향후 확장된 데이터 차원에서 높은 성능을 유지할 수 있을 것으로 예상되어 진다.