



# 딥러닝 얼굴 감정 인식 기술을 활용하여 실시간 학습자 집중 상태와 관련한 ITS 개발 및 사용성 평가\*

## Development and usability evaluation of ITS related to real-time learner concentration status using deep learning facial emotion recognition technology

이준형<sup>†</sup> · 송기상<sup>††</sup>  
Jun-Hyeong Lee<sup>†</sup> · Ki-Sang Song<sup>††</sup>

### 요약

인공지능 기술의 발달로 다양한 교육 지원 AI 기술들이 교육 현장에 도입되고 있다. 그러나 인공지능 기반 실시간 학습자 집중도 분석 시스템은 여전히 학교 현장에서 활용되기 어려운 상황이다. 이는 얼굴 감정을 단순히 집중도로 치환하는 연구와 프로그램이 다수 개발되었지만, 실제적으로 현장에 적용되기 위해서는 추가적인 자원과 이론적 준비가 필요하기 때문이다. 이에 본 연구는 얼굴 감정 인식을 기반으로 실시간 집중도를 파악하는 ITS 시스템을 설계·구현하고, 그 사용성을 평가하여 인공지능 기술이 교육자와 학습자에게 실시간 학습 보조 역할을 할 수 있음을 검증하고자 한다. 때문에 본 연구에서는 얼굴 감정 인식을 기반으로 하여 실시간 집중도를 파악하는 ITS 시스템을 설계 및 구현하고, 그 사용성을 평가하여 인공지능 기술이 교육자와 학습자에게 실시간 학습자 학습 보조의 역할을 할 수 있도록 한다.

**주제어** ITS, 얼굴 분류, 학습자 집중도, AI, CNN

### ABSTRACT

With the development of artificial intelligence technology, various educational support AI technologies are being introduced into the educational field. However, it is still difficult to utilize an AI-based real-time learner concentration analysis system in schools. This is because although many studies and programs have been developed to simply replace facial emotions with concentration, additional resources and theoretical preparation are required for practical application in the field. Therefore, this study designs and implements an ITS system that identifies real-time concentration based on facial emotion recognition, and evaluates its usability to verify that artificial intelligence technology can serve as a real-time learning assistant for educators and learners. Therefore, this study designs and implements an ITS system that identifies real-time concentration based on facial emotion recognition, and evaluates its usability to verify that artificial intelligence technology can serve as a real-time learning assistant for educators and learners.

**Keywords** ITS, face recognition, learner concentration, AI, CNN

†정회원 한국교원대학교 대학원 컴퓨터교육학과 박사수료  
††중신회원 한국교원대학교 사범대학 컴퓨터교육과 교수(교신저자)  
논문투고 2024년 10월 19일  
심사완료 2024년 12월 26일  
게재확정 2025년 01월 02일  
발행일자 2025년 01월 08일

\* 본 논문은 2024년 한국컴퓨터교육학회 학술대회에서 "AI 얼굴 감정 인식 기술을 기반으로 학습자의 집중 상태를 분류한 ITS 사용성 평가 사례 연구"의 제목으로 발표된 논문을 확장한 것임.

\* 이 연구는 교육부와 한국연구재단(NRF)의 기초과학연구사업(2021R111A305223413)의 지원을 받아 수행된 연구임.

## 1. 서론

학습자의 상태를 측정하기 위한 연구는 과거부터 지속되어져 왔다. 학습자의 상태를 측정해 즉시적이고 개별화된 피드백은 모든 교육 연구자들의 요구 사항이었다. 그러나 지속적인 관심과 연구에도 불구하고, 현대에 이르기까지 학습자를 위한 지능형 교수 시스템(ITS, Intelligent Tutoring System)[1]이 만족할만한 성과를 내지 못하였다. 그 이유는 기술을 사용하는 지능형 교수 시스템에 대한 교육 이론의 정립 미비와, 기술적인 구현의 어려움, 연산 장비의 연산 처리 속도 부족과, 그에 따른 막대한 시간과 비용 소모로 구현이 어려웠다.

그러나 2010년 이후 인공지능의 머신 러닝 알고리즘이 개선되고, 컴퓨터의 성능 또한 폭발적으로 개선됨에 따라 기존에 해결하지 못하던 문제들을 해결할 수 있게 되었다.

따라서 ITS 원리를 기반으로 하여 네 가지의 주요 요소인 도메인 모듈(Domain Module), 학생 모듈(Student Module), 교수 지도 모듈(Tutoring Module), 인터페이스 모듈(Interface Module)을 바탕으로 실시간 학습자 집중 상태 분류 ITS 서비스를 구현하여, 즉시적인 학습자 피드백 시스템을 구현해보고자 한다.

본 논문에서는 학습자 집중 분류에 딥러닝의 합성곱신경망[2]이 사용되어 학습자의 얼굴 감정 상태를 분류한다. 학습자의 얼굴 감정을 7가지로 분류하고, 학습자의 화면 응시 여부를 판단하여 종합적으로 학습자의 학습 집중 여부를 판단한다.

본 연구의 기여점은 다음과 같다. 첫째로 교육 ITS 이론을 적용한 실시간 학습자 집중 감지 시스템 설계를 구성하고 이를 개발에 적용하였다. 기존의 학습자 집중 감지는 단순히 PC 환경에서 본인의 집중도를 측정해보는 것에 그쳤으나, 본 연구에서는 사용자 인터페이스와 다중접속 네트워크 시스템을 구축하여 실시간으로 학습자와 교수자에게 실시간 피드백을 함께 제공하였다. 그리고 사용성을 평가해 봄으로써, 추후 교수자와 학습자에게 인공지능 학습 보조 서비스로써 도움이 될 것으로 기대한다.

## 2. 이론적 배경

### 2.1 지능형 교수 시스템(Intelligent Tutoring System)

지능형 교수 시스템은 인공지능을 이용하여 지식을 제시하고 학습자와 상호 작용하는 교육 보조 시스템이다[1]. 학습자의 특성에 따라 적절한 학습 내용을 적절한 전달 전략으로 제시하고, 이를 학습자의 학습 능력 및 성취 수준에 따라 융통성 있게 적용한다. 따라서 자체적인 지능적 추론 기능과 인지 모델링을 갖춘 교수 매체로 볼 수 있다[3].

일반적으로 ITS가 다음의 네 가지 구성요소를 갖춘다[4]. 전문가 모듈은 교사가 가르치고자 하는 학습에 관한 지식이 들어있다. 이때 학습에 관한 지식은 학습내용과 관련된 문제를 해결할 수 있는 지식 등을 포함하는 개념이다. 학습자에게 가르치고자 하는 주제에 관한 내용으로 구성된

다. 교수 모듈은 학습자에게 ‘언제 어떤 자료를 제시할 것인가’에 관한 내용을 통제하는 시스템으로 학습자에 대한 진단과 처방, 교수 전략 등의 정보를 갖는다. 학습자 모듈은 학습자들의 반응을 해석하고 평가하며, 이를 바탕으로 학습자에 관한 새로운 정보를 재구성한다. 인터페이스 모듈은 ITS와 학습자가 서로 정보를 주고받으며 학습을 진행할 때 도움을 주는 시스템이다. 즉, 인터페이스 모듈은 학습자와 컴퓨터 간에 양방향 커뮤니케이션을 제공한다[4]. 이러한 ITS를 바탕으로 다양한 교육 보조 시스템들이 개발되어지고 있다.

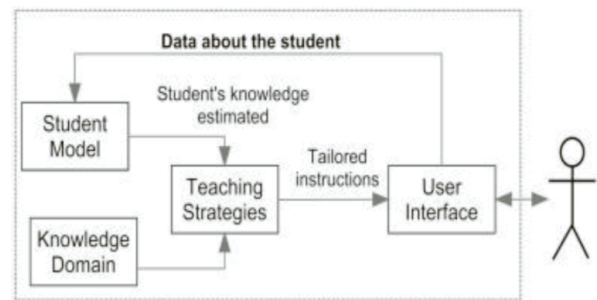


Figure 1. Basic structure of ITS[5]

### 2.2 딥러닝(Deep learning)

딥러닝은 인공신경망(ANN, Artificial Neural Networks)을 기반으로 하는 머신 러닝의 한 분야로, 깊은 신경망을 통해 데이터를 처리하고 학습하는 방식이다. 이러한 대량의 데이터를 통해 패턴을 학습하여, 이미지 인식, 자연어 처리, 음성 인식의 분야에서 활용된다. 이 딥러닝 알고리즘 중 하나인 합성곱 신경망(CNN, Convolution Neural Network)[2]은 인공신경망이 개선된 알고리즘이다. 인공신경망은 디지털 컴퓨터에 인간 두뇌의 신경 연결을 흉내 내어 설계된 것으로 입력 계층, 은닉 계층, 출력 계층으로 구성된다. 은닉 계층에는 인공 뉴런이 있으며 입력 계층으로부터 입력 받은 신호들을 하나의 값으로 조합하고 이를 출력으로 변환시키는 것을 활성화 함수를 통해서 동작하게 된다. 활성화 함수에서는 조합된 입력들이 임계점을 넘기 전까지는 매우 낮은 값을 유지하다가 입력들이 합쳐져서 임계점을 넘기게 되면 활성화되고 출력이 높아지게 된다.

CNN은 인공신경망에 합성곱 계층과 풀링 계층이 추가되어 동작하는 것이다. 합성곱 계층에서는 합성곱 연산을 처리하며 입력 데이터에 필터를 적용한다. 합성곱 연산은 연산의 구간을 일정 간격으로 이동하며 입력 데이터에 적용한다. 이때 편향을 필터로 적용하여 가중치에 더해진다. 풀링은 가로·세로 방향의 공간을 줄이는 연산이다. 이러한 합성곱 신경망은 이미지 인식 분야 뿐만 아니라 다양한 인공지능 분야에서 응용되고 있다. 이러한 계층 풀링을 통해 데이터에 대한 패턴을 학습 가능하게 되며, 현대 AI 기술의 근간이 되고 있다.

### 2.3 집중 상태의 조작적 정의

Cha[6]의 연구에서는 학습자의 집중이란 교수자의 교수에 따라 학생들의 행동으로, 수업에 참여하는 행동, 학습에 몰입하는 모습을 보였을 때 수업에 집중으로 하였다.

교육 기술 연구자는 학습자의 집중력 상태를 평가하기 위해 이미지 데이터 및 시선과 같은 지표를 사용한다. Keith[7]의 연구에 따르면 주의력이 높고 학습자가 집중되어 있을 때 집중 대상에 대해 시선 방향이 고정되어 있는 것으로 나타났다. 또한, 집중력이 높은 기간에는 사람들의 시선이 고정되어 있는 경향이 있지만, 집중력이 감소할 때 시선과 얼굴은 변동하였다[8]. 따라서 이러한 데이터는 학습자의 집중력과 주의력에 대해 유용한 데이터를 제공한다고 볼 수 있다[9].

Cha[6]의 온라인 학습자 집중도 분석 연구에서는 집중과 비집중의 기준을 다음과 같이 정의하였다. 집중이란 ‘학습에 집중하고 있는 상황으로 눈을 뜨고 있는 정면인 상태’로 정의하였고, 비 집중이란 ‘학습에 집중하고 있지 않은 상황으로 눈을 감고 있거나 고개를 숙이거나 고개를 좌우로 돌리는 상태’라고 정의하였다. 해외의 연구에서도 집중의 외향적 얼굴 상태에 대해 다음과 같이 분석하고 집중도 시스템 개발 연구를 진행하였다. 그 내용의 정리는 <표 1>과 같다.

Table 1. Details of papers on concentration classification using CNN

Title	Year	Details
[10]	2019	<ul style="list-style-type: none"> <li>Collect information on eye and head movements and facial emotions in e-learning to classify concentration levels.</li> <li>Divide into three levels: “very engaged,” “nominally engaged,” and “not engaged at all.”</li> </ul>
[11]	2019	<ul style="list-style-type: none"> <li>Classify learners’ concentration levels using FER2013 facial emotion classification</li> <li>Classify students’ concentration levels into three levels based on emotion</li> </ul>
[12]	2021	<ul style="list-style-type: none"> <li>Presents a deep learning-based algorithm that monitors learning ability based on students’ facial emotions</li> <li>Calculates real-time emotions such as anger, disgust, fear, happiness, sadness, and surprise, and calculates concentration weights</li> </ul>
[13]	2021	<ul style="list-style-type: none"> <li>Detects emotions and student engagement during the lecture</li> <li>Suggests student concentration by combining the degree of eye opening and the weight of emotions</li> </ul>
[14]	2021	<ul style="list-style-type: none"> <li>Limitations in real-time AI video processing and scalability for large-scale offline classes</li> <li>Set student facial expressions such as ‘boredom’, ‘confusion’, ‘concentration’, ‘disappointment’, ‘yawning’, and ‘sleepy’</li> </ul>

이를 토대로 다음과 같이 학습자의 집중 상태에 대한 조작적 정의를 <표 2>와 같이 정리하여 정의하였다.

즉, 학습의 집중 상태와 비 집중 상태의 구분은 정면, 고개 돌림, 고개 숙임으로 판단한다. 만약 정면이면 집중, 고개 돌림, 고개 숙임이면 비 집중으로 판단하였다.

Table 2. Operational definition of learner focus in computer vision

	Low Concentration	Normal Concentration			High Concentration		
Facial Recognition	Not Recognized	Face Recognized			Face Recognized		
Face Angle	Side (30 to 45 degrees)	Partial frontal (15 to 30 degrees)			Front (0 to 15 degrees)		
Emotion	Null	Aversion	Anger	Happy	Surprise	Sad	Normal

## 3 시스템 설계 및 구현

### 3.1 시스템 구조

#### 3.1.1 하드웨어와 소프트웨어의 구조

실시간 ITS 서비스 구축을 위해서 2대의 서버를 가동하여 시스템을 구현 하였다. 2대의 서버를 운영한 이유는 다양한 학생들의 이미지 데이터를 실시간 처리해야하므로, 이에 따른 네트워크 통신량의 확보와 통신의 안정적인 서버가 필수적으로 요구된다. 또한 실시간 딥러닝 결과 연산에 따른 고성능 GPU의 서버가 필수적이다.

웹 서버는 접속자의 인터페이스와 상호 작용하는 담당하는 프론트엔드 서버를 목적으로 개발하였다. 프론트 서버는 리눅스 CentOS8, 메모리 16G, CPU intel i9-7900X의 자원을 가지고 있다. 인터넷 속도는 20mbps 이다. 웹 서버 구현을 위한 소프트웨어는 운영 프로그램으로 Eclipse IDE, Tomcat, Mysql 로 구현하였다. 인터페이스 웹페이지는 Java로 구현하였다. Java는 반응형과 확장성을 제공하기 때문에 다양한 환경에서 동일하게 동작하여 다양한 접속 환경에 대응하는 인터넷 인터페이스 페이지에 알맞다. 집중도 분석 웹의 Springboot는 연산과 API환경에 특화되어 간편하게 특정 동작을 수행하는 앱 개발에 알맞다. 또한 자체적인 보안기능도 가지고 있기 때문에, 민감한 데이터를 다루거나 내부의 모델들이 유출에 보안이 필요할 경우 사용한다.

연산 서버는 학습자 이미지를 인공지능으로 분석하여 집중도 값을 연산하여 서버에 송신하는 역할을 한다. 사용자의 실시간 이미지를 처리하는 연산을 담당하도록 백엔드 서버로 구현하였다. 연산 서버의 하드웨어 스펙은 CPU i7-12700, 메모리 32G, GPU RTX 2080Ti 로 개발 되었다. 인터넷 속도는 500mbps 이다. 소프트웨어는 윈도우 10, 파이썬, Keras, Univecon API로 개발하였다.

연산 서버는 이미지를 수신하여 CNN 분석된 이미지를 분류 결과 값을 다시 웹 서버로 리턴 한다. 웹 서버에서는 수신된 결과 값을 데이터베이스에 저장한다.

각 서버에 대한 구조 설계도는 [그림 2]와 같다. 학습자는 웹 서버에 접속하여, 학습자 개인의 고유 데이터 및 이미지 데이터를 송수신한다. 웹 서버는 수신된 이미지 데이터를 연산 서버로 API 통신을 통해 전송하며, 이를 토대로 학습자와 교수자에게 전문 진단을 제공한다.

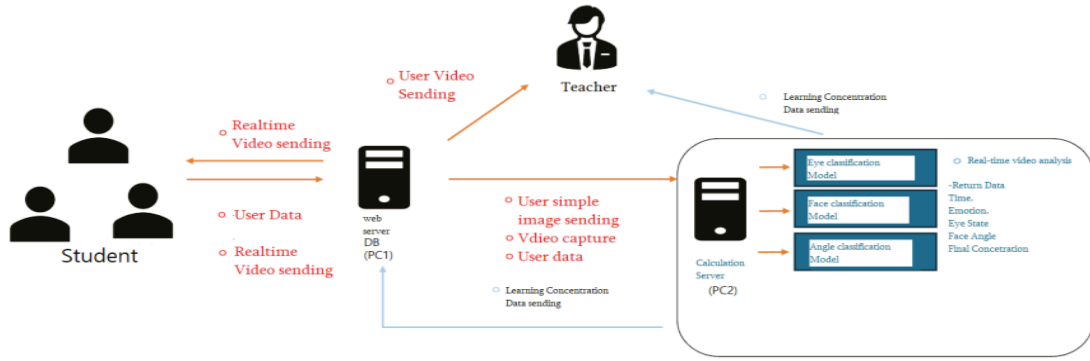


Figure 2. Real-time learner focused state system hardware architecture

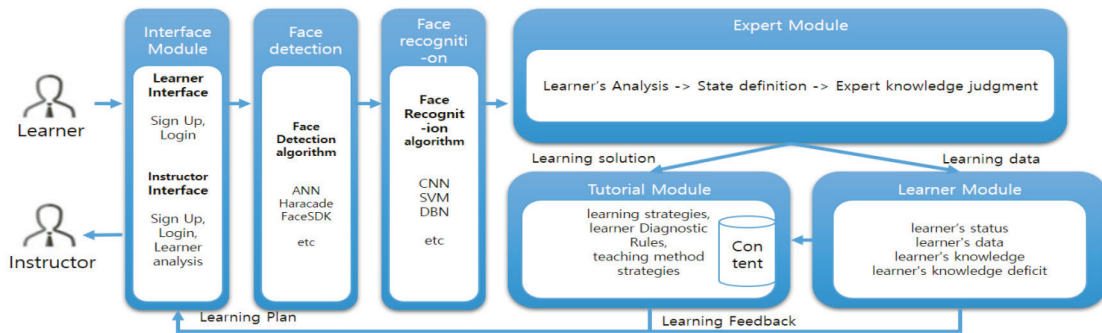


Figure 3. ITS-based real-time learner-focused state system operation structure

### 3.2 모듈 구조

본 연구에서는 ITS에 기반하여 시스템을 4가지 모듈로 나누어 시스템을 구성하였다. 세부 모듈별 내용 구성은 [그림 3]과 같다.

인터페이스 모듈은 학습자와 교수가 접근하여, 각각의 데이터의 수집과 확인을 목적으로 한다. 학습자와 교수는 개인의 고유한 ID를 가지며, 교수는 학습자 전체 및 개인의 집중도를 파악할 수 있도록 인터페이스를 제공한다.

전문가 모듈은 학습자의 집중 상태에 관한 학습자 분석, 상태 분석, 그리고 조작적 정의에 등에 관한 내용이다. 여기서 판단된 데이터는 학습자 모듈은 학습자 데이터 수집, 변인 등에 관련한 모듈이다. 이 시스템에서는 학습자의 얼굴 이미지 영상을 수집하여, 초 단위로 서버에 현재 상태를 데이터베이스에 저장한다.

교수 모듈은 전문가 모듈의 데이터를 토대로 학습자의 상태에 피드백 하는 단계이다. 학습자의 동기나 집중 환기가 될 수 있도록 제공하는 역할을 한다.

### 3.3 CNN 기반의 얼굴 감정 분류에 따른 집중도 값

학습자의 올바른 집중 상태를 측정하기 위해선 신뢰성 있는 모델의 구현이 먼저이다. 기존의 개발된 얼굴 감정 인식 모델들을 동양인에게 적용할 경우 그 인식률이 낮아진

다는 점을 발견했다[20]. Lee[15]의 연구에 따르면 대표적인 서양인으로 구성된 얼굴 감정 데이터셋인 FER2013 데이터셋으로 학습을 하였을 때, 그 예측율은 0.604로 나타나, 이 모델에 동양인의 데이터셋을 예측할 경우 예측율이 0.264로 낮아지는 것을 확인하였다. 즉, 기존의 데이터셋으로 학습할 경우 데이터의 편향성으로 인해 데이터 외의 이미지에 대해서는 이미지 분류 예측도가 낮아지는 것을 확인하였다. 이처럼 편향된 예측율은 Morales[21]과 Darabant[22]의 연구에서도 확인해 볼 수 있다.

때문에 이 문제를 해결하기 위해 이전 연구를 참고하여, 아시아인 얼굴 데이터 세트를 추가하는 것이 필요하거나, CNN 모델의 개선이 필요하다는 것이 확인 되었다[21,22].

또한, Lee[15]의 연구에서는 아시아인 데이터셋을 추가하는 것 뿐만 아니라, 기존의 CNN 모델을 비교 분석함으로써 서양인과 동양인을 포함한 얼굴 감정 인식율에 대해, 높은 예측율을 나타내는 CNN 모델과 데이터셋을 비교하는 것을 확인하였다.

본 연구에 학습으로 쓰인 데이터는 FER2013[16], Aihub[17]에서 제공하는 한국인의 감정 얼굴 이미지, 그리고 JAFFE[18]가 사용 되었다.

이 연구에서 CNN 모델 아키텍처의 입력 이미지는 48\*48 크기의 3개 회색 RGB 채널로 구성되었다. 학습 전에 사전 학습된 'ImageNet'가중치의 전이학습을 진행하였다. 다양한 이미지 입력에 대응하기 위해 이미지 증강을



이용해 0.3 배율, 0~20 각도, 0.3 기울기로 다양한 얼굴 입력 크기에 대응했다. 이 옵션은 모든 이미지 데이터 세트 증강에 동일하게 적용되었다.

각 모델은 배치 크기 64, 에포크 30, 학습율 0.00001 으로 학습되었다. 학습은 데이터 세트의 데이터 불균형에 따라 각 클래스에 가중치를 부여하여 수행되었다.

최종적으로 Lee[15]의 연구에서 FER2013과 동양인 얼굴에 대해 가장 높은 예측값을 보인 CNN 모델인, 개선된 Xception[6]을 사용하였다. 기존 Inception 모델과 달리 Xception 모델은 깊이별 분리형 합성곱을 사용하여 네트워크를 설계한다. 이를 통해 Inception 모델의 합성곱 연산이 더 효율적이 되어 높은 성능을 달성할 수 있다. Xception 모델의 구조는 [그림 4]와 같다. 이 연구에서 사용된 Xception의 개선된 모델의 구조는 [그림 5]와 같다.

Lee[15]의 연구에서 나타난 각 데이터세트와 모델별 얼굴 감정 분류 예측율 비교 표는 <표 3>과 같다.

Table 3. Comparison of facial emotion recognition accuracy by AI model

No	Mo-del	DataSet	FER 2013 Acc	KFE Acc	JAFPE Acc
1-1	Mobi-Net	FER2013	0.371	0.175	0.225
1-2		FER2013,CK+	0.354	0.137	0.235
1-3		FER2013,CK+,KFE	0.352	0.190	0.258
1-4		CK+	0.135	0.146	0.169
1-5		KFE	0.157	0.144	0.160
2-1	Xception	FER2013	0.604	0.264	0.272
2-2		FER2013,CK+	0.600	0.270	0.263
2-3		FER2013,CK+,KFE	<u>0.619</u>	0.328	<u>0.315</u>
2-4		CK+	0.298	0.177	0.235
2-5		KFE	0.326	0.240	0.192
3-1	VGG16	FER2013	0.627	0.235	0.291
3-2		FER2013,CK+	0.621	0.236	0.277
3-3		FER2013,CK+,KFE	0.626	0.314	0.324
3-4		CK+	0.285	0.158	0.254
3-5		KFE	0.371	0.305	0.286
4-1	Den-seNet121	FER2013	0.497	0.252	0.300
4-2		FER2013,CK+	0.509	0.257	0.300
4-3		FER2013,CK+,KFE	0.529	0.253	0.376
4-4		CK+	0.169	0.153	0.174
4-5		KFE	0.178	0.150	0.174
5-1	ResNet50 V2	FER2013	0.492	0.194	0.225
5-2		FER2013,CK+	0.491	0.194	0.224
5-3		FER2013,CK+,KFE	0.507	0.194	0.192
5-4		CK+	0.156	0.166	0.155
5-5		KFE	0.145	0.132	0.131

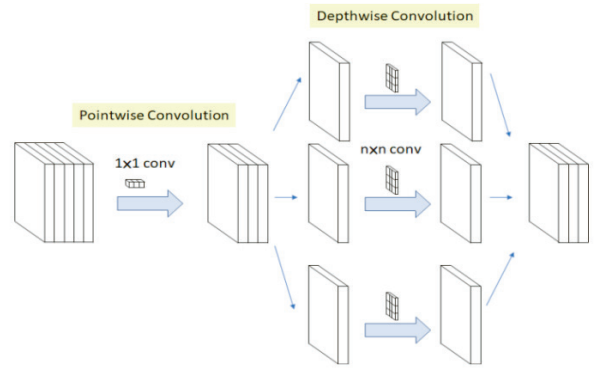


Figure 4. Xception's convolutional architecture

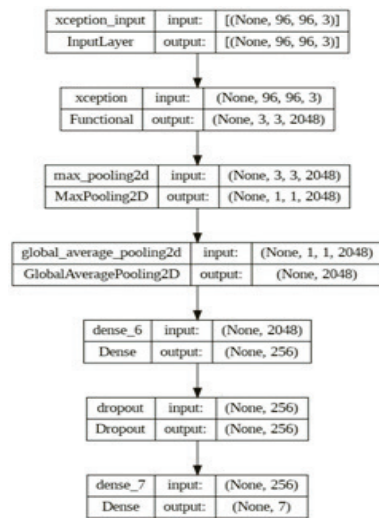


Figure 5. Improved Xception convolutional architecture for learning facial emotion image classification

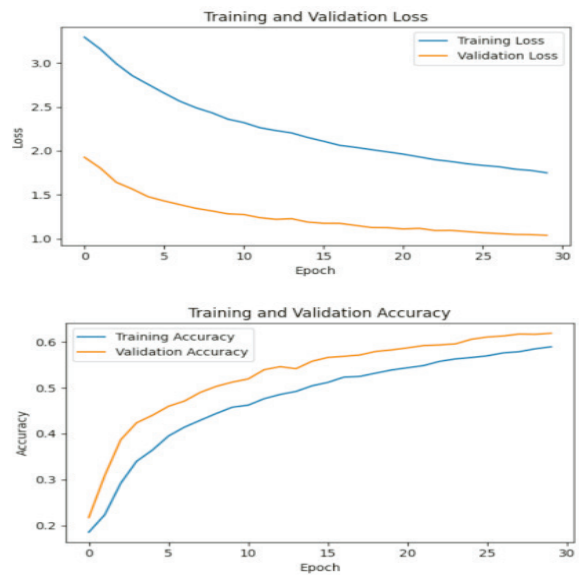


Figure 6. Loss rate and prediction rate of a facial emotion classification model trained using the improved Xception

개선된 Xception의 모델은, 기존의 Xception모델에서 최종 얼굴 감정의 7가지 클래스로 분류하기 위해 풀

링층과 밀집(Dense)층, 드롭아웃을 추가하였다. 기존의 imagenet 1000개의 클래스 분류를 위해 설계된 모델이므로, 최종 7개의 출력을 위해서는 기존의 학습된 설계 모델을 최대한 변형하지 않는 부분에서 전이학습을 시킬 필요가 있기 때문이다.

위 레이어 모델 구조에 따라서 얼굴 감정 이미지를 학습한 CNN 모델의 최종 학습 손실율과 예측율은 [그림 6]과 같이 제시되었다.

에포크가 반복될수록 손실율이 저하되고 예측율이 높아지고 있기 때문에, 위 모델과 데이터세트는 올바르게 학습되어, 얼굴 감정을 올바르게 분류하고 있다고 볼 수 있다. 이 모델에 대한 코드와 세부 결과는 대한 것은 다음 주소에서 확인할 수 있다.<sup>1)</sup>

이를 토대로 최종적으로 얼굴 감정 인식을 통한 집중도 값에 관한 수식은 다음과 같다. 각각의 변수에 대한 정의는 다음과 같다.

face\_class\_weight : 얼굴 감정에 따른 가중치  
 max\_value : 얼굴의 감정 수치  
 yaw : 얼굴의 각도

최종 집중도 값 수식 =

$$\left( \frac{\max\_value \times \text{face\_class\_weight}[\max\_index]}{2.5} \right) \times \left( \begin{cases} 1 & \text{if } |yaw| \leq 0 \\ 0.1 & \text{if } |yaw| \geq 25 \\ \frac{25 - |yaw|}{25} & \text{if } 0 < |yaw| < 25 \end{cases} \right)$$

이와 같은 수식을 통해, 인식된 얼굴은 0 ~ 100 사이의 학습자 집중 상태 수치를 나타내게 된다. 예를 들어 보통 표정에 측면을 바라 볼 경우, 보통의 집중 가중치에서, 측면의 각도를 0.1~1 사이의 비율로 표준화 하여 집중도 값을 나타낸다.

#### 4. 구현 및 평가

최종적으로 온라인 실시간 집중도 분류 ITS을 [그림 7]과 같이 구현하였다. 본 시스템은 실시간으로 학습자의 집중도를 분석하여, 그래픽 인터페이스로 집중도 수치를 표시하여 피드백을 준다. 학습자의 집중도가 25 이하로 낮아지게 되면, 화면 응시도가 낮아졌다는 메시지를 주어 학습자의 학습 지속성을 유발한다.

1) [https://github.com/ljh77/korean\\_FER\\_24/blob/main/ipynb/new\\_2-3\\_X.ipynb](https://github.com/ljh77/korean_FER_24/blob/main/ipynb/new_2-3_X.ipynb)

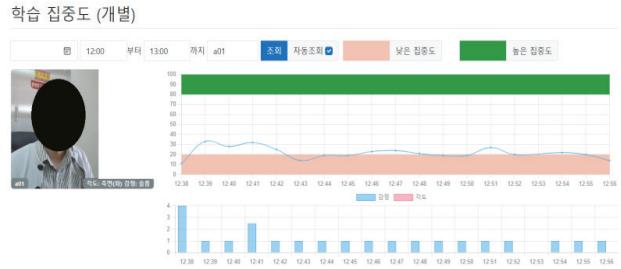


Figure 7. Online real-time concentration individual analysis web screen

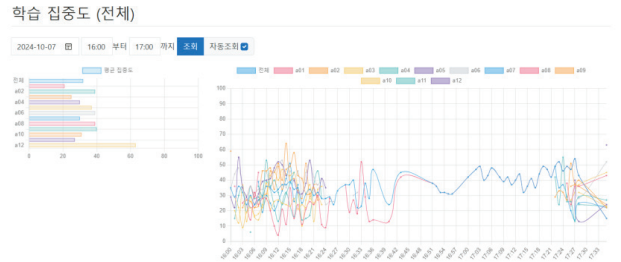


Figure 8. Full analysis of online real-time concentration

교수자는 학습자 전체의 실시간 집중도 상태를 한눈에 파악할 수 있도록 인터페이스를 제공한다.

위 시스템을 토대로 K대학 학생 20명에게 본 시스템의 사용성 평가와 인터뷰를 진행하였다. 사용성 평가 도구는 ‘디지털 학습 지원 플랫폼에 대한 사용성 평가 연구’을 진행한 Lim[19]의 연구를 바탕으로 제작된 시스템 사용성 평가 도구를 활용하였다.

본 시스템은 사용성 평가 도구를 적용하기 위해 시스템에 다음과 같은 사전 개선 작업을 수행하였다.

인터페이스의 구성은 전문 교사 2명, 컴퓨터 교육 전문가 2명, 웹 개발 전문가 2명, 실험 테스트 참여자 3명이 참여하였다. 각 참여자는 직관성, 접근성, 일관성을 기준으로 평가하였으며, 테스트 결과 도출된 주요 피드백인 표 위치 크기 조절, 메뉴 위치 구성, 단계별 접근 시나리오, 반응형 웹 등을 반영하여 개선을 구성하였다.



Figure 9. Web screen with improved interface

학습 지원 요소에는 전문 교사 2명, 컴퓨터 교육 전문가 2명으로 구성하였다. 주요 평가 기준은 수치 데이터의 적합성, 사용자의 학습 상태 지원 여부, 학습자의 데이터 접근성을 검토하였다. 내용으로는 학습자에게 적절한 학습 지원을 제공하는지 검토하였다. 이후 개선 사항인 집중 상태에 따른 자동 메시지 출력, 교수자 전용 페이지, 채팅 기능 등의 개선을 진행하였다.



Figure 10. Real-time learning support feedback screen

시스템의 안정성에는 웹 개발 전문가 2명, 시스템 테스트 3명으로 구성하였다. 주요 테스트에는 부하 테스트, 장애 복구 시뮬레이션, 응답 시간 분석을 진행하였다. 시스템 부하 테스트에서는 동시에 최대 응답 수신율을 확인하였다. 이후 웹 서버에는 다운시 자동 페일백 추가, API 시스템의 분리, 서버의 실시간 로그 모니터링 등의 기능을 추가하였다.

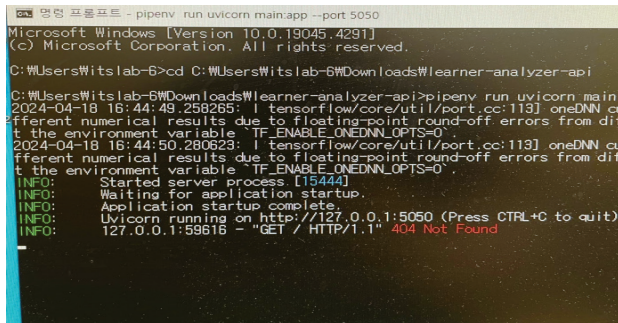


Figure 11. Operation server real-time monitoring log

이를 토대로 기능 사용성 평가 영역을 인터페이스 구성, 학습 지원 요소, 시스템 안정성의 세 가지 요소에 대해 묻는 질문으로 <표 4>과 같이 구성하였다.

Table 4. Usability Evaluation Elements of a Learner Focus State System

Context	Evaluation details
Interface	- Is it easy for users to use? - Does it work in accordance with users' common sense?
Learning Support	- Is it possible to understand the learner's status? - Is learning motivation improved?
System Stability	- Does the system operate stably? - Do the functions work properly?

연구에 참여한 학습자는 실제 강의를 듣고 학습을 진행하는 K대학교의 20명을 대상으로 진행하였다. 참여자는 실험 당시 만 20세에서 25세의 나이의 대학 1~4학년이었다. 참여자의 특징은 <표 5>와 같다.

Table 5. Participants' characteristics

Context	Details
Sex	male : 9 / female : 11
Education	University students

참여자가 응답한 크론바흐 알파 값은 0.834로 신뢰도가 있다고 볼 수 있다. 학습자 집중도 분류 ITS에 대한 사용성 평가 영역에 대한 응답 결과는 <표 6>와 같다.

Table 6. Evaluation scores of the learner focus state system

Context	Average	Standard Deviation
Interface	3.58	.874
Learning Support	3.82	1.17
System Stability	2.61	1.13

인터페이스 영역은 평균 3.58과 표준 편차 .874를 보였다. 세부 항목은 '필요한 기능을 찾기 쉽다(3.8)', '시스템의 정보를 쉽게 파악할 수 있다(3.7)', '직관적이고 사용하기 쉽다(3.65)', '시스템은 반응 속도는 빠르고 원활하다(3.1)'로 나타났다. 웹과 기능들은 직관적이고 편의성 있게 동작한 것으로 볼 수 있다. 그러나 시스템의 반응 속도는 원활하지 않아 개선이 필요한 것으로 보인다.

학습 지원 영역은 평균 3.82 표준 편차 1.17로, 대체로 긍정적인 반응을 보였으나, 개인에 따라 편차가 있는 것으로 나타났다. 세부 항목은 '시스템을 사용하는 동안 학습 동기가 향상되었다(3.9)', '시스템은 학습에 도움이 된다(3.7)'로 학습 동기와 학습 지원에 긍정적인 관점을 제시하고 있으나, 편차가 상대적으로 크기 때문에 개인에 따라 학습 지원에 대한 의견이 차이가 있는 것으로 볼 수 있다. 인터뷰 내용은 다음과 같다.

-교사의 입장에서 생각하면 학생들을 동시에 관리가 불가능할때 매우 유용할 것 같고 학생의 입장에서는 자신을 성찰하고 돌아보면 자신을 더 나은 방향으로 개선 가능 할 수 있을 것 같습니다. (ㄴ학생)

-학습자의 집중도를 손쉽게 점검할 수 있다는 점에서 긍정적이며, 앞으로 이러한 것을 실제 교육에 적용하여 사용하면 보다 통제 가능한 환경이 조성될 수 있다고 본다. (ㄷ학생)  
-자신의 모습을 수업 후에 확인할 수 있어 다시 자신의 학습 모습을 보며 성찰할 수 있다는 점이 만족스러웠다. (ㄹ학생)  
-정확도가 낮은 듯 하다. (ㄹ학생)

-화면을 보는 시간으로 집중력을 계산하는 것은 필기를 열심히 하거나 하는 학생들에게 수업에 대한 집중도가 낮은 것처럼 되어 억울할 수도 있겠다는 생각이 들었다. (ㄴ학생)

시스템 안정성은 2.61과 표준편차 1.13으로 낮은 것을 나타냈다. ‘시스템이 안정성 있게 동작한다(2.75)’, ‘시스템이 오류가 발생하지 않거나, 적다(2.35)’로 낮게 나타났다. 이는 하나의 서버에서 20명의 영상데이터를 일시에 처리할 뿐만 아니라, 다시 학습자에게 영상데이터를 재송신하여, 이에 따른 서버 부하가  $n^2$ 으로 기하급수적으로 늘어나게 된다. 이에 따라 서버가 일시적으로 멈추고 접근을 차단하는 현상을 확인하였다. 이에 따라 일시적으로 학습자는 시스템의 이용이 어려웠으며, 추가적인 기능 개선 및 설계가 필요해 보인다. 학생들의 인터뷰 내용은 다음과 같다.

- 접속이 많아지자 접속할 수가 없었다. (♂학생)
- 서버가 불안정 하였다. (♀학생)

## 5. 결론

본 연구에서는 학습자 집중 상태 분류를 위한 ITS를 설계하고 구현하여, 그 시스템의 사용성을 평가하였다. 연구 결과, 본 시스템은 학습자와 교수자 모두에게 실시간 집중 상태를 확인하고 피드백을 제공함으로써 학습 지원 및 강의 개선에 기여할 가능성을 확인하였다. 학습자는 자신의 집중 상태를 실시간으로 파악하여 학습 동기를 유지하거나 향상시킬 수 있었으며, 교수자는 강의 중 집중도 변화에 대한 데이터를 활용하여 강의 내용을 개선할 수 있었다. 그러나 다음과 같은 한계점을 갖는다.

첫째로, 다의적 행동에 대해서는 판별이 어렵다는 한계를 갖는다. 학습자가 필기를 위해서 고개를 숙이는 행동과 수업 산만의 표시로 고개를 숙이는 행동. 집중하는 무표정의 학습자의 정면 얼굴과, 무표정으로 정면 응시는 하고 있지만 다른 생각을 하고 있는 경우 등. 이러한 특정 행동은 의도에 따라 신중한 판별이 요구된다. 따라서 행동의 맥락적 의도를 포함한 추가적인 연구 분류 기준을 설정하고, 이를 반영한 연구 설계가 필요하다.

때문에 학습자의 다양한 행동 및 표정 데이터를 추가적으로 수집하고 이를 기반으로 보다 정밀한 집중 상태 분류 기준을 마련할 필요가 있다. 학습 중 다양한 감정 상태와 행동을 포함한 맥락적 데이터를 분석할 필요가 있겠다. 예를 들어 카메라를 통해 실시간으로 표정을 과 행동을 기록하며, 필요시 시선 추적 데이터나 심박수 등 생리학적 데이터를 병합할 수도 있다. 이러한 데이터를 기반으로 학습자가 집중하고 있는 상황과 산만한 상황을 구분하여 라벨링 작업을 수행하면 정밀한 집중 분류의 설정이 가능할 것이다.

둘째로, 온라인 실시간 시스템의 원활한 기능 구현이다. 실제적으로 20명의 동시적인 영상처리는 연산과 네트워크 통신에 부하를 가져와 원활한 시스템 구동이 어려운 점이 있었다. 학습자나 교수자가 원활히 시스템을 활용하도록 전문적인 네트워크 시스템 구현의 개선이나 최대 동시 처리 인원 에 대한 제한이 필요 된다.

향후 연구는 다음과 같다. 첫째로, ITS의 피드백 기능을 학습 지원 외에도 학습자의 학습 전략 수립이나 학습 개선

방향을 제안하는 방식으로 확장할 필요가 있다. 예를 들어, 학습자가 일정 시간 동안 집중 상태를 유지하지 못할 경우, 집중을 유도하는 구체적인 학습 활동을 추천하는 기능을 추가적으로 개발할 수 있다.

둘째로, 본 시스템을 대규모 온라인 강의, 소규모 토론 수업 등에서 테스트하여 실질적인 활용 가능성을 평가하고, 이를 기반으로 실용적인 서비스로 확장하는 연구가 필요할 것이다. 특히, 시스템 성능과 사용자 경험을 개선하여 ITS가 교육 현장에서 널리 활용될 수 있도록 해야 한다.

이와 같이 실시간 학습자 집중 상태 분류 ITS에 대해 설계하고 그 사용성을 평가해 보았다. 이러한 적시적인 피드백 시스템을 적용하여, 추후 교육 학습의 발전과 교육 보조 시스템에 대한 초석이 되길 기대한다.

## 참고문헌

- [1] Sleeman, D., & Brown, J. S. (1982). Intelligent tutoring system. Academic Press. [https://doi.org/10.1016/0004-3702\(85\)90033-5](https://doi.org/10.1016/0004-3702(85)90033-5)
- [2] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324. <https://doi.org/10.1109/5.726791>
- [3] Winn, W. (1989). Toward a rationale and theoretical basis for educational technology. *ETR&D* 37, 35-46. <https://doi.org/10.1007/BF02299044>
- [4] Burns, H., & Parlett, J.(1991). The evolution of intelligent tutoring systems: Evolutions in design. In *Intelligent Tutoring Systems*. Hillsdale, NJ. Lawrence Erlbaum Associates.
- [5] Butz, C. J., Hua, S., & Maguire, R. B.(2006). A web-based bayesian intelligent tutoring system for computer programming. *Web Intelligence and Agent Systems: An International Journal*, 4(1), 77-97.
- [6] Cha, S., & Kim, W. (2015). Analyze the learner's concentration using detection of facial feature points. *Adv. Sci. Technol. Lett*, 92, 72-76. <http://dx.doi.org/10.14257/astl.2015.92.15>
- [7] Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *Quarterly Journal of Experimental Psychology (Hove)*, 62(8), 1457-1506. <https://doi.org/10.1080/17470210902816461>
- [8] Doherty, K., & Doherty, G. (2018). Engagement in HCI: Conception, theory, and measurement. *ACM Computing Surveys*, 51(5), Article 99. <https://doi.org/10.1145/3234149>
- [9] Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data. *Marketing Science*, 23(4), 579-595. <https://doi.org/10.1287/mksc.1040.0073>
- [10] Sharma, P., Joshi, S., Gautam, S., Maharjan, S., Khanal, S. R., Reis, M. C., Barroso, J., & Filipe, V. M. de J. (2023). Student engagement detection using emotion analysis, eye tracking, and head movement with machine



- learning (Version 5). *arXiv*. <https://doi.org/10.48550/arXiv.1909.12913>
- [11] Sharma, P., Esengönül, M., Khanal, S. R., Khanal, T. T., Filipe, V., & Reis, M. (2019). Student concentration evaluation index in an e-learning context using facial emotion analysis. In *Technology and Innovation in Learning, Teaching and Education* (pp. 529-538). Springer. [https://doi.org/10.1007/978-3-030-20954-4\\_40](https://doi.org/10.1007/978-3-030-20954-4_40)
- [12] Bhardwaj, P., Gupta, P. K., Panwar, H., Siddiqui, M. K., Morales-Menendez, R., & Bhaik, A. (2021). Application of deep learning on student engagement in e-learning environments. *Computers & Electrical Engineering*, 93, 107-277. <https://doi.org/10.1016/j.compeleceng.2021.107277>
- [13] Sharma, P., Esengönül, M., Khanal, S., Khanal, T., Filipe, V., & Reis, M. (2019). Student Concentration Evaluation Index in an E-learning Context Using Facial Emotion Analysis. In *Technology and Innovation in Learning, Teaching, and Education* (pp. 529-538). Springer. [https://doi.org/10.1007/978-3-030-20954-4\\_40](https://doi.org/10.1007/978-3-030-20954-4_40)
- [14] Pabba, C., & Kumar, P. (2022). An intelligent system for monitoring students' engagement in large classroom teaching through facial expression recognition. *Expert Systems*, 39(1), e12839. <https://doi.org/10.1111/exsy.12839>
- [15] Lee, J. H., & Song, K. S. (2024). Comparison and Analysis of CNN Models to Improve a Facial Emotion Classification Accuracy for Koreans and East Asians. *International Journal on Advanced Science, Engineering & Information Technology*, 14(3), 811-817. <https://doi.org/10.18517/ijaseit.14.3.18078>
- [16] M. Saleem Abdullah, S., & Abdulazeez, A. M. . (2021). Facial Expression Recognition Based on Deep Learning Convolution Neural Network: A Review. *Journal of Soft Computing and Data Mining*, 2(1), 53-65. <https://publisher.uthm.edu.my/ojs/index.php/jscdm/article/view/7906>
- [17] Korea Institute for Artificial Intelligence & Society. *AIhub*. Retrieved October 2024, from <https://aihub.or.kr/>
- [18] Lyons, M., Akamatsu, S., Kamachi, M., & Gyoba, J. (1998). Coding facial expressions with Gabor wavelets. In *Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition* (pp. 200-205). Nara, Japan. <https://doi.org/10.1109/AFGR.1998.670949>
- [19] Lim, C., Song, H.-D., Lee, L., & Lee, Y. (2009). Development and implementation of a digital textbook platform usability assessment instrument. *Journal of Educational Technology*, 25, 127-157. <https://doi.org/10.17232/KSET.25.4.127>
- [20] Wang, M., & Deng, W. (2019). Mitigate bias in face recognition using skewness-aware reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 9322-9331). <https://doi.org/10.48550/arXiv.1911.10692>
- [21] Morales, A., Fierrez, J., Vera-Rodriguez, R., & Tolosana, R. (2020). SensitiveNets: Learning agnostic representations with application to face images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. <https://doi.org/10.48550/arXiv.1902.00334>
- [22] Darabant, A. S., Borza, D., & Danescu, R. (2021). Recognizing human races through machine learning—A multi-network, multi-features study. *Mathematics*, 9(2), 195. <https://doi.org/10.3390/math9020195>



이준형

- 2013년 순천대학교 컴퓨터교육학과 (학사)
- 2018년 순천대학교 컴퓨터교육전공 (석사)
- 2023년 한국교원대학교 컴퓨터교육 (박사수료)

✚ 관심분야 : 컴퓨터 교육, AI 교육, AI, 컴퓨터 과학  
✉ jhlee77@knue.ac.kr



송기상

- 1983년 아주대학교 학사
- 1985년 한국과학기술원 석사
- 1994년 U. of Washington, Ph. D.
- 1995~현재 한국교원대학교 교수
- 2007~현재 UNESCO, KOICA ICT in Education 국제컨설턴트
- 2021~ KOICA 캄보디아 중학교 ICT 역량강화사업 PM

✚ 관심분야 : 에듀테크, 인공지능 기반 학습체제, ICT 기반 교육을 통한 국제교육협력  
✉ kssong@knue.ac.kr, kssong1k4@gmail.com