

데이터셋 프레임 길이에 따른 집중도 예측 모델 성능 비교*

Comparison of Concentration Prediction Model Performance Based on Dataset Frame Length

김동욱[†] · 김수균^{††} · 윤여찬^{†††}

Donguk Kim[†] · Sookyun Kim^{††} · Yeochan Yoon^{†††}

요약

코로나19 이후 온라인 심리상담 수요 증가로, 비대면 환경에서 내담자의 감정 이해와 집중 유지가 중요해졌다. 이전 연구는 눈 깜빡임, 하품, 멍 때림 등을 기반으로 한 300프레임 길이의 DAiSEE(Dataset for Affective States in E-Environments) 데이터셋으로 집중도 분석을 했으나, 긴 프레임의 행동을 포착하는 데 한계가 있었다. 본 연구에서는 새로운 2000프레임의 데이터셋을 구축하여 머신러닝으로 집중도를 분석하고, 기존의 300프레임의 DAiSEE 데이터셋과 성능을 비교한다. 비교 결과 프레임 길이가 긴 데이터셋으로 학습할 때의 성능이 더 우수한 것을 확인한다. 또한, 상담자가 내담자의 상태를 직관적으로 이해할 수 있는 사용자 인터페이스를 개발해 상담 효과를 극대화할 것으로 기대된다.

주제어: 온라인 상담, 비디오 분석, 머신러닝, 얼굴 분석, 데이터셋

ABSTRACT

After COVID-19 pandemic, the demand for online psychological counseling has increased, making it important to understand the emotions and maintain the concentration of clients in non-face-to-face environments. Previous research used the DAiSEE (Dataset for Affective States in E-Environments) dataset, which is based on 300 frames capturing behaviors like blinking, yawning, and spacing out, to analyze concentration. However, this dataset had limitations in capturing longer frame behaviors. In this study, we constructed a new dataset of 2000 frames and analyzed concentration using machine learning, comparing its performance with the existing 300-frame DAiSEE dataset. The comparison results show that training with the longer frame dataset yields better performance. Additionally, the development of a user interface that intuitively helps counselors understand the client's state is expected to maximize the effectiveness of counseling.

Keywords: Online counseling, Video Analysis, Machine Learning, Face Analysis, Dataset

1. 서론

코로나19 팬데믹 이후 사회의 많은 부분이 급격한 디지털 전환을 경험했다. 특히, 심리상담 분야에서는 대면 상담의 제한 때문에 온라인 상담 서비스의 수요가 많이 증가했다.[1] 이러한 상황에서 상담의 질과 효과성을 결정하는 핵심 요소 중 하나인 내담자의 감

정 이해와 집중 유지는 더욱 중요해졌다.

내담자의 표정, 시선, 자세 변화 등의 비언어적 신호들은 상담 중에 발생하는 감정적 변화와 집중도를 반영하는 중요한 신호이다.[2] 이러한 신호들을 분석함으로써, 상담자는 내담자의 상태를 더 정확하게 이해하고, 적절한 상담 기법을 적용할 수 있게 된다. 이 과정에서 복잡한 데이터 패턴을 인식하고 다양한 신

[†]정 회 원: 제주대학교 소프트웨어학부 컴퓨터공학과

^{††}정 회 원: 제주대학교 소프트웨어학부 컴퓨터공학과 교수

^{†††}정 회 원: 제주대학교 소프트웨어학부 인공지능전공 교수(교신저자)

논문투고: 2023년 11월 29일, 심사완료: 2024년 01월 08일, 게재확정: 2024년 01월 10일

* 이 논문은 2023년도 제주대학교 교원성과지원사업에 의하여 연구되었음

호 간의 상호작용을 고려하여 더 정확한 예측을 가능하게 하는 머신러닝 모델의 적용은 상담 효과를 높이는 데 중요한 역할을 할 수 있다.

본 연구에서는 온라인 상담 환경에서 내담자의 집중도를 예측하기 위해 여러 가지 특징을 분석하였고 머신러닝 모델을 통해 집중도 예측하는 방법론을 탐구하였다. 이 과정에서 데이터셋의 프레임 수가 모델 성능에 미치는 영향을 살펴보고 효과적인 데이터셋을 선택한다. 이를 통해 높은 정확도로 내담자의 비언어적 신호를 분석할 수 있었다. 또한 이러한 분석 결과를 상담자가 직관적으로 이해할 수 있도록 사용자 인터페이스를 개발하였다.

본 연구는 다음과 같은 구조로 진행된다. 먼저, 2장에서는 딥러닝과 머신러닝이 집중도 분석이나 얼굴 특징 분석 분야에 어떻게 적용될 수 있는지에 대한 기존 연구를 검토하고, 본 논문에서 선택한 방법을 소개한다. 또한 기존의 사용하던 데이터셋을 소개하고 이러한 데이터셋의 한계점을 제시하고 본 연구에서 사용할 데이터셋의 구축 방향에 대해 소개한다. 이어서, 3장에서는 얼굴 이미지에서 추출하는 특성들을 소개하고, 각 특성을 어떻게 추출하는지 설명한다. 4장에서는 기존 연구에서 사용하던 데이터셋과 새로 구축한 데이터셋의 전처리 과정과 특징들의 상관관계에 대해 설명한다. 5장에서는 교차 검증을 통해 효율적인 데이터셋을 선택하고 이를 학습한 모델을 이용한 사용자 인터페이스를 소개한다. 마지막으로 6장에서는 연구 결과를 요약하고 향후 연구 방향을 제안한다.

이 연구를 통해, 상담자는 더욱 미세하고 정교한 내담자의 신호들을 파악할 수 있게 되며, 이는 상담 과정의 질을 향상하고, 내담자의 치유 과정에 더욱 심층적으로 참여할 수 있는 기회를 제공할 것으로 기대된다.

2. 관련 연구

2.1 딥러닝 기반 측정 모델

심리상담 영역에서 비언어적 신호를 해석하는 데 있어 딥러닝(deep learning) 기술은 주목할 만한 발전을 이루어왔다. 특히, 컴퓨터 비전과 자연어 처리 분야에서의 성공은 심리상담의 비언어적 신호 분석에도 영향을 미쳤다. CNN, RNN, LSTM과 같은 모델들은 표정 인식, 음성 분석, 시선 추적 등에서 탁월한 성능을

보여주며 상담 상황에서의 내담자의 감정과 집중도를 평가하는 데 사용되었다. Chirra [3]의 연구는 운전자의 얼굴 이미지에서 눈 부분을 추출해 CNN(Convolutional Neural Networks)을 활용하여 졸고 있는지를 분석하였다. Magan [4]의 연구에서는 얼굴 이미지에 대해서 CNN을 사용하고 모호한 대상을 다루는 퍼지 논리를 적용해 사용자의 졸음운전을 판별하고 알려주는 시스템을 구현하였다. Su [5]의 연구에서는 교실 내 학생의 집중도를 관찰하기 위해 학생의 얼굴 특징들과 자세 특징들을 추출하여 CNN을 통해 집중도 분석을 하였고 이를 통해 교사에게 학생의 학습 상태에 대한 통찰력을 제공하는 데 목적이 있다. Solnaki [6]은 DAISEE 데이터셋을 이용하여 집중도 분석을 하여 높은 성능을 기록하였다.

그러나 이러한 모델들의 적용에는 몇 가지 문제점이 존재한다. 데이터의 양과 질에 대한 요구가 높을 뿐만 아니라, 학습과 예측 과정에서 필요한 계산 자원도 상당하다. 또한, 딥러닝 모델은 '블랙 박스'로 여겨져서 그 결정 과정이 상담자에게 명확하게 설명되기 어렵다는 단점이 있다. 이러한 해석 가능성의 부족은 상담 상황에서의 실질적인 의사결정 과정에서 신뢰성을 떨어뜨릴 수 있다.

2.2 머신러닝 기반 측정 모델

머신러닝 모델은 복잡한 딥러닝 방식과는 대조적으로, 직관적이고 해석 가능한 결과를 제공한다. 이러한 접근법은 주어진 데이터셋에서 패턴을 학습하고 이를 기반으로 새로운 데이터에 대한 예측이나 분류를 수행한다.[7] 예를 들어, Liu [8]의 연구는 눈동자 움직임 패턴을 머신러닝(SVM)을 통해 분석하고 이를 통해 자폐 스펙트럼을 식별하는 방법에 대해 다룬다. 이 방법은 복잡한 데이터 분석이 필요하지 않으며, 사용자 인터페이스를 통해 얻은 눈동자 데이터만으로도 자폐 스펙트럼의 식별이 가능하다는 점에서 혁신적이다. 이는 머신러닝 방법으로도 얼굴 특징 분석이 가능함을 시사한다.

상담자는 이러한 모델을 사용하여 내담자의 신체 언어, 표정 등을 분석하고, 상담 과정 중 발생할 수 있는 다양한 심리적, 감정적 상태를 판단할 수 있다. Villaroya [9]의 연구에서는 학생들의 표정을 분석하여 참여도를 자동으로 인식하는 방법을 탐색한다. 학생들의 표정을 분류하고, 이 데이터를 머신러닝을 사용하여 분석한다. 이 연구는 교육적 환경에서 학생들의 표

정을 통해 참여도를 인식하는 자동 시스템의 가능성을 보여준다. 이러한 시스템은 교사들이 온라인 교육 환경에서 학생들의 참여도에 대한 실시간 피드백을 받는 데 도움이 될 수 있다.

이러한 머신러닝 모델들은 실시간 분석의 필요성과 상담자의 해석 및 응용 능력을 강조하는 상담 분야에 특히 적합하다. 상담자는 이 모델들을 활용하여 내담자의 행동과 감정을 더욱 빠르고 직관적으로 이해할 수 있으며, 상담 과정을 더욱 효과적으로 이끌어갈 수 있다.

따라서, 본 연구에서는 내담자의 비언어적 신호를 분석하고 평가하는 데 있어 머신러닝 기반의 접근법을 채택하였다. 이는 교육자의 입장에서는 딥러닝 모델의 블랙박스 특성보다는 머신러닝이 제공하는 실시간 해석이 더 유용하다는 판단에 기반한다. 머신러닝 모델은 딥러닝에 비해 상대적으로 낮은 계산 복잡성을 가지며, 상담자에게 이해하기 쉽고 실시간 상황에 적합한 해석을 제공한다.

본 연구는 이러한 머신러닝 모델의 구현과 실제 상담 환경에서의 적용 가능성에 대해 집중적으로 탐구할 것이며, 상담의 질을 향상하는 데 기여할 것으로 기대된다.

2.3 집중도 분석 데이터셋

기존의 여러 연구에서는 얼굴 특징을 분석하여 집중도나 졸음을 판단하는 다양한 데이터셋을 활용하였다.

YawDD(Yawning Detection Dataset)[10]은 운전자의 졸음 상태를 감지하기 위해 하품을 탐지하는 데 중점을 둔 데이터셋이다. 이 데이터셋은 다양한 얼굴 특성을 가진 운전자들이 두 가지 조건에서 촬영된 비디오로 구성되어 있다. 첫 번째 데이터셋은 차량 전면 거울 아래에 설치된 카메라로 촬영되었으며, 각 참가자는 정상, 대화/노래, 하품 등의 다양한 입 상태를 보여주는 세 개 또는 네 개의 비디오를 가지고 있으며 총 320개의 비디오가 있다. 각 비디오는 평균적으로 약 700프레임의 길이로 구성되었다. 두 번째 데이터셋은 운전자 앞 대시보드에 설치된 카메라로 촬영되었으며, 29명의 참가자는 다양한 입 상태를 포함한 한 개의 약 2000프레임으로 촬영하였다.

DAiSEE(Dataset for Affective States in E-Environments)[11]는 온라인 학습 환경에서 사용자의 지루함(Bored), 어리둥절함(Confused), 집중(Engaged), 짜증

(Frustrated) 등의 정서 상태를 인식하기 위해 개발된 데이터셋이다. 112명의 사용자로부터 캡처된 9,068개의 비디오로 구성되어 있으며, 이는 지루함, 어리둥절함, 집중, 짜증 상태를 네 가지 수준(매우 낮음, 낮음, 높음, 매우 높음)으로 분류된다. 데이터셋의 각 비디오는 대략 300프레임 길이이다.

기존 데이터셋들은 각각의 특성을 가지고 있지만, 장기간에 걸친 연속적인 표정 변화나 집중력 변화의 분석에는 한계가 있다. yawDD의 첫 번째 데이터셋은 평균 700프레임, 두 번째 데이터셋은 약 2000프레임으로 구성되었으나, 샘플 수가 한정적이다. 또한, DAiSEE의 각 비디오는 약 300프레임으로, 짧은 시간 동안의 상태 변화만을 포착한다. 이에 반해 본 연구에서는 300프레임보다 긴 시간 동안의 데이터를 포함하는 새로운 데이터셋을 구축하여, 더욱 긴 시간 동안의 변화를 포착하여 집중도와 얼굴 특징 분석의 정밀도를 향상하고자 한다. 이를 통해 DAiSEE와 새로운 데이터셋의 프레임 길이에 따른 모델 성능 차이를 비교 분석하며, 프레임 길이가 모델의 정확도에 미치는 영향을 평가할 것이다.

3. 얼굴 특징 분석

집중도 분석을 위한 얼굴 특징 분석은 얼굴 랜드마크를 기반으로 한다. 이 랜드마크들은 눈, 코, 입, 턱선 등 얼굴의 중요한 부위들을 정확하게 식별하는 데 사용되며, 이러한 랜드마크들을 통해 얼굴의 다양한 특징점들을 파악한다. 특징점들은 행동 패턴을 탐지하는 데 사용된다.

3.1 68개의 얼굴 랜드마크

얼굴의 68개 특징점[12]은 얼굴 인식 및 분석 분야에서 중요한 도구이다. 이는 정밀한 방식으로 얼굴의 주요 부위들을 탐지하고 추적하는 데 사용되며, 이를 통해 얼굴의 기하학적 구조를 깊이 있게 분석한다. 특히 눈, 코, 입, 턱선 등의 위치를 정확하게 식별하는 데 매우 유용하며, 이를 통해 얼굴의 다양한 표정과 감정 상태를 세밀하게 파악할 수 있다.

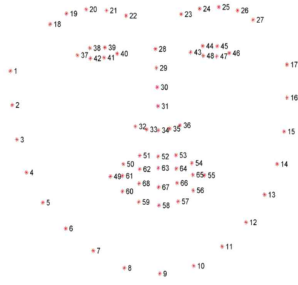


Figure 1. 68 Face Landmark

이러한 68개의 얼굴 특징점들을 활용하여 상담 상황에서 발생할 수 있는 다양한 비언어적 신호들을 추출하고 분석하고자 한다.

3.2 눈 깜빡임(blink) 탐지

눈 깜빡임의 빈도와 지속 시간은 피로도도와 집중도의 중요한 지표이다. dlib[13]의 특징점을 사용하여 눈 주변의 변화를 감지하고, 이를 분석하여 내담자의 깜빡임 패턴을 추출한다. dlib의 face landmark에 따르면, 36번부터 41번까지의 특징점은 왼쪽 눈을, 42번부터 47번까지는 오른쪽 눈을 나타낸다.

분석과정에서는 이 특징점들에 해당하는 이미지를 (24,24) 크기로 조정하여 처리한다. 전처리된 이미지를 사전에 학습된 모델[14]을 사용하여 눈의 개폐 상태를 판단한다. 양쪽 눈이 모두 감긴 상태를 눈을 감은 걸로 판단하였고, 눈을 감았다고 판단되는 프레임이 연속으로 2개 이상 나타나면 눈을 감기 시작했다고 판단했다.

3.3 하품(yawn) 탐지

하품 탐지는 MAR(Mouth Aspect Ratio) 계산법[15]을 통해 측정하였다. MAR 계산은 입 부분의 특징점들을 이용하여 입의 형태 변화를 정량적으로 분석하는 방법이다. 이를 위해 dlib face landmark 중 입과 관련된 부분을 추출하고, 이를 기반으로 MAR을 계산한다.

Figure 2에는 하품 탐지 과정에서 사용되는 요소들이 나타나 있다. Figure 2는 입 부분에 해당하는 facial landmark를 보여주며, 이 특징점들을 활용하여 입의 형태 변화를 감지한다. (식 1)은 이러한 특징점들을 기반으로 한 MAR 계산식을 나타낸다. MAR 계산식은

입의 가로 길이와 세로 길이의 비율을 계산하여 입이 얼마나 크게 열렸는지를 수치화한다.

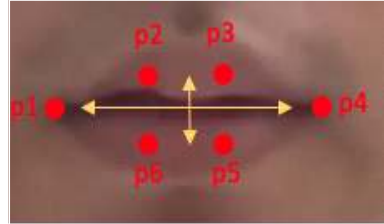


Figure 2. MAR 계산을 위한 랜드마크

$$MAR = \frac{\|p_2 - p_6\| + \|p_3 - p_5\|}{2\|p_1 - p_4\|}$$

(식 1) MAR 계산식

본 연구에서는 MAR 값의 임계값을 0.6으로 설정하였다. 이는 입을 벌렸다고 판단하는 기준으로 사용되며, MAR값이 이 임계값을 초과할 경우 입을 벌렸다고 판단한다. 또한, 입을 계속 벌리고 있는지를 판단하기 위해 연속 프레임 임계값을 10으로 설정하였다. 즉, 연속적으로 10개 이상의 프레임에서 MAR 값이 0.6을 초과할 경우 하품으로 분류한다.

3.4 멍때림(space out) 탐지

내담자의 멍때림을 탐지하기 위해서 동공의 위치를 파악한다.[16] 이 과정은 양쪽 눈 동공의 위치(x,y 좌표)를 정밀하게 추출하고, 연속적인 프레임에서 이 위치의 변화를 분석하여 판단한다. 이전 프레임과의 동공 위치 사이의 유클리드 거리를 계산하여 양쪽 눈 모두 계산값이 3보다 작은 경우, 동공의 움직임이 없다고 간주한다. 이는 내담자가 멍하니 있거나 집중하지 않고 있다는 신호로 해석될 수 있다. (식 2)에서는 이러한 분석을 위해 사용되는 유클리드 거리 계산식을 나타낸다. 이 계산식은 연속적인 프레임 사이에서 동공 위치의 변화를 정량적으로 측정하는 데 사용된다.

$$d(P, Q) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

(식 2). 유클리드 거리 계산식

동공 움직임이 없는 상태가 1800프레임(약 60초) 동안 지속될 경우, 내담자가 멍때리고 있다고 판단한다. 이러한 긴 시간 동안 동공 움직임이 없는 것은 일반적인 주의 집중 상태와는 다르며, 피로 또는 집중력 저하의 가능성을 나타낸다.

3.5 얼굴 탐지 백분율

본 연구에서는 dlib을 이용해서 얼굴 탐지를 진행하는데, 이를 통해 얼굴이 탐지된 프레임의 비율을 계산한다. 비디오의 각 프레임을 분석하여 얼굴이 탐지되었는지를 확인하고, 이 데이터를 바탕으로 전체 프레임 대비 얼굴이 탐지된 프레임의 비율을 계산한다. 이 변수에 저장된 값은 상담 상황에서 내담자의 얼굴이 카메라에 얼마나 자주 노출되었는지를 나타내는 지표이다. 이 값이 낮은 경우, 내담자가 카메라에서 자주 벗어나거나 얼굴이 제대로 포착되지 않고 있음을 의미한다. 이는 내담자의 집중도가 낮거나 상담에 적극적으로 참여하지 않고 있음을 예상해 볼 수 있다.

3.6 오일러 각도

오일러 각도는 3차원 공간 내에서 물체의 방향을 3개의 서로 수직인 X,Y,Z축 각도로 표현하는 방법으로, 본 연구에서는 이를 통해 내담자의 머리 움직임을 분석한다.[1] 이 특징은 내담자의 주의 집중 및 관심도 변화를 정량적으로 평가하는 데 중요하다. 예를 들어, 수직축(yaw) 각도는 좌우로 머리를 돌리는 움직임을 반영하고, 횡축(pitch) 각도는 상하로 머리를 움직이는 동작을 나타내고 종축(roll) 각도는 머리가 양 옆으로 기울어진 정도를 나타낸다. 이러한 각도별 분석을 통해 내담자의 행동 패턴과 집중력 변화를 보다 세밀하게 이해할 수 있으며, 상담의 질을 향상시키는 데 기여할 수 있다.

각 축에 대한 head turn counts 변수를 설정하여, 머리가 바라보는 방향의 변화를 수치화한다. 이는 각 축에 대해 설정된 임계치를 기준으로 판단되며, 방향이 변경될 때마다 해당 축에 대한 특징값이 증가한다. Figure 3에서 수직축, 횡축, 종축에 해당하는 축을 나타내고 Table 1에서는 각 축별 임계치를 나타낸다. 또한 head turn counts를 계산할 때, 머리의 방향이 5프레임 이상 변경되어야 새로운 방향 전환으로 카운트된다. 이를 통해 짧은 시간 동안의 미세한 움직임보다

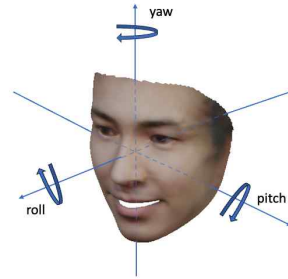











Figure 3 . Euler 각도

Table 1. 각 축별 임계값

yaw		
		
Looking left	Looking forward	Looking right
$yaw < -30$	$-30 \leq yaw \leq 30$	$30 < yaw$
pitch		
		
Looking up	Looking forward	Looking down
$0 < pitch < 170$	$170 \leq pitch < 180$ $-180 \leq pitch < -170$	$-170 < pitch < 0$
roll		
		
Tilted left	Looking forward	Tilted right
$10 < roll$	$-10 \leq roll \leq 10$	$roll < -10$

는 의미 있는 머리 움직임의 변화를 정확하게 포착한다.

또한, 각 축별로 중앙을 바라보고 있는 프레임의 수의 백분율 역시 중요한 데이터이다. 이 특징값들은 내담자가 각 축에 대해 중앙 방향을 얼마나 오랫동안 유지하는지를 나타낸다.

3.7 머리 기울기 평균

본 연구에서는 Euler 각도 분석과 별도로 머리의 기울기 평균값을 추가적인 특징으로 사용하였다. Euler 각도 분석을 통한 바라보는 방향 분석은 내담자의 정면 시선을 바라보는 백분율과 바라보는 방향이 달라지는 횟수를 파악하는 데 중요하지만, 머리 기울기 평균은 머리의 기울기를 평균적으로 측정하여 더 세부적인 자세 변화를 포착한다. 이는 내담자가 장시간 집중하거나 피로할 때 나타나는 미묘한 머리 기울기 변화를 정량적으로 분석하는 데 유용하다.

측정은 face landmark를 바탕으로 왼쪽 눈과 오른쪽 눈의 중앙 좌표를 정확하게 파악하는 데서 시작한다. 두 눈의 중앙 좌표 간의 x,y좌표의 차이를 계산하고, 이 차이를 (식 3)에 나와있는 arctan함수에 적용하여 얼굴의 기울기 각도를 계산한다. 이렇게 계산된 얼굴 각도의 절대값이 큰 경우, 내담자가 얼굴을 기울이고 있다는 것을 뜻하므로 집중하지 않고 있다고 예측할 수 있다.

$$\arctan(x) = \tan^{-1}(x)$$

(식 3) arctan 함수

4. 연구 방법

4.1 데이터 수집 및 전처리

본 연구에서는 DAiSEE 데이터셋과 새로 구축한 데이터셋을 비교 분석하여, 프레임 길이가 모델 성능에 미치는 영향을 평가하고자 한다. 이를 통해 짧은 프레임 데이터셋과 긴 프레임 데이터셋 간의 차이점과 이로 인한 분석 결과의 차이를 명확히 이해할 수 있을 것으로 기대된다.

DAiSEE 데이터셋은 온라인 학습 환경에서 학생들의 감정 상태를 분석하기 위해 개발되었으며, 각 비디

오는 약 300프레임의 길이로 구성되어 있다. 데이터 전처리 과정에서는 각 상태별 가중치 중 최대값이 3 이상인 데이터만 선별하였다. 그 중에서 집중(Engage d) 상태의 가중치가 최대값인 경우에만 Target 값을 0으로 설정하고, 나머지 상태의 가중치가 최대값인 경우에는 Target 값을 1로 설정하였다. 이러한 전처리 과정을 거친 결과, Train dataset에서는 총 2,388개의 비디아가 선택되었으며, Test dataset에서는 총 763개의 비디아가 선택되었다.

비교 분석을 위해 새로 구축한 데이터셋(New Attention Analysis Dataset, NAAD)은 총 19명의 학생들을 대상으로 수집되었으며, 이들 각각에게 집중하는 상황과 집중하지 않는 상황에 해당하는 두 가지 유형의 영상을 수집하였다. 각 학생은 5분 길이의 집중하는 영상과 5분 길이의 집중하지 않는 영상을 촬영했으며 총 190 분량의 비디오 데이터가 수집되었다. 영상은 초당 30프레임의 비디오를 웹캠으로 캡처하였다.



(a)Engagement

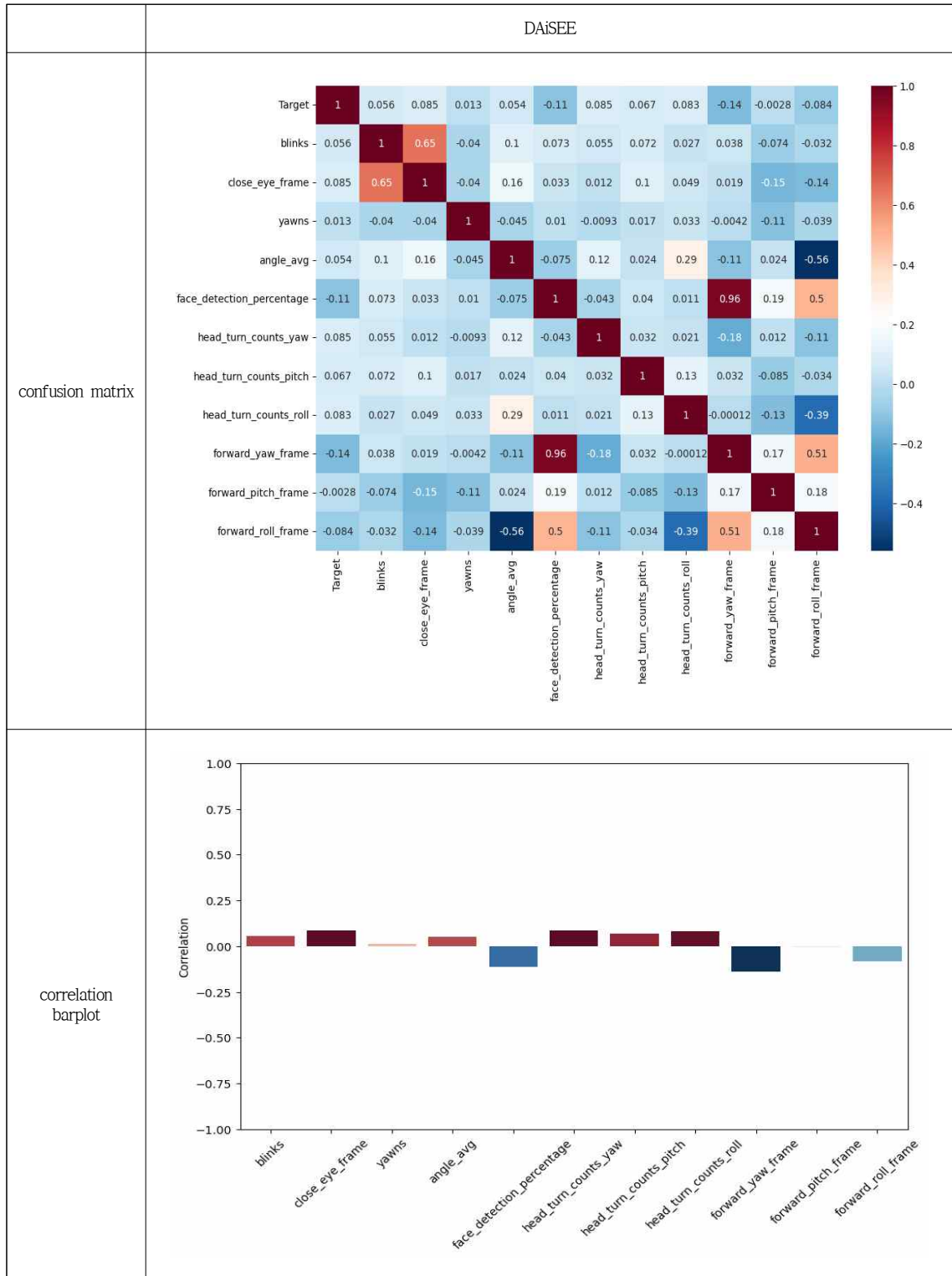


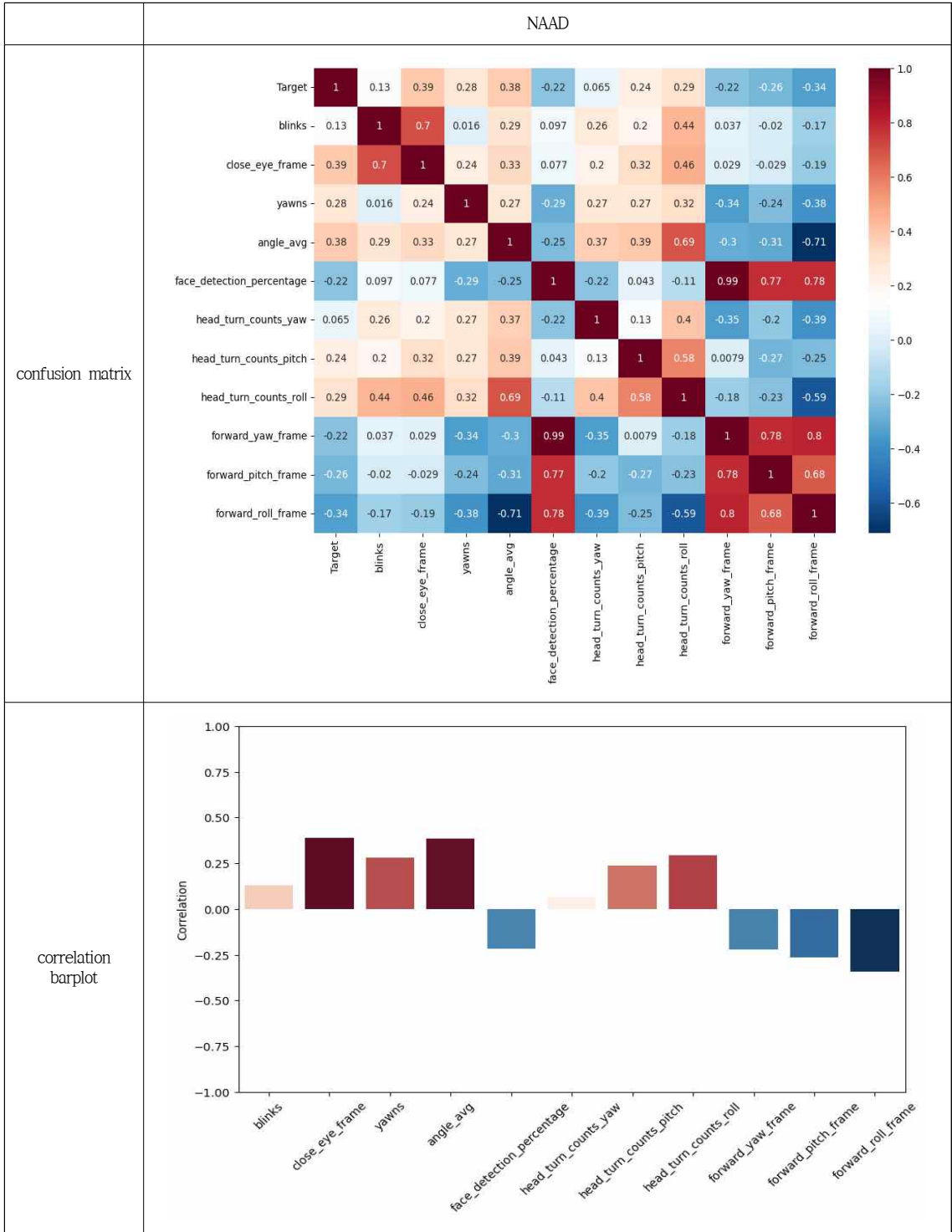
(b)non-Engagement

Figure 4. NAAD

이 원본 영상을 2000프레임 단위의 서브 영상들로 나눈다. 이때 step size는 300프레임으로 설정하였다. 집중한 영상이면 Target 값을 0으로 설정하고, 집중하지 않은 영상이면 Target 값을 1로 설정하였다. 수집한 데이터들 중 4명에 해당하는 서브 비디오 180개의 영상들을 test dataset으로 지정하고, 나머지 인원에 대한 740개의 영상들을 train dataset으로 지정하였다.

전처리 과정을 거친 두 데이터셋(DAiSEE, NAAD)에 대해, 모든 영상들에서 다양한 특징값들을 추출하여 학습을 진행하였다. 추출된 특징값들은 머신러닝 모델을 이용한 집중도 예측에 사용된다.





(b) NAAD 상관관계 분석

Figure 5. 데이터셋별 특징 분석

4.2 데이터셋 특징 비교

DAiSEE와 새로 구축한 데이터셋의 특징값들과 Target 값 간의 상관관계를 분석하여, 두 데이터셋의 성능과 특징 추출 능력의 차이를 명확히 평가하고자 한다.

DAiSEE 데이터셋에서 상관 분석을 진행한 결과, 대체적으로 낮은 상관관계 값을 보여준다. 가장 높은 상관관계 값이 매우 낮아, 이 데이터셋의 짧은 프레임 길이

가 복잡한 얼굴 특징의 변화를 정확하게 포착하는데 한계가 있음을 나타낸다.

NAAD에서의 상관관계 분석 결과는 DAiSEE 데이터셋과 비교하여 상대적으로 더 높은 상관관계를 보여준다. 가장 높은 상관관계를 나타낸 눈을 감고 있는 프레임 수(close_eye_frame)와 머리 기울기 평균(angle_avg)은 상당히 높은 수치의 양의 상관관계를 나타낸다. 특히, 눈을 감은 프레임 수를 기록하는 close_eye_frame과 하품 횟수를 기록하는 yawns 특징값들이 DAiSEE 데이터셋과 비교했을 때 상관관계 수치가 커졌음을 확인할 수 있다. close_eye_frame이나 yawns 특징들은 긴 프레임동안 포착되는 특징들로, 이는 긴 프레임 수가 연속적인 행동 패턴을 더 정확하게 포착하는 데 유리하다는 것을 시사한다.

제 양성 케이스 중 모델이 올바르게 양성으로 예측한 비율을 나타내며, 모델이 양성 케이스를 감지하는 능력을 측정한다. 높은 recall은 거짓 음성(false negative)의 수가 적음을 의미한다. F1-score는 precision과 recall의 조화 평균을 나타내며, 이는 두 지표의 균형을 나타내는 중요한 척도이다. 높은 F1-score는 모델이 정밀도와 재현율을 모두 고려하여 더 균형 잡힌 성능을 보여줌을 의미한다. Accuracy(정확도)는 모든 예측 중 올바른 예측의 비율을 나타내며, 모델의 전체적인 성능을 간단하게 파악할 수 있도록 해준다.

이러한 다양한 지표들을 사용하는 것은 모델의 성능을 다각도에서 평가하는 데 유용하기 때문에 본 연구에서는 모델의 성능을 보다 포괄적이고 균형 잡힌 방식으로 평가하고자 하였다.

Table 2. 학습 데이터-평가 데이터별 성능

Train dataset	Test dataset	pre.	recall	f1-score	Acc.
DAiSEE	NAAD	0.76	0.53	0.41	0.53
NAAD	DAiSEE	0.87	0.60	0.70	0.60
NAAD	NAAD	0.86	0.82	0.82	0.82
DAiSEE	DAiSEE	0.85	0.87	0.86	0.87

Table 2에는 Train dataset으로 학습된 모델을 Test dataset을 통해 성능을 나타내봤을 때, precision, recall, f1-score, accuracy를 나타내고 있다. 이 결과에 따르면, NAAD로 학습한 후 DAiSEE으로 평가했을 때의 성능이 더 우수했다. 이는 DAiSEE로 학습하여 NAAD로 평가한 경우보다 모든 성능 분석 지표에서 높은 수치를 나타냈다. 이러한 결과는 긴 프레임 데이터셋이 연속적인 행동 패턴을 더 정확하게 포착하여 모델 성능을 향상시키는 데 기여함을 시사한다.

5. 연구 결과

5.1 교차 검증

본 연구에서는 교차 검증 방법 중에서도 외부 검증을 사용하여 데이터셋의 프레임 길이에 따른 모델 성능의 영향을 평가하였다. 이 방식을 통해, 모델이 특정 데이터셋에 지나치게 최적화되지 않고 다양한 상황에서도 일관된 성능을 발휘하는지를 평가하고자 하였다. 이를 위해 기존 연구[9]를 참고하여 선택한 Random Forest 모델을 선택하고, NAAD와 DAiSEE 데이터셋을 대상으로 학습과 평가를 진행하였다.

모델의 성능을 평가하기 위해 precision, recall, f1-score, 그리고 accuracy 지표들을 사용하였다. Precision(정밀도)는 모델이 양성으로 분류한 케이스 중 실제로 양성인 경우의 비율을 나타내며, 이는 모델이 얼마나 정확하게 양성을 예측하는지를 측정한다. 높은 precision은 모델이 적은 수의 거짓 양성(false positive) 결과를 생성함을 의미한다. 반면, Recall(재현율)은 실

5.2 테스트 UI

본 연구의 중요한 단계 중 하나는 집중도 분석을 위한 사용자 인터페이스(UI)의 개발이다. 이 UI의 주요 목적은 사용자로부터 영상을 입력받아 해당 영상 내 개인의 집중도를 효과적으로 분석하고 결과를 제공하는 것이다. 이를 통해 집중력 분석 과정을 사용자에게 보다 접근하기 쉽고 효율적으로 만드는 것을 목표로 하였다.

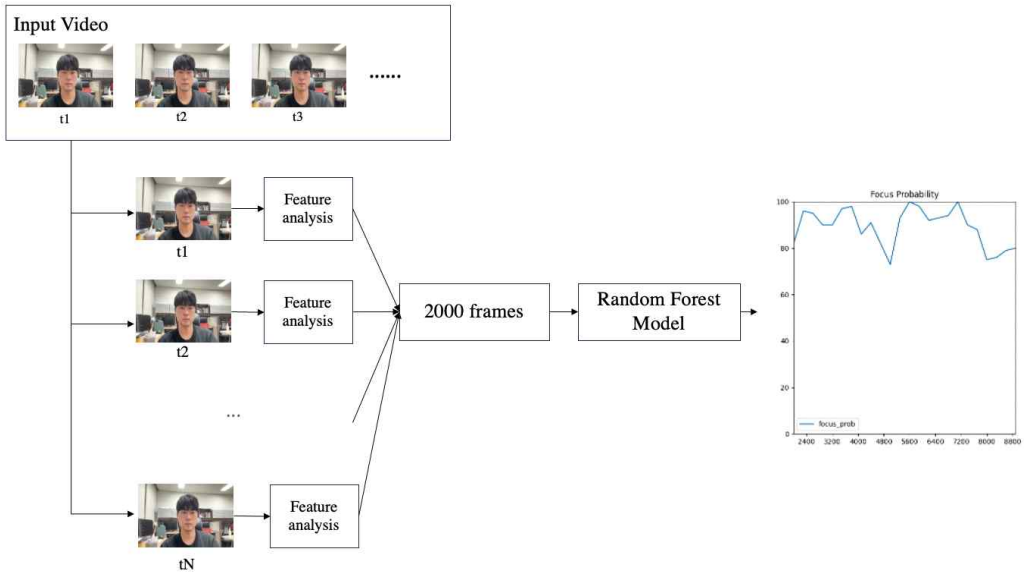


Figure 6. 집중도 분석 과정

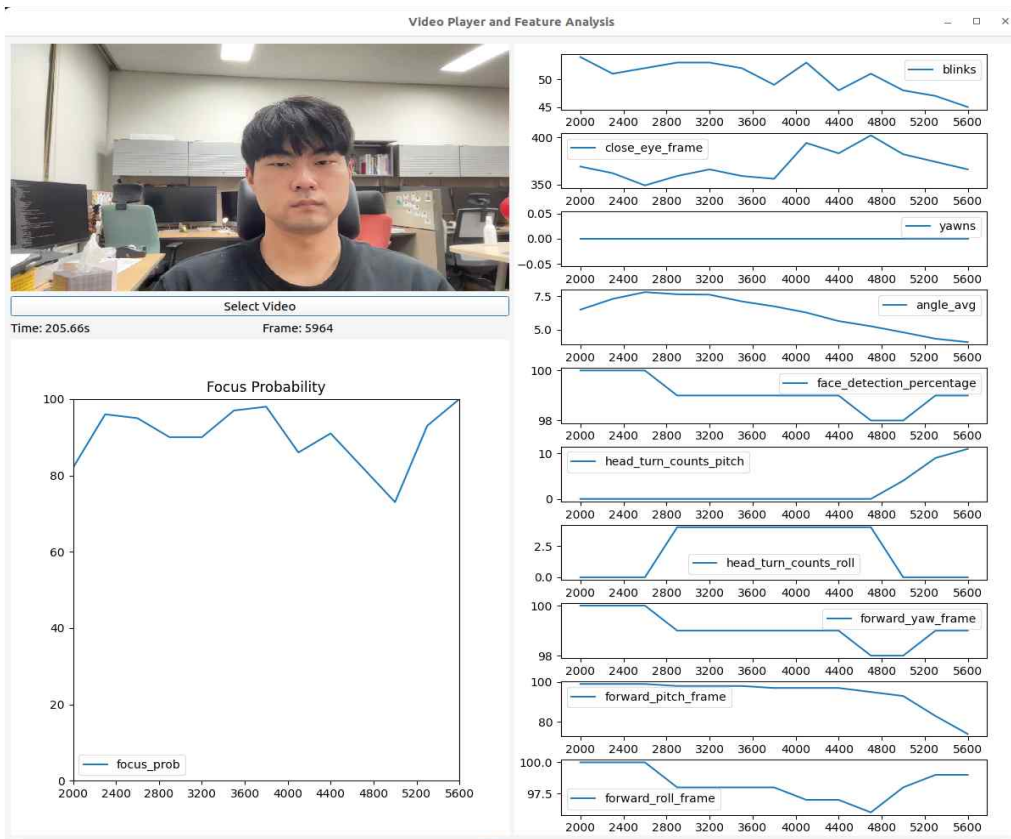


Figure 7. 집중도 분석 UI

사용자는 이 인터페이스를 통해 영상 파일을 업로드할 수 있다. Figure 6에 나타난 분석 과정을 보면 업로드된 영상에 대해서 2000프레임 동안 얼굴 특징 분석을 한다. 2000프레임 동안 얼굴 특징 분석을 마치면 모인 데이터를 기반으로 학습된 Random Forest 모델을 통해 집중도 분석을 진행한다. 분석 과정은 영상의 처음 2000프레임 동안 모인 데이터를 기반으로 초기 집중도 점수를 예측한다. 매 프레임 마다 집중도 예측을 하는 것이 아닌 2000프레임마다 집중도 예측을 하는 이유는 영상 내에서 연속성을 가지는 특성값들을 충분히 수집하고 집중도를 보다 정확하게 판단하기 위함이다. 얼굴 분석을 시작한 후 매 300프레임마다 새롭게 2000프레임 동안 얼굴 분석 데이터를 모으고, 모인 데이터를 바탕으로 집중도 점수를 다시 예측하고 이를 사용자에게 그래프 형태로 시각적으로 표시한다. 이렇게 함으로써 사용자는 영상 내에서 집중도의 변화를 파악할 수 있다.

Figure 7에 나타난 UI 화면에서 왼쪽 아래 그래프는 예측된 집중도 점수에 대한 그래프를 나타내고, UI의 오른쪽에는 각 특성값들에 대한 그래프를 나타낸다.

6. 결론

본 논문에서는 온라인 상담을 받는 학생들의 집중도를 평가하기 위해 DAiSEE 데이터셋처럼 300 프레임의 짧은 데이터셋보다 새로 구축한 2000 프레임의 긴 데이터셋을 사용하는 것이 더 효율적임을 제안하였다. 이 연구는 얼굴 인식 기술을 기반으로 하며, 얼굴 이미지 프레임에서 다양한 특징값들을 추출하여 사용한다. 주요 특징으로는 눈 깜빡임, 하품, 동공의 움직임, 얼굴의 각도, 얼굴 탐지 비율, 오일러 각도 등이 포함되며, 이들을 통해 집중도의 변화를 정확하게 파악할 수 있다.

데이터셋의 길이에 따른 모델 성능을 비교하기 위해 외부 검증 방법을 사용하였다. 검증 결과 2000 프레임의 데이터셋으로 학습된 모델이 DAiSEE 데이터셋으로 학습된 모델보다 높은 성능을 보여줌으로써, 영상 길이가 긴 데이터셋이 모델의 성능을 향상시키는 데 기여함을 확인할 수 있었다. 이러한 결과는 연속적인 행동 패턴을 포착하는데 긴 프레임 데이터셋이 중요한 역할을 한다는 점을 강조한다.

또한 사용자 인터페이스(UI)를 개발하여, 영상을 업

로드하고 집중도를 효과적으로 분석 및 표시할 수 있게 하였다. 이 UI는 사용자에게 영상 내의 집중도의 변화를 프레임에 따라 세밀하게 파악할 수 있는 기능을 제공한다.

향후 연구에서는 많은 데이터셋을 수집하여 모델의 적용성과 정확도를 향상시킬 예정이다. 또한 현재 연구에서 주로 활용된 머신러닝 접근법에 더하여, 복잡한 데이터 패턴과 대규모 데이터셋을 효율적으로 처리할 수 있는 딥러닝 모델을 탐색하고 적용해 볼 예정입니다. 이러한 접근은 다양한 데이터와 문제 상황에 대응하는 능력을 강화할 것입니다.

참고문헌

- [1] Yurayat, P., Seechaliao, T. (2022), Undergraduate Students' attitudes towards online counseling since the COVID-19 pandemic. *Higher Education Studies*, 12(1), 72. DOI: 10.5539/hes.v12n1p72.
- [2] Stommel, W., van der Houwen, F. (2014). Complaining and the management of face in online counseling. *Qualitative Health Research*, 24(2), 183-193. DOI: 10.1177/1049732313519706.
- [3] Chirra, V., ReddyUyyala, S., KishoreKolli, V. (2019). Deep CNN: A machine learning approach for driver drowsiness detection based on Eye State, *Revue d'Intelligence Artificielle*, 33(6), 461-466. DOI: 10.18280/ria.330609.
- [4] Magán, E. et al. (2022). Driver drowsiness detection by applying deep learning techniques to sequences of images. *Applied Sciences*, 12(3), 1145. DOI: 10.3390/app12031145.
- [5] Su, M.-C. et al. (2021). A video analytic in-class student concentration monitoring system. *IEEE Transactions on Consumer Electronics*, 67(4), 294-304. DOI: 10.1109/tce.2021.3126877.
- [6] Solanki, N., Mandal, S. (2022). *Engagement Analysis Using daisee dataset*. 2022 17th International Conference on Control, Automation, Robotics and Vision (ICARCV). doi:10.1109/icarcv57592.2022.10004250.
- [7] Janiesch, C., Zschech, P. and Heinrich, K. (2021). Machine learning and deep learning. *Electronic Markets*, 32(3), pp. 685-695. DOI: 10.1007/s12525-021-00475-2.
- [8] Liu, W., Li, M. and Yi, L. (2016). Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework. *Autism Research*, 9(8), pp. 888-898. DOI: 10.1002/aur.

1615.

- [9] Villaroya, S.M. et al. (2022). *Real-time engagement detection from facial features*. 2022 IEEE International Conference on Development and Learning (ICDL) [Preprint]. DOI: 10.1109/icdl53763.2022.9962228.
- [10] Abtahi, S. et al. (2014). *Yawdd* Proceedings of the 5th ACM Multimedia Systems Conference [Preprint]. DOI: 10.1145/2557642.2563678.
- [11] Gupta, A. et al. (2022). Daisee: Towards user engagement recognition in the wild. *arXiv.org*. Available at: <https://arxiv.org/abs/1609.01885>
- [12] Sagonas, C. et al. (2016) ‘300 faces in-the-Wild Challenge: Database and Results’, *Image and Vision Computing*, 47, pp. 3-18. DOI: 10.1016/j.imavis.2016.01.02.
- [13] Davis E. King. (2009). Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10, 1755-1758
- [14] yoshidan. (2019). *pytorch-eyeblick-detection*. San Francisco(CA) : Github; <https://github.com/yoshidan/pytorch-eyeblick-detection>
- [15] Zhu, T. et al. (2022). Research on a real-time driver fatigue detection algorithm based on facial video sequences. *Applied Sciences*, 12(4), p. 2224. DOI: 10.3390/app12042224.
- [16] antoinelame. (2022). *GazeTracking*. San Francisco (CA) : Github; <https://github.com/antoinelame/GazeTracking>
- [17] Wang, Y., Zhou, W. and Zhou, J. (2023). 2dheadpose: A simple and effective annotation method for the head pose in RGB images and its dataset. *Neural Networks*, 160, 50-62. DOI: 10.1016/j.neunet.2022.12.021.

김 동 욱



2019년 ~ 현재 제주대학교 컴퓨터공학전공
관심분야: 컴퓨터비전, 자연어처리
E-mail: bradl0809@naver.com

김 수 균



2006년 고려대학교 컴퓨터학과 졸업(박사)
2006-2008년 삼성전자 책임연구원
2008-2020년 배재대학교 게임공학과

2022년 ~ 현재 제주대학교 컴퓨터공학전공 조교수
관심분야: 그래픽스, 컴퓨터비전
E-mail: nicesk@gmail.com

윤 여 찬



2004년 고려대학교 컴퓨터학과 졸업(학사)
2007년 고려대학교 컴퓨터학과 졸업(석사)
2020년 고려대학교 컴퓨터학과 졸업(박사)
2007-2022년 한국전자통신연구원 책임연구원

2022년 ~ 현재 제주대학교 인공지능전공 조교수
관심분야: 자연어처리, 멀티모달 딥러닝
E-mail: ycyoon@jejunu.ac.kr