



Research on a statistics education program utilizing deep learning predictions in high school mathematics

Hyeseong Jin¹, Boeuk Suh^{2*}

¹Teacher, Daejeon Dunsan Girls' High School

²Professor, Chungnam National University

ABSTRACT

The education sector is undergoing significant changes due to the Fourth Industrial Revolution and the advancement of artificial intelligence. Particularly, the importance of education based on artificial intelligence is being emphasized. Accordingly, the purpose of this study is to develop a statistics education program using deep learning prediction in high school mathematics and to examine the impact of such statistically problem-solving-centered statistics education programs on high school students' statistical literacy and computational thinking. To achieve this goal, a statistics education program using deep learning prediction applicable to high school mathematics was developed. The analysis revealed that students' understanding of context improved through experiencing how data was generated and collected. Additionally, they enhanced their comprehension of data variability while exploring and analyzing various datasets. Moreover, they demonstrated the ability to critically analyze data during the process of validating its reliability. In order to analyze the impact of the statistics education program on high school students' computational thinking, a paired sample t-test was conducted, confirming a statistically significant difference in computational thinking between before and after classes ($t=-11.657$, $p<0.001$).

Keywords High school mathematics education, AI convergence education, Statistics education, Deep learning

서 론

4차 산업혁명으로 사회의 다양한 분야에서 다양한 변화가 나타나고 있다. 특히 4차 산업혁명의 여러 가지 기술 중에 가장 주목받고 있는 분야는 인공지능(Artificial Intelligence)이다. 딥러닝(deep learning)의 발전으로 인공지능이 인간의 지적 기능을 모방하고 수행하는 데까지 발전하였다. 2022 개정 교육과정 총론에서도 빠르게 변화하는 디지털 전환에 대응하기 위해 AI 교육환경에 맞는 교수·학습 및 평가체제 구축을 교육과정 개정의 방향으로 설정하였다(Ministry of Education, 2022a). 인공지능으로 인한 산업의 변화로 인공지능을 기반으로 하는 교육의 중요성이 강조되고 있다(Kwon et al., 2023; Lee, 2023; Oh, 2023). 인공지능 활용 교육의 핵심은 문제해결이며 비정형데이터를 수집하고 이를 분석하여 예측하는 활동을 통해서 학

Received March 4, 2024; Revised March 29, 2024; Accepted May 8, 2024

*Corresponding author Boeuk Suh

E-mail eukeuk@cnu.ac.kr

2000 Mathematics Subject Classification 97U70

본 논문은 진혜성의 충남대학교 석사학위논문을 바탕으로 작성되었음.



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

습자는 실제적 문제를 탐구할 수 있다(Han et al., 2021). 딥러닝의 비선형성(nonlinear) 모델은 실생활의 복잡한 비선형적 관계를 학습할 수 있으며, 학습된 모델은 다양한 문제상황에서 쉽게 적용할 수 있다. 따라서 인공지능 활용 교육에서 딥러닝은 실생활과 관련된 문제를 표현하고 이를 해결하는 유용한 도구로 사용될 수 있다.

2022 개정 수학과 교육과정에서는 지식·이해, 과정·기능, 가치·태도를 통합적으로 학습하여 수학 교과 역량을 함양하는 것을 총괄 목표로 설정하였다. 과정·기능 범주에서는 학생의 주도성 강화를 위해 일상생활을 포함한 다양한 맥락의 문제를 해결하는 데 도움이 되는 관점으로 공학 도구를 활용한 '통계적 문제해결 과정'을 강조하였다. 자료를 이용하여 통계적 문제해결 과정을 실천하고 생활 속의 가능성을 탐구하는 것은 미래를 예측하고 합리적인 의사결정을 하는 데 기반이 된다(Ministry of Education, 2022b). 또한, 2022 개정 정보과 교육과정에서는 컴퓨팅 사고력을 교과의 역량으로 설정하였다. 컴퓨터를 활용한 문제해결을 전제로 문제를 발견, 분석하여 실생활과 다양한 학문 분야의 문제를 해결하기 위한 새로운 방법론을 제시할 수 있는 능력을 컴퓨팅 사고력으로 정의하였으며 컴퓨팅 사고력 함양을 정보 교과의 목표로 제시하였다(Ministry of Education, 2022c). 그러나, 2022 개정 수학과 교육과정에서 통계적 탐구 활동을 통해 실생활 문제를 해결할 수 있도록 안내하고, 공학 도구의 활용은 교육과정에서 꾸준히 강조되고 있음에도 불구하고, 교과서에는 여전히 의사소통, 비판적 태도, 공학 도구의 활용이 미비하게 다루어지고 있다(Lee & Rim, 2021). 2022 개정 수학과 교육과정은 통계적 문제해결 과정을 중요한 내용 요소로 포함하고 있으며 통계적 소양을 개발하는 것을 통계교육의 목적으로 명시하고 있다. 또한, 2022 개정 정보과 교육과정은 정보 교과의 목적으로 컴퓨팅 사고력을 강조하고 있다. 따라서, 통계적 문제해결 과정을 통해 통계적 소양과 컴퓨팅 사고력을 함양하는 통계교육 프로그램이 필요하다.

본 연구의 목적은 고등학교 수학에서 딥러닝 예측을 이용한 통계교육 프로그램을 개발하고 통계적 문제해결 과정 중심의 통계교육 프로그램이 학생의 통계적 소양 및 컴퓨팅 사고력에 어떤 영향을 미치는지를 분석하는 것이다. 이러한 본 연구의 목적을 달성하기 위하여 첫째, 고등학교 수학에서 딥러닝 예측을 이용한 통계교육 프로그램은 어떻게 개발될 수 있는지, 둘째, 딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생들의 통계적 소양 및 컴퓨팅 사고력에 어떤 영향을 미치는지 살펴보는 것으로 설정하였다.

이론적 배경

1. 통계적 소양과 통계적 문제해결 과정

통계적 소양은 정보화 사회에서 사람들에게 기대되는 핵심 능력이며 학교 통계교육의 목적이 되고 성인들의 수리 능력 및 문해력의 필수 요소로 자주 언급된다. 그러나 그 의미와 구성 요소에 대해 명확한 주목을 받지 못하고 있다(Gal, 2002). 통계적 소양이란 다양한 맥락에서 마주치게 될 수 있는 통계적 정보 및 데이터와 관련 주장 또는 확률 현상을 해석하고 비판적으로 통계적 정보를 평가하는 능력과 이러한 통계적 정보에 대한 반응을 논의하거나 의사소통하는 능력을 의미한다(Gal, 2002). Watson과 Callingham (2003)은 통계적 소양이 계층적 구조를 갖는다는 사실을 증명하였다. 통계적 소양의 계층적 구조는 총 6가지 수준으로 구분된다. 1수준(부적절한 수준), 2수준(비형식적인 수준), 3수준(비일관적인 수준), 4수준(일관적이거나 무비판적인 수준), 5수준(비판적인 수준), 6수준(비판적이고 수학적인 수준)으로 각 단계가 학습자의 이해와 능력을 나타내며, 교육 과정 설계와 학습자 평가에 유용한 기준을 제공한다. 이에 본 연구에서는 다양한 맥락에 마주치게 될 수 있는 통계적 정보 및 현상을 해석하고 비판적으로 평가하여 현실의 문제를 해결하는 능력을 통계적 소양으로 본다. 또한, 선행 연구 분석 결과를 토대로 인공지능을 활용한 통계교육 프로그램이 학생들의 통계적 소양에 미치는 영향을 살펴보고자 한다.

통계적 문제해결 과정은 체계적이고 절차적인 통계적 방법의 전략으로 구성되는데, 연구자들은 4단계(Franklin et al., 2007; Graham, 2006) 또는 5단계(Mackay & Oldford, 1994; Wild & Pfannkuch, 1999)로 이루어진 통계적 조사 과정을 제시한다. 이들 통계적 조사 과정의 각 단계는 순환적 구조를 가지며 상호 영향을 미친다(Ko & Park, 2017). GAISE 보고서(Franklin et al., 2007)는 통계적 문제해결 과정을 Table 1과 같이 문제 설정, 자료 수집, 자료 분석, 결과 해석의 네 가지 구성 요소로 이루어진 조사 과정으로 정의하였다.

2. 딥러닝을 활용한 인공지능교육

인공지능은 인간의 지능적인 행위를 흉내 낼 수 있게 하는 소프트웨어 시스템을 말한다. Hong과 Choi (2020)의 '학교 교육에서 인공지능의 개념 및 활용'에 따르면 교육에서 인공지능 개념에 대한 인식 조사 결과 '초·중등 교육에서의 인공지능교육'

Table 1. Statistical problem solving is an investigative process (Franklin et al., 2007, p.11)

Process	Characteristics
Formulate questions	-Clarify the problem at hand -Formulate one (or more) questions that can be answered with data
Collect data	-Design a plan to collect appropriate data -Employ the plan to collect the data
Analyze data	-Select appropriate graphical and numerical methods -Use these methods to analyze the data
Interpret results	-Interpret the analysis -Relate the interpretation to the original question

은 ‘인공지능에 대한 교육’과 ‘인공지능 활용 교육’을 포괄하는 개념으로 정의한다. 인공지능에 대한 교육은 ‘내용으로서의 AI’를 강조하며 인공지능 활용 교육은 ‘도구로서의 AI’를 강조하는 개념으로 정의할 수 있다(Kim & Choi, 2022; Park et al., 2022; Sim et al., 2023). 또한, Han 외 (2021)는 인공지능 교육을 크게 네 개의 영역, 즉 인공지능 기반교육, 인공지능 이해교육, 인공지능 활용 교육, 인공지능 융합교육으로 구분하였다.

인공지능교육에 관한 선행 연구(Hong & Choi, 2020; Lee et al., 2023a; Ryu & Han, 2022)는 주로 인공지능을 이해하는 데 목적을 두고 있으며, 초등학생과 중학생을 대상으로 하는 교육으로 기본적인 수준의 인공지능 기술을 도구로 활용하는 데 중점을 두고 있다. Ryu와 Han (2022)은 인간과 인공지능이 협력하여 창의적으로 문제를 해결할 수 있는 디지털 기반의 창의협력지능 교수학습 모형을 설계하고, 이에 따른 모형의 효과성을 검증하는 연구를 수행했다. 이 모형을 기반으로 초등학교와 중학교에서 각각 11종과 7종의 총 18종의 교수학습 프로그램을 개발하였으며, 시범 수업 적용 결과 만족도, 수업 내용의 흥미도, 협력적 문제해결태도, 협력적 창의 태도, 협력적 도구 활용 태도에서 유의미한 결과가 나타났다(Ryu & Han, 2022). 초등학생과 중학생을 대상으로 한 이 연구 결과를 일반화하기 위해 고등학생들을 대상으로 한 후속 연구가 필요하다. Lee 외 (2023)는 효과적인 빅데이터 통계 프로젝트 기반 과제를 설계하기 위한 일련의 가이드라인을 제안하였다. 도출된 가이드라인은 다음과 같다. (1) 지식과 기술을 국가 학교 수학 교육과정에 맞추고, (2) 전처리된 대규모 데이터 셋을 사용하며, (3) 데이터 과학자의 문제 해결 방법을 사용하고, (4) 의사 결정을 장려하며, (5) 공학도구를 활용하고, (6) 협업 학습을 촉진한다. 이 기준에 따라 인공지능 수학 교과서에 실린 최적화 단원 과제를 평가한 결과 가이드라인 2에 해당하는 요소를 프로젝트 과제에 통합하지 못하고 있는 것으로 나타났다(Lee et al., 2023a; Lee et al., 2023c). 이러한 가이드라인을 바탕으로 인공지능교육이 이루어진다면 빅데이터를 기반한 새로운 통계 교육의 기회가 제공될 것으로 기대된다. 따라서 빅데이터를 사용하는 통계 교육에서 공학적 도구의 효과적인 활용에 대한 다양한 후속 연구가 필요하다.

인공지능은 사고나 학습 등 인간이 가진 지적 능력을 컴퓨터로 구현하는 기술이며 컴퓨터가 스스로 학습하여 새로운 정보를 얻어내거나 예측하는 기술인 기계학습(machine learning)을 포함한 상위 개념이다(Lee et al., 2023b). 기계학습의 한 종류인 딥러닝이란 인간의 뉴런을 모방한 인공 신경망을 이용하여 기계학습을 구현하는 기술이다. 기계학습의 방법으로는 지도학습, 비지도학습, 강화학습이 있다. 지도학습이란 문제와 정답을 모두 알려주고 학습시키는 방법으로 예측값으로 연속적인 값을 출력하는 회귀(regression)와 이산적인 값을 출력하는 분류(classification)로 나눌 수 있다(Park, 2020). 비지도학습은 정답을 가르쳐주지 않고 학습을 시키는 방법으로 연관 규칙, 군집에 활용된다. 마지막으로 강화학습은 보상을 통해 상을 최대화하고 벌은 최소화하는 방향으로 행위를 강화하는 학습 방법이다. 인공 신경망(ANN, artificial neural network)이란 인간이 뇌를 통해 문제를 처리하는 방법처럼 문제를 해결하기 위해 컴퓨터에서 인위적으로 구성된 구조이다. 인공 신경망은 다수의 퍼셉트론(perceptron)으로 구성된 구조로, 입력값에 가중치를 곱하여 출력값을 결정한다. 가중치는 입력값의 중요도를 조절하며, 학습 과정에서 가중치를 조정함으로써 모델은 점점 정확한 결과를 도출하도록 발전한다. 퍼셉트론은 하나의 뉴런을 본떠서 만들어진 계산 방식이다. 신경망 모델은 뉴런이 모여 신경계를 이루는 것처럼 퍼셉트론을 연결하여 형성한다. 신경망 모델은 퍼셉트론과 다르게 입력층(input layer)과 출력층(output layer) 외에도 은닉층(hidden layer)을 추가로 사용한다. 은닉층이 생성됨으로써 신경망 모델은 여러 층이 생기게 되고 여러 개의 은닉층으로 이루어진 인공 신경망이 바로 심층 신경망(deep neural network)이라 한다(Park, 2020). 퍼셉트론은 출력값으로 선형함수(linear function)를 가지지만, 심층 신경망은 비선형함수(non-linear function)를 적용함으로써 비선형적 문제를 효과적으로 다룰 수 있다. 딥러닝은 심층 신경망을 이용하여 다양한 맥락에서 발생하는 패턴을 분석하고 기계학습을 통해 문제에 맞는 최적의 결과값을 예측한다. 이러한 인공지능의 기술

을 통계교육에 활용한다면 학생들이 일상생활에서 마주하는 현실의 문제를 효과적으로 다루며 이를 분석하고 예측함으로써 합리적인 의사결정을 하는 데 필요한 기반을 갖출 수 있을 것으로 기대된다. 이에 본 연구에서는 전문가들이 현실의 문제를 해결하기 위해 인공지능을 이용하는 것처럼 고등학생을 대상으로 한 인공지능교육에서 인공지능이 어떻게 활용될 수 있는지 구체적으로 살펴보고자 한다.

3. 컴퓨팅 사고력

컴퓨팅 사고력(computational thinking)은 Wing 교수에 의해 2006년부터 본격적으로 언급되기 시작한 개념이다(Park, 2023). Wing (2006)은 컴퓨팅 사고력은 컴퓨터 과학자뿐만 아니라 모든 사람에게 필수적인 기술로 읽기, 쓰기, 산술처럼 기본능력으로 설명하였다(Wing, 2006). 미국의 CSTA (Computer Science Teachers Association)와 ISTE (International Society for Technology in Education)는 K-12 교육을 위한 컴퓨팅 사고력의 조작적 정의 요소를 제안하였고 이를 정리하면 Table 2와 같다. 2022 개정 정보과 교육과정은 인공지능으로 정의되는 사회에서 데이터와 정보로 인한 디지털 세상의 변화를 인식하고, 정보의 사회적 가치를 탐구하며, 정보를 처리하는 다양한 원리와 기술에 기반한 사고력으로 컴퓨팅 사고력을 정의하였다(Ministry of Education, 2022c). 이에 본 연구에서는 컴퓨터 과학의 기본 개념을 이해하고, 문제 해결에 적용하는 능력을 컴퓨팅 사고력으로 정의하고 인공지능을 활용한 통계교육 프로그램이 학생들의 컴퓨팅 사고력에 미치는 영향을 살펴보고자 한다.

Table 2. CT vocabulary and progression chart (CSTA & ISTE, 2011, pp. 8-9)

	Definition
Data collection	The process of gathering appropriate information
Data analysis	Making sense of data, finding patterns, and drawing conclusions
Data representation	Depicting and organizing data in appropriate graphs, charts, words, or images
Problem decomposition	Breaking down tasks into smaller, manageable parts
Abstraction	Reducing complexity to define main idea
Algorithms	Series of ordered steps taken to solve a problem or achieve some end
Automation	Having computers or machines do repetitive or tedious tasks
Simulation	Representation or model of a process. Simulation also involves running experiments
Parallelization	Organize resources to simultaneously carry out tasks to reach a common goal

연구방법

1. 연구대상 및 절차

본 연구는 OO광역시에 있는 일반계 고등학교에 재학 중인 2학년 수학동아리 학생 12명을 대상으로 연구를 진행하였다. 수학동아리는 수학 교구를 만들어 교내 수학 체험전을 운영하고 수학과 관련된 다양한 활동에 참여하며 수학적 역량을 기르는 것을 목표로 한다. 수학동아리 학생들은 대부분 수학에 흥미가 높으며 실생활과 연결지어 수학의 실제적 탐구에 관심이 많다. 연구 수업에 참여한 학생들의 수학 학업 성취도는 중상위권에 속해 수학의 기본적인 개념과 수학적 원리를 학습하고 이해하는 데 어려움을 느끼지 않았다. 그러나 탐구형 과제를 설정하고 해결하는 데 어려움을 보였으며, 수학적 개념과 원리를 제한적으로 활용하였다. 사전 설문 결과에 따르면 대다수 학생들은 문제 해결을 위해 적합한 자료를 수집하고 패턴을 분석하여 결과를 도출할 수 있다고 응답했지만, 프로그래밍 언어를 이용하여 문제를 해결한 경험이 없으며 프로그램을 구현하는 데 어려움을 느낀다고 응답하였다. 연구 수업은 동아리 시간을 활용하여 진행되었으며, 단일집단 연구로 각 50분간 총 8차시의 통계교육 프로그램 활동이 이루어졌다. 본 연구의 방법과 절차를 정리하면 Figure 1과 같다.

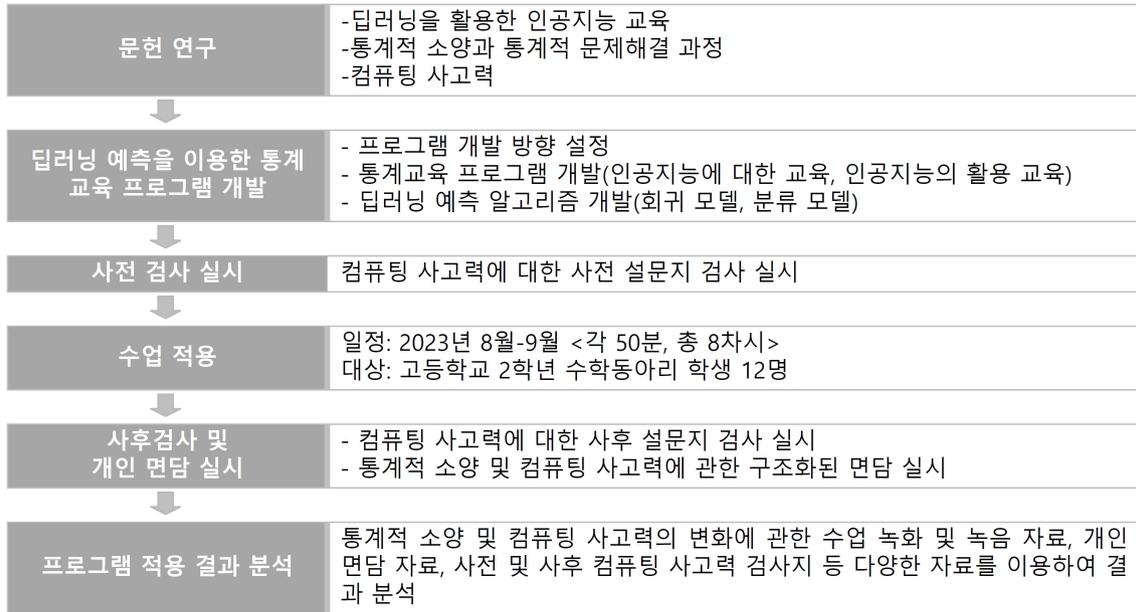


Figure 1. Research methods and procedures.

2. 분석 도구 및 분석 방법

(1) 통계적 소양 분석틀

본 연구에서는 선행연구(Franklin et al., 2007; Gal, 2002; Ju et al., 2018; Ko & Park, 2017; Ryu & Han, 2022; Watson & Callingham, 2003)를 참고하여 통계적 문제해결 과정에 따른 통계적 소양 분석틀을 개발하였다. 통계적 문제해결 과정은 문제 설정(1), 자료 수집(2), 자료 분석(3), 결과 해석(4)으로 보았다. 그리고 Gal (2002)의 통계적 소양 요소인 지식 요소(기본 소양, 통계·수학적 지식, 맥락적 지식, 비판적 질문)와 성향 요소(신념과 태도, 비판적 자세) 및 GAISE 보고서(Franklin et al., 2007)에서 강조한 변이성을 통계적 소양 요소로 정의하였다. 통계적 소양의 수준은 Watson과 Callingham (2003)의 6가지 수준을 하 수준(1, 2수준), 중 수준(3, 4수준), 상 수준(5, 6수준)의 세 가지 수준으로 재구성하여 Table 3과 같이 통계적 문제해결 과정에 따른 통계적 소양 분석틀을 개발하였다. 문제 설정(1)은 Ko와 Park (2017)이 제시한 분석틀을 참고하였으며 자료 수집(2), 자료 분석(3), 결과 해석(4)의 목표 달성 확인은 Ju 외 (2018)가 제시한 통계적 소양 분석틀을 참고하였으며 결과 해석(4)의 결과 공유 및 피드백, 상호 학습 확인, 협력지능 업데이트는 Watson과 Callingham (2003)이 제시한 통계적 소양 계층 구조를 참고하였다(Jin, 2024).

(2) 컴퓨팅 사고력 분석 도구

Kim과 Kim (2018)이 개발한 문제해결 기반의 컴퓨팅 사고력 측정도구(PSB-CTQ; Problem Solving based Computational Thinking Questionnaire)를 이용하여 통계교육 프로그램에 참여한 수학동아리 학생 12명의 문제해결 활동에 대한 컴퓨팅 사고력을 측정하였다. 총 31개 문항으로 문제인식 11문항, 해결방안탐색 7문항, 프로그래밍 13문항으로 구성되며, Likert식 4단계 척도(매우 그렇지 않다=1점-매우 그렇다=4점)로 응답을 구성하였다. 각 영역별 계수(Cronbach's α)는 문제인식 0.948, 해결방안탐색 0.939, 프로그래밍 0.957이고, 전체 문항 31개에 대한 신뢰도 Cronbach α 는 0.97로 나타났다. 연구 도구의 구성 및 설문 문항 예시는 다음 Table 4와 같다(Jin, 2024).

(3) 분석방법

연구 수업에 참여한 모든 학생의 동의를 얻어 연구 수업의 모든 차시에서 영상 녹화와 모둠별 음성 녹음을 진행하였다. 또한, 연구 수업 이후에는 구조화된 면담을 하고 이를 녹음하여 전사하였다. PSB-CTQ 설문지를 활용하여 연구수업 전과 후에 설문하고 설문지 자료를 수집하였으며, 차시별 학생들의 활동 결과물을 문서 자료로 수집하기 위해 모둠 및 개별 학습자료를 제작하여 이를 작성하도록 안내하였다. 통계적 소양의 자료 분석은 본 연구에서 개발한 통계적 소양 분석틀을 이용하여 딥러닝 예

Table 3. Statistical literacy framework for statistical problem-solving process

	AI convergence educational	Activities key	Level of statistical literacy		
			Low level	Medium level	High level
Formulate questions (1)	Problem identification	Context Variability Critical questioning	Unable to describe the content of the investigation	Although aware of the content being investigated, unable to articulate it clearly	Clearly articulate the content of the investigation
	Problem exploration	Context Variability Statistical • mathematical knowledge	Failure to define the meanings of terms in the problem	Understanding the meanings of terms in the problem but failing to present them manipulatively	Present the meanings of terms in the problem in a manipulative manner
	Goal setting	Context Variability	Setting the goal of problem solving but failing to identify factors that affect the results	Setting the goal of problem solving and recognizing influencing factors	Set the goal of problem solving and identify factors that influence the results, and control such factors
Collect data (2)	Solution exploration	Context Variability Statistical • mathematical knowledge	Collecting data without considering its reliability and validity	Utilizing statistically representative data collection methods but failing to consider their relevance to the research question	Utilize statistically representative data collection methods and collect data considering their relevance to the research question
	Collaboration methods and resources				
Analyze data (3)	Data processing (A)				
	Solution determination	Context Variability Statistical • mathematical knowledge	Omitting crucial aspects of the data or expressing quantitative features that do not match the investigation problem	Sing tables or graphs to represent statistical features of the data but not comprehensively expressing the core content of the research topic	Comprehensively visualize and represent the statistical features of data that are relevant to solving the research topic and problem
	Data processing (B)				
Execute collaborative activities	Execution of collaborative activities	Context Variability	Using methods that are statistically irrelevant or applying statistical principles and concepts that do not match the investigation problem	Analyzing data in relation to statistical principles and concepts but inconsistently considering the key features of the research problem	Analyze the data in relation to statistical principles and concepts, considering the key features of the research problem
Interpret results (4)	Goal achievement confirmation	Context Variability Critical perspective	Presenting results unrelated to the research topic and problem	Deriving valid results related to the research topic and problem but including some unjustified findings	Explain the relevance of the variables used in the analysis to the research topic and problem
	Results sharing and feedback	Context Variability Critical perspective	Holding intuitive and non-statistical beliefs relevant to the context and focusing on single factors in complex concepts and situations	Thinking contextually but not critically, and using statistical concepts qualitatively without appropriate justification	Understand the uncertainty of predictions and raise critical questions in various contexts
	Mutual learning confirmation				
	Collaborative intelligence update				

측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 학생들의 통계적 소양에 미치는 영향을 분석하였다. 또한, 딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학교 학생들의 컴퓨팅 사고력에 미치는 효과성에 대해 알아보기 위해 PSB-CTQ를 이용하여 사전-사후 검사 대응 표본 t-검정을 시행하였다(Jin, 2024).

Table 4. Composition of research tools and examples of survey questions

Domain	Question number	Question
Problem identification	1	I can determine what data is needed to solve the problem
	11	I can understand and articulate the problem
Solution exploration	1	I can identify the conditions necessary to solve the problem
	7	I can establish the sequence of steps for problem-solving
Programming	5	I can use programming languages
	13	I have experience in improving upon projects completed by others

연구 결과 및 분석

1. 딥러닝 예측을 이용한 통계교육 프로그램 개발

(1) 프로그램 개발 방향

본 연구에서는 선행연구 분석 결과를 바탕으로 고등학교 수학과 교육과정에 따른 인공지능 융합교육 프로그램의 개발 방향을 다음과 같이 설정하였다. 첫째, 인공지능의 개념을 이해하고 학습 방식을 수학적으로 해석할 수 있도록 한다. 인공지능을 사용하여 합리적으로 문제를 해결하고 이를 수학적으로 해석함으로써 수학의 유용성을 인식하는 경험을 제공하려 한다. 둘째, 데이터 분석 분야에서 널리 사용되는 프로그래밍 언어인 파이썬(python)과 고수준 신경망 API인 케라스(keras) 및 데이터 분석 라이브러리인 판다스(pandas)를 이용하여 딥러닝 예측 알고리즘을 개발한다. 셋째, 통계교육 프로그램에서 자신의 진로와 관련지어 현실의 문제를 탐구 주제로 선정하고 딥러닝을 이용하여 미래의 경향성을 예측할 수 있도록 한다. 넷째, 통계교육 프로그램은 대규모 데이터셋의 전처리를 과정 포함하며, 데이터의 편향성을 고려하여 데이터의 공정성을 추구하도록 한다. 한정된 표본을 사용하는 통계교육은 오늘날 정보 사회에서 흔한 불확실성을 효과적으로 다루지 못할 수 있다(Lee et al., 2023a). 다섯째, 통계적 문제해결 과정에서 맥락 및 변이성을 강조하며 협력 활동을 통해서 데이터 너머를 바라볼 수 있도록 한다.

(2) 통계교육 프로그램 개발

본 연구의 수업 활동은 큰 틀에서 두 가지로 구분할 수 있다. 첫째, ‘인공지능에 대한 교육’이며, 인공지능 개념을 학습하고 파이썬을 사용하여 직접 인공지능을 다루는 과정을 수행한다. 둘째, ‘인공지능의 활용 교육’으로 인공지능을 적용하여 통계적 문제해결 과정에 따라 비정형 문제를 해결하는 활동이다. 인공지능에 대한 교육은 3차시로 진행되며, 학생들이 제한된 시간 안에 케라스를 공학 도구로서 완전히 활용하는 것은 현실적으로 어려운 과제이다. 그러므로 딥러닝을 공학 도구로 활용할 수 있는 능력을 개발하기 위해 플립러닝 방식을 수업에 도입하였다. 자료 분석 활동에 앞서, 딥러닝 예측 알고리즘의 사용 방법에 대한 영상 강의를 학생들에게 제공하여 온라인 수업을 진행하였다.

‘인공지능에 대한 교육’은 세 가지 활동으로 구성되어 있다. 첫째, 이해 활동으로 인공지능의 개념을 이해하고 학습 방식을 수학적으로 분석한다. 둘째, 실습 활동으로 코랩(colab)을 사용하여 브라우저 내에서 파이썬의 기초 문법을 배우고 간단한 알고리즘을 직접 작성한다. 셋째, 시연 활동으로 코랩을 이용해 브라우저 내에서 파이썬으로 작성된 딥러닝 예측 알고리즘을 실행시켜 보고 결과를 확인한다. 그리고 ‘인공지능의 활용 교육’은 통계적 문제해결 과정을 통해서 현실 세계의 문제를 탐구하는 프로젝트 수업으로 구성하였다. GAISE 보고서(Franklin et al., 2007)의 통계적 문제해결 과정을 바탕으로 통계교육 프로그램을 설계 및 구성하였다. Ryu와 Han (2022)의 디지털 기반의 창의협력지능(CCI) 교수학습 모형을 수정하여 통계적 문제해결 과정에 따른 인공지능 융합교육 활동을 설계 및 구성하였다. CCI 수업의 Data processing은 데이터를 수집, 가공, 분석, 표현하는 활동으로 이뤄진다. 통계적 문제해결 과정에서는 데이터 수집은 자료 수집 단계에서 이루어지고, 데이터의 가공, 분석 그리고 표현은 자료 분석 단계에서 수행된다. 따라서, CCI 수업의 Data processing을 재구성하여 Data processing (A)에서는 데이터를 수집하는 활동으로 Data processing (B)에서는 데이터를 가공, 분석 그리고 표현하는 활동으로 조정되었다. 통계교육 프로그램의 차시별 세부 내용은 Table 5와 같다.

본 연구에서는 인공지능을 수학과 연계하기 위해서 데이터 표현, 데이터 분석, 비지도학습, 신경망, 예측, 최적화를 인공지능 기능으로 활용한다. 딥러닝 예측을 이용한 통계교육 프로그램에서 다루는 수학 내용을 구체적으로 정리하면 Table 6과 같다.

Table 5. Detailed contents for each session of the statistics education program

	Content	Detailed content
1	Understanding I	Machine learning <ul style="list-style-type: none"> -Concept of machine learning: relationship between artificial intelligence, machine learning, and deep learning -Methods of machine learning: supervised learning, unsupervised learning, reinforcement learning -Types of supervised learning: regression, classification
		Deep learning <ul style="list-style-type: none"> -Understanding perceptrons and the learning principle of artificial neural networks -Constructing a perceptron (predicting continuous output)
2	Understanding II	Deep learning <ul style="list-style-type: none"> -Linear regression -Loss function (mean squared error) -Understanding deep neural networks
	Practice I	Python <ul style="list-style-type: none"> -Access google colab -Data types, numerical operations, variables and inputs, lists and indexes
3	Practice II	Python <ul style="list-style-type: none"> -Creating functions and quadratic function algorithms
	Demonstration	Keras <ul style="list-style-type: none"> -Deep learning prediction algorithm result verification <ol style="list-style-type: none"> 1. Regression: predicting sales based on advertising expenses by medium 2. Classification: predicting diabetes based on patient diagnostic measurements
4	Formulate questions	Problem identification <ul style="list-style-type: none"> -Analyzing the given problem scenario, identifying and recognizing the issues
		Problem exploration <ul style="list-style-type: none"> -Researching to specify and define the problem
		Goal setting <ul style="list-style-type: none"> -Setting goals for problem-solving and confirming the ultimate objective to achieve -Creating questions for which answers can be obtained from the data <ol style="list-style-type: none"> 1. Selecting independent and dependent variables 2. Crafting a problem
5, 6	Collect data	Exploration of solutions <ul style="list-style-type: none"> -Understanding and analyzing the current situation of the problem -Exploring various solutions
		Collaboration methods and resources <ul style="list-style-type: none"> -Identifying required data and collaborators -Understanding deep learning prediction algorithms: <ol style="list-style-type: none"> 1. Regression model: predicting sales based on advertising expenses by medium 2. Classification model: predicting diabetes based on patient diagnostic measurements
		Data processing (A) <ul style="list-style-type: none"> -Data collection -Recording the data collection methods and sources of collected data
7	Analyze data	Decision making on solutions <ul style="list-style-type: none"> -Selecting data analysis tools (regression model, classification model)
		Data processing (B) <ul style="list-style-type: none"> -Processing and analyzing collected data using pandas and presenting it
		Execution of collaborative activities <ul style="list-style-type: none"> -Collaborating in the workspace to engage in problem-solving activities -Analyzing data using appropriate graphs and numerical methods, utilizing deep learning to address real-world problems and evaluate performance
8	Interpret results	<ul style="list-style-type: none"> -Interpret the analysis results, evaluate them in relation to the original problem, and interpret the analysis results -Share and evaluate the results and outputs -Confirm the participants' learning through collaborative activities -Continuously modify, enhance, and expand the results

Table 6. The mathematical content covered in a statistics education program utilizing deep learning predictions

AI functions		Related mathematical content
Data representation	Matrix	In the activity 'Constructing Perceptron' in the first session of understanding I, matrices are introduced as a method of representing data collected
Data analysis	Correlation coefficient	Correlation coefficient is used to measure and understand the degree of association between two variables. In this study, the importance of correlation analysis is emphasized in the second session of understanding II, 'Linear Regression'
Neural network	Linear function	In the 'Understanding Perceptron' process of understanding I, linear functions with two or more input values as independent variables are discussed
	Summation of sequences	In the second session of understanding II, 'Loss Functions,' summation of sequences is used to mathematically represent the loss function
Supervised and unsupervised learning	Regression analysis	The method of modeling the linear relationship between independent and dependent variables in 'Linear Regression' of the second session of understanding II is discussed
Prediction	Graph	In the 'Loss Functions' of the second session of understanding II, graphs are used to represent data as scatter plots and to draw trend lines to find predicted values
Optimization	Loss function	The concept of mean squared error is used to explain the principle of artificial intelligence learning in the 'Loss Functions' of the second session of understanding II

① 인공지능의 이해 I (1차시)

'인공지능 이해 I'의 학습 목표는 인공지능의 개념을 이해하고 학습 방식을 수학적으로 해석하는 것으로 설정하였다. 기계학습 이해에서는 기계학습의 개념, 방법, 유형을 학습한다. 딥러닝 이해 활동에서는 퍼셉트론 구성하며, 입력값, 가중치, 활성화 함수, 출력값의 의미를 이해하고, 예측값과 실제값의 차이를 최소화하는 방향으로 가중치를 조정하는 과정을 통해서 딥러닝 학습 메커니즘을 이해하도록 한다.

② 인공지능의 이해 II 및 실습 I (2차시)

'인공지능의 이해 II'의 학습 목표는 인공지능의 개념을 이해하고 학습 방식을 수학적으로 해석하는 것으로 설정하였다. '인공지능의 실습 I'의 학습 목표로는 파이썬의 기초 문법을 이해하고 간단한 코드를 작성하는 것이다. 다중 선형 회귀분석은 여러 개의 독립변수와 한 개의 종속변수 간의 선형적 관계를 다룬다. 손실함수, 특히 평균제곱오차(MSE)는 예측값과 실제값의 차이를 측정하는데 사용되며, 이를 통해 모델의 성능을 평가한다. 신경망의 이해에서는 단층 퍼셉트론과 다층 퍼셉트론의 차이를 구분하고, 더 복잡한 상황을 다루기 위해 노드(node)의 수와 은닉층(hidden layer)의 수를 조절해야 함을 강조한다. 인공지능의 실습 I에서는 코랩을 활용하여 브라우저 내에서 파이썬의 기초 문법을 가르친다. 자료형, 출력함수, 숫자 연산, 변수와 입력, 리스트(list)와 인덱스(index)를 주제로 파이썬의 사용 방법을 습득한다.

③ 인공지능의 실습 II 및 시연(3차시)

'인공지능의 실습 II'의 학습 목표는 파이썬의 기초 문법을 이해하고 간단한 코드를 작성하며, 공학 도구를 사용하여 데이터의 경향성을 추세선으로 나타내고 이를 예측에 이용하는 것이다. 인공지능의 실습 II에서는 실습 I에서 학습한 파이썬의 기초 문법을 활용하여 이차함수를 만들어 함수값을 출력해보고 순서쌍을 연결하여 그래프를 그린다. 딥러닝으로 데이터를 분석하고 예측할 수 있는 두 가지 딥러닝 예측 알고리즘을 개발하였다. 캐글(kaggle)의 데이터셋을 활용하여 광고비용을 기반으로 매출을 예측하는 회귀 모델과 환자의 진단 측정값을 기반으로 당뇨병을 예측하는 분류 모델이다. 문제의 특성에 따라 회귀 모델(연속 값 예측) 또는 분류 모델(이산값 예측) 중 하나를 선택하고 알고리즘을 간단히 수정하여 모델을 활용할 수 있다. 인공지능의 시연에서는 딥러닝 예측 알고리즘을 학생들에게 단계별로 시연하면서 딥러닝의 구조와 학습에 필요한 요소를 설명한다.

④ 문제 설정(4차시)

4차시는 통계적 문제해결 과정 중 문제 설정 단계로 해결해야 할 문제를 명확히 하고 자료를 가지고 답을 얻을 수 있는 문제를 만드는 것을 학습 목표로 설정하였다. 문제 설정을 위해서는 주어진 문제 상황을 자세히 검토하고 문제점을 인식하며, 문제를 구체화하고 정의하기 위한 조사를 수행한다. 또한, 문제해결의 목표를 설정하고 달성하고자 하는 최종 목표를 명확히 확인

해야 한다(Ryu & Han, 2022). 문제 설정 단계에서는 자료를 가지고 답을 얻을 수 있는 하나 이상의 문제를 만들어야 한다. 이 과정에서 통계적 문제를 구성할 때, 중요한 것은 결정론적인 답을 예측하는 문제와 데이터에 따라 답이 변동될 것으로 예상하는 문제의 차이를 이해하는 것이다(Franklin et al., 2007). 본 연구에서는 딥러닝 모델을 이용하여 데이터 분석을 수행한다. 딥러닝은 예측값과 실제값의 차이를 최소화하는 방향으로 입력 변수와 출력 변수 사이의 관계를 찾는 것을 목표로 한다. 따라서, 두 변수 사이의 관계를 이용하여 해결할 수 있는 제한된 문제 상황을 전제로 목표를 설정하고 문제를 구체화할 수 있도록 한다.

⑤ 자료 수집(5, 6차시)

5, 6차시는 통계적 문제해결 과정 중 자료 수집 단계로 적합한 자료를 수집하기 위한 계획을 설계하고 해당 계획을 사용하여 자료를 수집하는 것을 학습 목표로 설정하였다. 문제 상황을 파악하고 분석하며, 다양한 해결책을 찾는 과정을 통해 자료 수집에 필요한 데이터를 결정한다. 또한, 데이터가 편향될 가능성을 고려하고, 편향의 원인을 파악하여 데이터의 정확성을 높이는 방안을 모색해야 한다. 또한, 데이터의 변이성을 이해하고 변이성을 최소화하기 위한 고려가 필요하다(Franklin et al., 2007). 자료 수집 단계에서는 필요한 자료를 수집하고 해결책에 따른 협력 방법과 도구를 살펴보고 이해한다(Ryu & Han, 2022). 본 연구에서는 딥러닝 예측 알고리즘을 활용하므로, 출력값의 연속성을 고려하여 회귀 모델과 분류 모델을 구별할 수 있어야 한다. 3차시에서 시연한 딥러닝 예측 알고리즘의 단계별 알고리즘을 자세히 설명한다.

⑥ 자료 분석(7차시)

7차시는 통계적 문제해결에서 자료 분석 단계로 적절한 그래프 및 수치적인 방법을 선택하여 자료를 분석하고, 딥러닝을 활용하여 현실적인 문제를 해결하고 모델의 성능을 평가하는 것을 학습 목표로 설정하였다. 문제 설정 단계에서 설정한 최종 목표에 맞게 회귀 모델, 분류 모델 중 하나를 데이터 분석 도구로 선택한다. 수집한 데이터의 독립변수와 종속변수를 정의하고, 판다스를 활용하여 데이터를 체계적으로 분석한다. 대규모 데이터셋을 전처리하며, 데이터의 편향성을 고려하여 데이터의 공정성을 유지하려 노력한다. 이러한 과정을 통해 딥러닝을 활용하여 실제 문제를 해결하고 모델의 성능을 개선하는 데 초점을 맞추었다.

⑦ 결과 해석(8차시)

8차시는 통계적 문제해결 과정에서 결과 해석 단계로 분석 결과를 해석하고 해석을 원래 문제와 관련지어 분석 결과를 평가한다. 문제 이해 단계에서 설정한 목표를 달성하였는지 확인하고, 결과 및 산출물을 다른 이들과 공유하며 평가한다. 또한, 협력 활동을 통한 참여자의 학습을 확인하고 결과를 계속해서 수정, 보완, 확대한다. 이 과정에서 데이터 내의 변이성을 허용하여 데이터를 일반화하기 위해 데이터 너머의 관점을 강조한다. 데이터 너머를 바라본다는 것은 데이터의 변이성을 허용함을 의미한다(Franklin et al., 2007).

(3) 딥러닝 예측 알고리즘 개발

딥러닝 예측 알고리즘은 다음과 같이 총 13단계의 알고리즘으로 설명할 수 있으며, 분류 모델일 때 카테고리형 데이터를 수형 데이터로 변환해야 하는 4-2단계가 추가된다.

1단계, 엑셀에 입력한 data.csv 파일 업로드

```
1 from google.colab import files
2 uploaded=files.upload()
```

2단계, 판다스를 이용하여 CSV파일을 데이터 프레임으로 저장

```
1 import pandas as pd
2 import numpy as np
3 df=pd.read_csv('Advertising.csv') #'(name.csv)'안에는 파일의 주소를 입력한다.
4 print('1. 데이터프레임 shape')
5 print(df.shape)
```

```

6 print("\n2. 데이터프레임 정보')
7 print(df.info()) #데이터의 정보
8 print("\n3. 데이터프레임 기초통계량')
9 print(df.describe())
10 print("\n3. 데이터프레임 상위 5개 파일 출력')
11 print(df.head()) #df의 상위 5개 파일 출력

```

3단계, 데이터의 시각화

```

1 import matplotlib.pyplot as plt
2 x=df['TV'] # x축 변수값을 입력한다.
3 y=df['sales'] # y축 변수값을 입력한다.
4 fig=plt.figure(figsize=(6,4))
5 plt.scatter(x, y,color='violet',marker='o')
6 plt.title('Graph')
7 plt.xlabel('X')
8 plt.ylabel('Y')
9 plt.legend(['(x,y)'])
10 plt.show()

```

4-1단계, 상관관계수 찾기

- Seaborn(SNS)를 사용한 파이썬 데이터 시각화.
- 상관관계수는 -1부터 1 사이의 값으로 1에 가까울수록 양의 상관관계이고 -1 가까울수록 음의 상관관계라 한다.
- 밝을수록 양의 상관관계이고 어두울수록 음의 상관관계이다.

```

1 import seaborn as sns
2 sns.heatmap(df.corr(),annot=True)

```

4-2단계(분류 모델), 카테고리형 데이터를 수치형으로 변형

```

1 from sklearn.preprocessing import LabelEncoder
2 le=LabelEncoder()
3 target=le.fit_transform(df['Outcome'])
4 df['Outcome']=target
5 print(target)
6 print(le.classes_)

```

5단계, 데이터 전처리(data preprocessing) (1)

- 결측치 제거
- IQR 방식을 이용한 이상치 제거

```

1 print('1. 각 열에 포함된 결측값의 수')
2 print(df.isnull().sum()) # 각 열에 결측치가 몇 개 포함되어 있는지 확인
3 df.dropna(inplace=True) # 각 열에 결측값 제거
4 col=['TV','radio'] # ['독립변수1', '독립변수2']을 입력한다.
5 for k in col:
6     q1=df[k].quantile(0.25)
7     q3=df[k].quantile(0.75)
8     IQR=q3-q1
9     rev_range=1.5 # 일반적으로 1.5를 사용
10    df=df[(df[k]<=q3+(rev_range*IQR))&(df[k]>=q1-(rev_range*IQR))]
11    df=df.reset_index(drop=True)
12    print('\n2. 데이터프레임 shape')
13    print(df.shape)

```

6단계, 사이킷런으로 훈련 세트와 테스트 세트 나누기

- `train_test_split()`은 데이터를 훈련 데이터와 테스트 데이터로 나누는 함수이다.
- 테스트 세트로 나눌 비율은 `test_size` 매개변수에 지정할 수 있으며 기본값은 0.25 (25%)이다.

```
1 from sklearn.model_selection import train_test_split
2 df_train, df_test = train_test_split(df, test_size=0.2)
```

7단계, 데이터 전처리(data preprocessing) (2)

- `StandardScaler` 클래스를 사용해 훈련 세트와 테스트 세트에 표준화 전처리한다.
- 스케일(특성의 값이 놓인 범위)이 다른 특성의 값을 일정한 기준으로 맞춰준다.
- 표준점수(Z 점수): 각 특성값이 평균에서 표준편차의 몇 배만큼 떨어져 있는지를 나타냄.

```
1 from sklearn.preprocessing import StandardScaler
2 scaler=StandardScaler()
3 train_input=df_train[['TV','radio']] #df_train[['독립변수_1','독립변수_2']]을 입력한다.
4 test_input=df_test[['TV','radio']] #df_test[['독립변수_1','독립변수_2']]을 입력한다.
5 scaler.fit(train_input)
6 train_input=scaler.transform(train_input)
7 test_input=scaler.transform(test_input)
8 train_target=df_train['sales'] #df_train['종속변수']을 입력한다.
9 test_target=df_test['sales'] #df_test['종속변수']을 입력한다.
10 print('훈련 데이터 shape:', train_input.shape)
```

8단계, 심층신경망 모델 만들기

- `Sequential`은 케라스에서 신경망을 만드는 클래스이다.
- `Input layer`는 입력층으로 데이터의 크기를 입력한다.
- `Dense`는 신경망에서 밀집 층을 만드는 클래스이다. `Dense(뉴런의 개수, activation='활성화함수')`
- 출력층(Output)의 뉴런의 개수 및 활성화 함수의 종류
 - 회귀: 뉴런의 개수=1, 활성화 함수 없음
 - 분류: 뉴런의 개수=(class 개수), `activation='softmax'`

```
1 from tensorflow.keras.layers import Input, Dense
2 from tensorflow.keras.models import Sequential, Model
3 from tensorflow.keras.utils import plot_model
4 model=Sequential([Input(shape=(2,), name='input'),
5                   Dense(16, activation='relu', name='Dense1'),
6                   Dense(8, activation='relu', name='Dense2'),
7                   Dense(1, name='output')]) # 회귀 문제이므로 활성화 함수 없음
8 model.summary()
9 print('\n<심층 신경망 모델>')
10 plot_model(model)
```

9단계, 신경망 모델 설정하기

- `model.compile()`을 이용하여 최적화 알고리즘(adam) 및 손실함수와 훈련과정에서 계산하고 싶은 정확도 지표를 지정한다.
- `model.compile(optimizer='adam', loss='손실함수', metrics='측정값')`
 - 회귀: `loss='mse', metrics=['mse']`
 - 분류: `loss='sparse_categorical_crossentropy', metrics=['accuracy']`

```
1 from tensorflow.keras.optimizers import Adam
2 model.compile(optimizer=Adam(learning_rate=0.001), loss='mse', metrics=['mse'])
```

10단계, 훈련 fit()

- epoch=n : 훈련데이터 모두를 n번 반복훈련한다.
- batch_size=n : 훈련데이터를 batch size 크기로 쪼개어 학습한다.

```
1 history=model.fit(train_input, train_target,
2                   epochs=20,
3                   batch_size=4)
```

11단계, 손실함수의 시각화

```
1 history_dict=history.history
2 loss=history_dict['loss']
3 epoch=range(1,len(loss)+1)
4 fig=plt.figure(figsize=(5,2))
5 plt.plot(epoch, loss, color='red', label='train_loss')
6 plt.title('Train Loss')
7 plt.xlabel('Epoch') #x 축: epoch(훈련 횟수)
8 plt.ylabel('loss') #y 축: loss(손실)
9 plt.grid()
10 plt.legend()
```

12단계, 모델의 성능평가(evaluate)

- 테스트 데이터를 이용하여 모델의 성능을 평가한다.

```
1 model.evaluate(test_input,test_target,verbose=2)
```

13단계, 예측(predict)

- predict() 메서드로 새로운 데이터의 정답을 예측한다.
- predict() 메서드의 변수는 훈련 데이터와 같은 2차원으로 입력한다.

```
1 predict=model.predict(scaler.transform([[23,15]])) # 2차원으로 입력
2 print('예측값:',predict)
```

2. 딥러닝 예측을 이용한 통계교육 프로그램 적용 결과 분석

(1) 통계적 소양 분석

본 연구에서는 통계적 문제해결 과정에 따른 통계적 소양 분석틀을 이용하여 학생들의 통계적 소양 수준을 분석하였다. 인 공지능의 활용 교육에서 수집한 문서 및 음성 녹음 자료, 면담, 동영상 녹화 자료를 분석하였다. 모둠별 탐구 주제를 정리하면 Table 7과 같다.

Table 7. Group inquiry topics

Output	Student	Research topic
A	T1 ¹⁾ -T3 본 연구에 참여한 학생 12명을 T1-T12으로 나타낸다.	What is the correlation between sleep quality?
B	T4-T6	Correlation between OTT viewing rate and cable viewing rate
C	T7-T9	What is the relationship between the trend of COVID-19 pandemic and the OTT usage rate?
D	T10-T12	What is the hourly wage for women by country?

Table 3을 이용하여 통계적 문제해결 과정에서 나타난 학생들의 통계적 소양의 수준을 상, 중, 하로 구분하여 분석한 결과는 Table 8과 같다. 아래에서는 통계적 문제해결 과정의 단계별로 학생들이 보인 반응을 구체적으로 살펴보고자 한다.

Table 8. Analysis results of statistical literacy levels

	AI-integrated educational activities	Statistical literacy levels		
		Low	Intermediate	High
Formulate questions	Problem identification	0	5	7
	Problem exploration	0	6	6
	Goal setting	0	8	4
Collect data	Exploration of solutions	12	0	0
	Collaboration methods and resources			
Analyze data	Data processing (A)			
	Decision making on solutions	11	0	1
	Data processing (B)			
Interpret results	Execution of collaborative activities	6	5	1
	Verification of goal achievement	0	8	4
	Sharing results and feedback	3	9	0
	Confirmation of mutual learning			
	Update on collaborative intelligence			

① 문제 설정

(가) 통계적 소양 수준 분석

‘문제 설정’은 해결해야 할 문제를 명확히 하고 자료를 가지고 답을 얻을 수 있는 문제를 만드는 과정이다. 문제 설정 단계는 문제 인식, 문제 탐구, 목표 확인 단계로 세분화하여 통계적 소양을 평가하였다. 첫째, 문제 인식 단계는 맥락에서 문제 상황을 인식하고 문제점을 찾아 조사하고자 하는 내용을 명확하게 기술하는 과정이다. 분석 결과, 상 수준이 가장 많이 나타났다. 문제 인식 단계에서 학생들은 조사하고자 하는 문제 상황을 명확히 이해하고 이를 기술했지만, 문제점을 찾는 것에는 어려움을 나타내었다. 학생 T9는 개별 학습 자료에서 ‘코로나19가 확산하여 사회적 거리 두기로 인해 집에 있는 시간이 늘어나 콘텐츠 소비 문화가 영화관에서 OTT 사이트 구독으로 바뀌어 가고 있다’라는 사회현상에 대해 명확하게 기술하였으므로 상 수준으로 분류하였다. 둘째, 문제 탐구 단계는 문제를 구체화하고 정의하기 위해 설정한 문제에 포함된 용어들의 의미를 명확하게 제시하는 과정을 의미한다. 분석 결과, 하 수준의 결과물은 나타나지 않았다. 중 수준으로 나타난 결과물 A는 ‘잠의 질’을 주관적인 평가로 1부터 10의 범위로 평가하여 나타내었다. ‘잠의 질’을 수치화했지만, 이는 평가 기준이 명확하지 않은 주관적인 평가로 용어의 의미를 올바르게 제시하지 못하였다. 셋째, 목표 확인 단계는 문제해결의 목표를 설정하는 과정으로 설정한 문제와 관련하여 결과에 영향을 미치는 요인을 나타내었는지를 분석하였다. 분석 결과, 중 수준이 가장 많이 나타났다. 결과물 A에서는 상관 계수를 이용하여 결과에 영향을 미치는 요인을 이해하고 변인을 설정했기 때문에 상 수준으로 분류하였다.

(나) 질적분석

첫째, 일반적으로 결과에 영향을 미치는 요인을 단일 요인만으로 인식하고 이외 다양한 요인에 대한 분석 및 통제가 부족했다.

〈사례 1〉

학생 T1: 쇼핑물 상품 추천 알고리즘을 주제로 정하면 어때?

학생 T2: 그러면 독립변수가 뭐야?

학생 T1: 함수 관계로 나타내면 독립변수는 X고 종속변수는 Y.

학생 T2: 음.

〈사례 1〉의 학생들의 대화에서 나타나는 바와 같이, 학생들은 주어진 문제 상황에서 독립변수와 종속변수를 찾고 이를 표현하는 데 어려움을 겪고 있다. 학생들은 이미 주어진 한 가지 변수를 사용하여 다른 변수를 표현하는 방법에만 중점을 두고 있었다.

둘째, 학생들은 자신들이 직접 경험한 현실을 바탕으로 문제를 설정하기보다는 이미 수집된 자료를 바탕으로 문제를 정하는 경향이 있다.

〈사례 2〉

학생 T11: 주제는 무엇으로 할까?

학생 T10: 융합인재교육의 효과 어때? 융합인재교육을 받은 사람과 아닌 사람들의 효과성을 분석하자

학생 T11: 자료가 있을지 모르겠어. 융합인재교육의 효과 검색해 볼게.

학생 T12: 너무 뭐가 안 나온다. 데이터가 잘 나올만한 주제로 선정하자.

학생 T11: 그러면 자료가 많이 나오는 것으로 하자. 캐글에서 주제를 찾아보자.

학생 T10, T11, T12는 융합인재교육의 효과성을 알아보고자 융합인재교육의 효과성을 주제로 선정하였다. 그러나 융합인재교육과 관련된 자료를 쉽게 찾을 수 없었고, 만약 자료를 찾는 데에 성공한다 해도 데이터를 정리하는 과정이 어려울 것으로 판단하였다. 데이터 분석이 가능하도록 자료를 정리하여 제공하는 포털 사이트를 통해 주제를 선정할 때는 문제 설정을 위해 깊이 고민할 필요가 없고 자료를 쉽게 분석할 수 있는 이점이 있다. 그러나 자료를 수집하는 과정에서 얻을 수 있는 다양한 경험을 놓치게 될 수 있다는 단점도 고려해야 한다.

② 자료 수집

(가) 통계적 소양 수준 분석

‘자료 수집’은 적합한 자료를 수집하기 위한 계획을 설계하고 해당 계획을 사용하여 자료를 수집하는 과정이다. 이 과정에서 자료의 신뢰성과 타당성을 확인하고 충분한 근거를 제시하는 것이 중요인데, 이에 중점을 두어 분석을 진행하였다. 분석 결과 모든 과제물을 하 수준으로 평가했다. 모든 과제는 2차 자료로서 포털 사이트 및 신문 등을 이용하여 간접적으로 수집한 자료를 활용하였다. 자료 수집의 출처는 언급되었으나 모집단과 표본 추출 방법 등에 대해서는 언급하지 않았다. 이는 학생들이 모집단과 표본에 대한 통계적 이해가 부족한 것으로 해석된다. 이러한 문제를 해결하기 위해서는 통계 수업에서 모집단과 표본의 관계를 맥락으로 이해하고 이를 실생활에 적용하는 경험을 체계적으로 제공하는 것이 필요하다.

(나) 질적분석

첫째, 학생들은 수집한 자료가 갖는 자료의 형태(범주형 자료, 수치형 자료)와 척도를 이해하지 못하여 자료 수집에 어려움을 보였다.

〈사례 3〉

연구자: 자료는 찾아봤니?

학생 T4: OTT 서비스를 못 찾겠어요.

학생 T5: 이 자료는 케이블 방송 가입률을 찾은 거예요. 연도별로 찾았어요.

학생 T4: 연도별로 하려고 해도 OTT가 시작한 지 몇 년 안 되어서 데이터 수가 적고 달별로 되어 있는 자료는 찾기 어려워요.

자료를 수집하기 위해서는 먼저 자료의 특성을 정확히 구분할 수 있어야 한다. 자료가 갖는 특성은 척도에 따라 명목척도, 순서 척도, 구간 척도, 비율척도로 구분한다. 〈사례 3〉에서는 시청 비율을 비율척도로 나타낼 수 있다. 이때, 비율척도는 절대 영점을 가지며 사칙연산이 가능한 특성이 있다. 그러나 〈사례 3〉에서 학생 T5는 연도를 매개로 하여 케이블 시청 비율과 OTT 시청 비율의 자료를 수집하였지만, 수집한 두 자료의 절대 영점이 달라서 이를 비교하기 어렵다.

둘째, 학생들은 대화에서 딥러닝을 활용한 통계수업과 일반적인 통계 수업의 가장 큰 차이점으로 자료의 특성(원자료)을 지적했다.

〈면담 1〉

연구자: 딥러닝을 활용한 통계 수업이 통계적 정보를 해석하고 비판적으로 평가하는 능력에 어떤 영향을 주었던 것 같아?

학생 T6: 통계자료를 보면 몇 개는 오류가 있거나 과하게 크거나 작게 나타나는 때도 있는데 프로그래밍을 활용하면 오류값을 쉽게 제거할 수 있었어요.

학생 T6의 답변을 통해서 확인할 수 있듯이, 본 연구 수업은 기존의 통계수업과 달리 원자료(raw data)를 직접 수집하고 활용하였다. 또한, 원자료를 데이터 분석 목적과 방법에 맞는 형태로 처리하여 자료를 정리하였다. 이를 데이터 분석에 활용하여 설정한 문제를 해결하도록 안내하였다. 이러한 접근 방식은 학생에게는 다양한 자료에 대한 경험을 제공했고, 자료 수집 및 전처리의 필요성을 깨닫게 했다.

③ 자료 분석

(가) 통계적 소양 수준 분석

‘자료 분석’은 탐구 문제해결을 위해 관련성 있는 자료의 통계적 특징을 시각화하고 딥러닝을 활용하여 자료를 분석하는 과정을 의미한다. 이 과정을 해결책 결정 및 데이터 처리, 협력 활동 실행 단계로 세분화하여 통계적 소양을 평가하였다. 첫째, 해결책 결정 및 데이터 처리 단계는 탐구 문제와 관련성 있는 자료를 표나 그래프로 통계적 특징을 표현하는 과정을 의미한다. 분석 결과, 하 수준이 가장 많이 나타났다. 이는 탐구 문제와 관련성이 높은 변수를 빠뜨렸거나 탐구 문제와 불일치한 양적 특징을 표현한 것이다. 예를 들어, 결과물 D의 경우 탐구 문제로 ‘국가별 여성의 시간별 급여는 어떻게 될까?’로 설정하였다. 연도와 시간별 급여의 관계를 산포도로 나타내었으며 여성의 시간별 급여가 시간이 지날수록 높아지고 있다고 해석하였다. 이는 국가라는 양적 특징을 빠뜨렸으며 탐구 문제와 불일치한 양적 특징을 표현한 것으로 하 수준으로 평가하였다. 하지만 학생 T6는 개별 학습 자료에서 OTT 시청 비율과 케이블 시청 비율의 관계를 산포도로 나타내고 두 변수 사이의 관계를 정확히 표현하여 상 수준으로 평가하였다. 둘째, 협력 활동 실행 단계에는 통계학적 원리 및 개념과 관련지어 자료를 분석하는 활동을 한다. 분석 결과, 하 수준이 가장 많이 나타났다. 공통적으로 탐구 문제의 주요 특징을 일관적으로 고려하지 않고 주제와 관련성이 없는 내용을 표현하였다. 예를 들어 과제물 C의 경우 코로나19 유행 추세와 OTT 사용률의 관계를 분석하기 위해 코로나19 신규 감염자수와 모바일 넷플릭스 사용자의 상관관계를 분석하였다. 그러나 문제 설정 단계에서부터 OTT 사용률의 의미를 모바일 넷플릭스 사용자로 제한하여 탐구 문제와 일관성이 없는 탐구 활동이 이루어졌다.

(나) 질적분석

인포그래픽(inforgraphic)이란 인포메이션 그래픽(information graphics)의 줄임말로 자료를 시각적으로 표현한 것을 의미한다. 본 연구 활동에서는 수집한 자료를 빠르게 이해하고 변수 간의 관계를 다양한 이미지 정보로 표현해 주기 위해서 맷플롯립(matplotlib)과 씨본(seaborn) 프로그래밍 패키지를 활용하였으며 이를 정리하면 Figure 2와 같다.

첫째, 학생들은 인포그래픽을 활용한 수업에서 긍정적인 태도를 보였다. 〈면담 2〉에서 알 수 있듯이, 학생들은 그래프를 활용

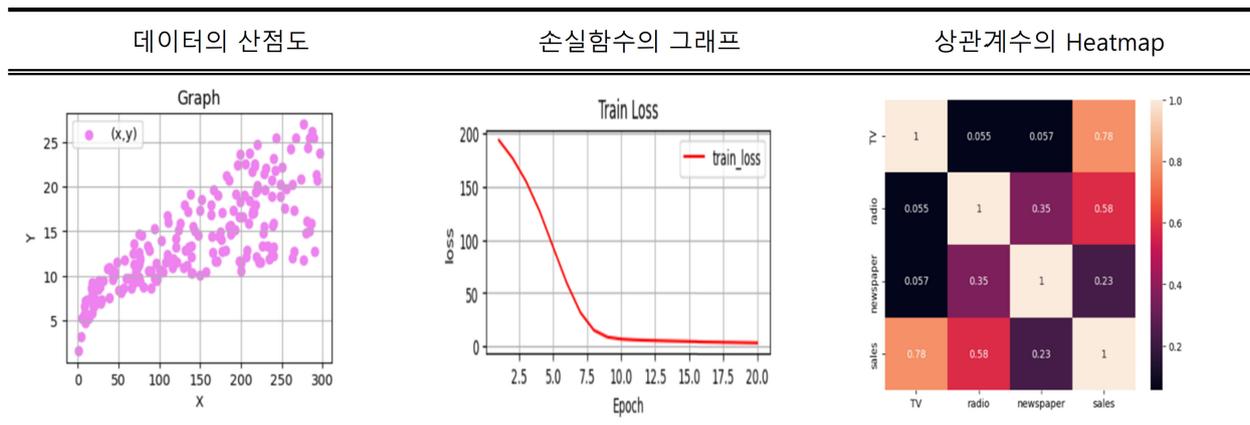


Figure 2. Infographics using programming packages.

하여 모뎀원의 기댓값을 예측하는 등 인포그래픽을 효과적으로 활용하는 모습을 보였다.

〈면담 2〉

연구자: 모뎀원 기대 예측값은 어떻게 추측했어?

학생 T6: 시간과 사용률을 축으로 OTT와 케이블 사용률을 각각 산점도로 표현하고 추세선을 만들어 구체적인 수치를 추측했어요.

학생 T6는 'OTT 시청 비율과 케이블 시청 비율의 상관관계'를 문제로 설정하여 OTT 시청 비율에 따른 케이블 시청 비율을 예측하는 회귀 모델을 개발했다. 〈면담 2〉에서 알 수 있듯이, 학생 T6는 모델의 성능 향상을 위해 손실함수의 그래프를 활용하고 지속해서 모델을 평가하고 수정함으로써 최적의 모델을 개발하였다. 또한, 수집한 데이터를 시간과 사용률을 축으로 하는 산포도로 표현하고 추세선을 만들어 모뎀원 기대 예측값을 찾는 데 그래프를 효과적으로 활용했다.

둘째, 학생들은 자료의 경향성을 오로지 선형적인 관점에서만 바라보는 경향을 보였다. 결과물 A, B, C에서 산점도의 추세선을 직선으로만 표현함으로써, 선형적 문제에만 집중하는 한계를 드러냈다.

〈사례 4〉

학생 T10: 우리가 지금 해야 하는 게 뭐지?

학생 T11: '무엇에 따라 증가한다 또는 감소한다' 이런 자료를 찾아봐야 하는 거지? 그런데 자료가 많이 나오는 게 뭐 있을까?

학생 T11는 두 변수 사이의 관계를 '무엇에 따라 증가하거나 감소하는지'와 같이 제한적으로 정의하였다. 이는 두 변수 사이의 관계를 한 변수의 값이 커짐에 따라 다른 하나의 변수도 커지는 양의 상관관계 또는 한 변수의 값이 커짐에 따라 다른 변수의 값이 작아지는 음의 상관관계로 제한하여 문제를 설정한 결과로 해석된다.

④ 결과 해석

(가) 통계적 소양 수준 분석

'결과 해석'은 분석 결과를 해석하고 문제 설정 단계에서 만들었던 문제와 관련지어 다양한 맥락에서 비판적인 의문을 제기하는 과정이다. 결과 해석은 다음과 같은 기준으로 세분화하여 평가하였다. 첫째, 목표 달성 확인 단계는 분석 결과를 해석하는 과정으로 탐구 문제와 일치하게 분석 결과를 설명하였는지를 평가하였다. 분석 결과, 중 수준이 가장 많이 나타났다. 과제물 B, C, D는 탐구 주제와 일치하지 않는 주장을 포함하고 있으므로 중 수준 또는 하 수준으로 평가하였다. 둘째, 결과 공유 및 피드백, 상호 학습 확인, 협력기능 업데이트 단계는 결과물을 공유하고 평가하며 협력 활동을 통해 결과물을 보완하고 확대하는 과정을 나타낸다. 맥락과 관련된 생각을 하고 비판적 의문을 제기하는 능력이 중점적으로 평가되었다. 분석 결과, 상 수준의 결과물은 나타나지 않았다. 맥락과 관련지어 문제를 설정하였지만 비판적 의문을 제기하는 경우는 거의 나타나지 않았다.

(나) 질적분석

학생들은 딥러닝 예측 알고리즘을 이용하여 데이터를 예측하는 과정에서 예측한 결과의 불확실성을 인지하고 모델을 지속해서 수정하여 보다 좋은 예측값을 구하기 위해 노력하는 모습을 보였다.

〈면담 3〉

연구자: 탐구활동에서 흥미로웠던 활동이 있었니?

학생 T1: 조금 더 예측값이 잘 나오기 위해서 어떻게 훈련 횟수를 늘려야 하는지 알게 되었고, 손실함수로 예측값을 더 정확하게 구할 수 있는지 확인할 수 있었어요. Batch size와 노드 값을 바꾸는 게 어려웠지만 흥미로웠고 활동을 하면서 어떻게 프로그램을 조정해야 더 좋은 예측값을 얻을 수 있는지 배우고 새로운 걸 알아가서 좋았어요.

〈면담 3〉에서 알 수 있듯이, 학생 T1은 수집한 데이터와 전처리 과정에 따라 예측값이 변동될 것으로 예상하며 프로그램을 통해서 다양한 결과값을 만들어 낼 수 있음을 활동을 통해서 깨닫게 되었다. 이는 자료의 변이성에 대한 이해를 보인 것으로 판단할 수 있다.

(2) 컴퓨팅 사고력 분석

딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생들의 컴퓨팅 사고력에 미치는 영향을 분석하고자 대응 표본 t-검정을 시행하였고 결과를 정리하면 Table 9와 같다. 분석 결과 수업 전과 후의 컴퓨팅 사고력 ($t=-11.657, p<0.001$)은 통계적으로 유의한 차이를 나타냈다. 따라서 딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램은 학생들의 컴퓨팅 사고력을 높이는 긍정적인 영향을 주었음을 알 수 있다. PSB-CTQ 영역별 사전·사후 분석 결과를 살펴보면, 문제인식($t=-6.832, p<0.001$), 해결방안탐색($t=-4.024, p<0.001$), 프로그래밍($t=-8.985, p<0.001$)에서 통계적으로 유의한 차이를 보였다. 컴퓨팅 사고력의 하위영역 전체에서 사전 평균 점수보다 사후 평균 점수가 유의미하게 증가하였다. 특히, '프로그래밍' 영역의 평균 점수 2.40에 3.12로 가장 높은 점수의 상승을 보였다. 이는 인공지능의 이해와 활용을 중점으로 하는 통계 프로그램이 프로그래밍에 대한 이해와 자신감을 높이는 데 영향을 주었음을 판단할 수 있다.

컴퓨팅 사고력 영역의 하위 요소별 수업 전과 후에 실시한 결과를 정리하면 Table 10과 같다. 문제인식 영역에서는 자료수집

Table 9. Pre- and post-inspection results for PSB-CTQ overall and by domain

Domain	Pre-inspection		Post-inspection		t	p
	M	SD	M	SD		
Computational thinking	2.67	.700	3.18	.590	-11.657***	<0.001
Problem identification	2.85	.540	3.25	.533	-6.832***	<0.001
Solution exploration	2.88	.533	3.16	.486	-4.024***	<0.001
Programming	2.40	.809	3.12	.678	-8.985***	<0.001

*** $p<0.001$.

($t=-3.469, p=0.001$), 자료분석($t=-2.996, p=0.005$), 자료표현($t=-5.277, p<0.001$) 문항에서 통계적으로 유의한 결과를 나타낸다. 해결방안탐색 영역에서는 추상화($t=-3.389, p=0.002$) 문항에서만 유의한 차이가 나타났다. 프로그래밍 영역에서는 자동화($t=-4.061, p<0.001$), 시뮬레이션($t=-4.053, p<0.001$), 병렬화($t=-4.905, p<0.001$) 문항에서 통계적으로 유의한 결과가 나타났다.

Table 10. Pre- and post-inspection results by subcomponents of computational thinking domain

Domain	Subcomponents	Pre-inspection		Post-inspection		t	p
		M	SD	M	SD		
Problem identification	Data collection	2.97	.385	3.29	.459	-3.469*	0.001
	Data analysis	2.97	.559	3.33	.585	-2.996*	0.005
	Data representation	2.64	.601	3.16	.558	-5.277***	<0.001
Solution exploration	Problem decomposition	2.86	.542	3.08	.554	-1.848	0.073
	Abstraction	2.86	.542	3.22	.421	-3.389*	0.002
Programming	Algorithms	2.58	.767	3.25	.601	-4.848***	<0.001
	Automation	2.25	.840	3.02	.774	-4.061***	<0.001
	Simulation	2.33	.761	3.16	.637	-4.053***	<0.001
	Parallelization	2.37	.841	3.06	.696	-4.905***	<0.001

* $p<0.05$, *** $p<0.001$.

① 문제인식

문제인식 영역은 해결해야 하는 문제에 적합한 자료를 모으는 과정인 자료 수집 문항과 자료를 이해하고 패턴을 찾아 결론을 도출하는 자료분석 문항 그리고 적절한 그래프, 차트, 글, 그림 등으로 자료를 정리하는 자료표현 문항으로 총 11문항으로 구성하였다. 문제인식 영역의 모든 문항에서 통계적으로 유의한 평균 차이를 보였다. 특히, 자료표현에서 차이가 가장 높게 나타났다. 학생들은 자료를 수집하고 분석하는 것보다 자료를 그림, 표, 식으로 표현하는 것을 상대적으로 어렵게 느끼고 있으며 수업 후에도 같은 어려움을 겪고 있음을 알 수 있다. 이러한 문제를 해결하기 위해서는 분석한 자료를 다양하게 표현하고 표현한 자료를 모둠 안에서 그리고 모둠 간에서 논의하는 집단적 협력 활동의 필요성을 시사한다.

② 해결방안 탐색

해결방안 탐색 영역은 문제를 해결 가능한 수준의 작은 문제로 나누는 문제분해 문항과 문제해결을 위해 필요한 핵심 요소를 파악하고 복잡함을 단순화하는 추상화 문항과 절차 문제를 해결하거나 어떤 목표를 달성하기 위해 수행되는 일련의 단계인 알고리즘 문항으로 총 7가지 문항으로 구성하였다. 해결방안 탐색 영역의 추상화 문항에서 통계적으로 유의미한 결과가 나타났다. 이는 딥러닝 예측 알고리즘의 4단계에서 Seaborn 프로그램 패키지를 사용하여 데이터의 상관관계를 시각화함으로써 결과에 영향을 미치는 핵심 요소를 쉽게 파악하고 5단계에서 판다스를 이용하여 결측치 및 이상치를 제거했던 경험에 의한 결과로 해석된다. 반면, 문제분해의 모든 문항에서는 점수가 향상하였지만, 통계적으로 유의미한 차이는 나타나지 않았다. 이는 통계적 문제해결 과정의 문제 설정 단계에서 해결해야 할 문제를 명확히 하고 문제를 만드는 과정에서 문제를 분해하는 것에 어려움을 경험한 것으로 해석된다.

③ 프로그래밍

프로그래밍 영역은 절차 문제를 해결하거나 어떤 목표를 달성하기 위해 수행되는 일련의 단계인 알고리즘 문항과 컴퓨팅 시스템이 수행할 수 있는 형태로 해결책을 나타내는 자동화 문항, 그리고 문제를 해결하기 위하여 만든 모델을 실행시켜 결과를 파악하는 시뮬레이션 문항으로 이루어져 있다. 또한, 목표를 달성하기 위한 작업을 동시에 수행하도록 자원을 구성하는 병렬화 문항으로 총 13문항으로 구성하였다. 본 연구는 딥러닝이라는 공학적 도구를 이용하여 현실의 문제를 해결하는 것을 목표로 설정했다. 즉, 통계교육 프로그램 활동을 통해 학생들에게 결과에 영향을 미치는 다양한 요소를 찾고 이를 딥러닝을 활용하여 분석함으로써 문제의 해결책을 찾는 경험을 학생들에게 제공했다. 이로써, 프로젝트 활동을 통해 학생들이 모둠별로 완성한 프로그램을 활용하여 실제 문제를 해결할 수 있다는 긍정적 태도를 형성하게 된 것으로 판단된다.

결론 및 제언

1. 요약 및 결론

4차 산업혁명과 인공지능의 발전으로 교육 분야에서 많은 변화가 일어나고 있다. 특히, 인공지능을 기반으로 하는 교육의 중요성이 강조되며 이러한 교육은 문제해결을 중심으로 진행되어 실제적인 문제를 해결하고 학습자들에게 유의미한 경험을 제공한다. 딥러닝은 인공 신경망을 사용하여 기계학습을 구현하는 기술로, 데이터의 복잡한 패턴을 인식하고 학습하여 정확한 예측을 생성할 수 있다. 이러한 딥러닝 기술은 실생활 문제해결에 활용할 수 있으며, 특히 통계교육 분야에 유용하게 활용될 수 있다. 본 연구의 목적은 고등학교 수학에서 딥러닝 예측을 이용한 통계교육 프로그램을 개발하고 이러한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생들의 통계적 소양 및 컴퓨팅 사고력에 미치는 영향을 알아보는 데 있다. 연구의 목적을 실현하기 위해 고등학교 수학에 적용할 수 있는 딥러닝 예측을 이용한 통계교육 프로그램을 개발하였다.

딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생의 통계적 소양에 미치는 영향은 다음과 같다. 첫째, 맥락에 대한 이해도가 향상되었다. 인공지능을 활용한 자료 분석은 실제 데이터를 사용하여 분석하는 경험을 제공함으로써 데이터가 어떤 맥락에서 생성되고 수집되었는지 직접 경험하게 된다. 따라서 학생들은 데이터의 맥락을 보다 깊이 있게 이해할 수 있었다. 둘째, 자료의 변이성에 대한 이해도가 높아졌다. 인공지능을 활용하여 다양한 데이터 셋을 탐색하고 분석하는 과정에서 데이터가 어떻게 분포되어 있는지 이해하고 시각화 도구를 이용하여 데이터의 분포와 패턴을 쉽게 파악함으로써 자료의 변이성에 대한 이해도가 높아졌다. 또한, 수집한 데이터를 전처리하는 과정에서 전처리된 데이터가 분석 결과에 미치는 영향을 이해하고, 다양한 결괏값을 만들어 낼 수 있음을 활동을 통해서 깨달았다. 셋째, 데이터 분석 활동에서 비판적 사

고력을 발휘했다. 자료 수집 및 전처리 과정의 필요성을 이해하고, 자료의 신뢰성을 검증하는 과정에서 비판적인 태도를 보였다. 결과에 영향을 미치는 다양한 변수를 고려하여 문제를 분석함으로써 데이터를 비판적으로 분석하는 능력을 키울 수 있었다. 따라서, Lee 외 (2023)가 제안한 빅데이터 프로젝트 과제 가이드라인을 바탕으로 설계한 통계교육 프로그램은 고등학생들의 통계적 소양에 유의미한 영향을 주었으며, 학생들은 빅데이터를 전처리하여 사용하는 경험을 통해서 빅데이터에 대한 올바른 이해를 갖게 되었다.

딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생들의 컴퓨팅 사고력에 미치는 영향을 분석하고자 대응 표본 t-검정 시행하였고 분석 결과, 수업 전과 후의 컴퓨팅 사고력($t=-11.657, p<0.001$)은 통계적으로 유의한 차이가 나타났으며 딥러닝 예측을 이용한 통계적 문제해결 과정 중심의 통계교육 프로그램은 학생들의 컴퓨팅 사고력을 높이는 긍정적인 영향을 주었음을 알 수 있다. 따라서, Ryu와 Han (2022)의 디지털 기반의 창의협력지능(CCI) 교수학습 모형은 고등학생을 대상으로한 수업으로 일반화할 수 있음을 확인할 수 있다.

이와 같은 연구 결과를 기반으로 결론을 도출하면 다음과 같다.

첫째, 현실의 다양한 문제를 다루기 위해, 2022 개정 수학과 교육과정의 ‘공통수학1’의 행렬 단원에 ‘다변수 함수’를 선형함수의 형태로 도입을 검토할 수 있다. <사례 1>에서 학생들의 대화를 통해 알 수 있듯이 학생들은 현실 문제에서 영향을 주는 요인(독립변수)과 영향을 받는 요인(종속변수)를 구분하는 데 어려움을 겪고 있다. 실생활의 다양한 문제를 해결하기 위해서는 다양한 변수를 스스로 설정하고, 변수 간의 상호작용을 이해하는 것이 필요하다. 이를 위해 다변수 함수를 선형함수로 다루는 접근이 필요하다. 다변수 함수를 행렬과 연결 지어 선형함수로 다루는 방식은 학생들이 현실의 문제를 보다 잘 이해하고 해결하는 데 도움이 될 수 있다. 둘째, 고등학교 통계 수업에서 가공되지 않은 원자료(raw data)를 다루는 것이 필요하다. 2015 개정 수학과 교육과정에 따르면, 확률과 통계에서는 자료를 수집하고 정리하여 결과를 분석하고 추정하는 것으로 통계를 정의한다(Ministry of Education, 2015). 그러나 확률과 통계에서는 확률분포와 통계적 추정을 지식 및 이해 중심으로 다루며 통계적 문제해결 과정은 다루지 않는다. <면담 1>에서 알 수 있듯이, 학생들은 자료를 수집하고 원자료를 데이터 분석 목적과 방법에 맞는 형태로 처리하는 과정을 통해 통계 자료를 다양하게 경험하고 자료 수집 및 전처리 과정의 필요성을 이해했으며, 자료의 신뢰성을 검증하는 과정에서 비판적인 태도를 기를 수 있었다. 셋째, 고등학교의 수학교육에서 코딩과 연계한 수학 내용을 구성할 필요가 있다(Lee et al., 2022). 연구 수업에 참여하는 모든 학생은 파이썬을 처음 배웠으며 기본적인 코딩 지식도 부족했다. 해석과 대수의 내용을 알고리즘 관점에서 해결하는 활동을 제공함으로써, 학생들의 수학적 사고력 및 컴퓨팅 사고력을 키울 수 있다. 예를 들어, ‘for 반복문’을 이용하여 등차수열의 합을 구하는 알고리즘 문제를 ‘수열 단원’에 포함해 코딩과 연계한 수학 내용을 지도를 고려할 수 있다.

```

1 def sum_of_arithmetic_sequence(a, d, n):
2     sum=0
3     for i in range(n):
4         sum+= a+i * d
5     return sum
6 sum_example = sum_of_arithmetic_sequence(1, 2, 5)
7 print('첫째항이 1, 공차가 2, 등차수열의 첫째항부터 제5항까지의 합:', sum_example)

```

첫째항이 1, 공차가 2, 등차수열의 첫째항부터 제5항까지의 합: 25

2. 제언

본 연구 결과를 바탕으로 제언을 제시하면 다음과 같다.

첫째, 고등학교 수학에서 딥러닝을 이용한 교수학습 방법에 관한 추가 연구가 필요하다. 선행연구를 분석한 결과, 초등학교 수학교육에서는 인공지능 및 딥러닝을 활용한 교수학습 방법에 관한 다양한 연구 결과가 있었지만, 중등학교 수학교육에서는 연구가 부족한 것으로 나타났다. 초등학교 수학교육에서는 인공지능과 딥러닝에 대한 기초적인 개념을 소개하고, 중등학교 수학교육에서는 더 깊은 이해와 응용을 강조하는 교육활동으로 연계되어야 한다. 둘째, 다양한 맥락에서 비판적으로 의문을 제기하는 태도를 형성할 수 있는 교수학습 방법에 관한 연구가 필요하다. 통계적 문제해결 과정의 결과 해석단계에서 통계적 소양 평가는 중 수준이 가장 많이 나타났으며, 학생들이 비판적으로 의문을 제기하는 태도는 거의 보이지 않았다. 통계적 정보와 데이터를 다양한 맥락에서 마주할 때, 비판적으로 의문을 제기하는 태도가 필요하다. 셋째, 프로그래밍의 어려움을 극복할 수 있

는 교수학습 방법에 관한 연구가 필요하다. 본 연구 결과 고등학교 학생들의 프로그래밍에 관한 자신감이 낮다는 결과가 나타났다. 따라서 프로그래밍으로 현실의 문제를 해결하는 태도를 학생들에게 심어줄 수 있는 교수학습 방법에 관한 연구가 필요하다.

Endnote

¹⁾본 연구에 참여한 학생 12명을 T1-T12으로 나타낸다.

ORCID

Hyeseong Jin: <https://orcid.org/0009-0006-3544-3479>

Boeuk Suh: <https://orcid.org/0000-0003-3468-4319>

Conflict of Interest

The authors declare that they have no competing interests.

References

- Computer Science Teachers Association (CSTA) & International Society for Technology in Education (ISTE). (2011). *Computational thinking in K-12 education teacher resource (2nd ed.)*. ISTE. https://cdn.iste.org/www-root/2020-10/ISTE_CT_Teacher_Resources_2ed.pdf
- Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report pre-K-12*. American Statistical Association.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1-25. <https://doi.org/10.1111/j.1751-5823.2002.tb00336.x>
- Graham, A. (2006). *Developing thinking in statistics*. The Open University in Association with Paul Chapman Publishing.
- Han, S. K., Ryu, M. Y., & Kim, T. R. (2021). *Artificial intelligence education for al thinking*. Sungandang.
- Hong, S. J., & Choi, I. S. (2020). *Concept and application of artificial intelligence in school education*. KICE.
- Jin, H. S. (2024). *Study research on a statistics education program utilizing deep learning predictions in high school mathematics* [Master's thesis, Chungnam National University].
- Ju, M. K., Kim, S. Y., Bae, K. T., Jeong, H. S., & Jing, S. Y. (2018). An analysis of high school students' statistical literacy: students' achievement and difficulties of statistical inquiry. *School Mathematics*, 20(4), 661-683. <https://doi.org/10.29275/sm.2018.12.20.4.661>
- Kim, M. Y., & Kim, S. Y. (2018). Korean secondary students computational thinking based on problem solving. *The Journal of Learner-Centered Curriculum and Instruction*, 18(16), 807-830. <http://dx.doi.org/10.22251/jlcci.2018.18.16.807>
- Kim, S. Y., & Choi, M. K. (2022). AI-based educational platform analysis supporting personalized mathematics learning. *Communications of Mathematical Education*, 36(3), 417-438. <https://doi.org/10.7468/jksmee.2022.36.3.417>
- Ko, E. S., & Park, M. S. (2017). Pre-service elementary school teachers' statistical literacy related to statistical problem solving. *School Mathematics*, 19(3), 443-459.
- Kwon, O. N., Oh, S. J., Yoon, J. E., Lee, K. W., Shin, B. C., & Jung, W. (2023). Analyzing mathematical performances of ChatGPT: Focusing on the solution of national assessment of educational achievement and the college scholastic ability test. *Communications of Mathematical Education*, 37(2), 233-256. <https://doi.org/10.7468/jksmee.2023.37.2.233>
- Lee, J. H., Han, C. R., & Lim, W. (2023a). Guidelines for big data projects in artificial intelligence mathematics education. *The Mathematical Education*, 62(2), 289-302. <https://doi.org/10.7468/mathedu.2023.62.2.289>
- Lee, J. H., Lee, S. G., & Ham, Y. M. (2022). Case study on college calculus Education for vocational high school graduates with coding. *Communications of Mathematical Education*, 36(4), 611-626. <https://doi.org/10.7468/jksmee.2022.36.4.611>
- Lee, J. Y., & Rim, H. M. (2021). Analysis of <Probability and Statistics> textbooks on statistical problem-solving process and statistical literacy. *Journal of the Korean School Mathematics Society*, 24(2), 191-216. <http://doi.org/10.30807/>

ksms.2021.24.2.002

- Lee, S. G., Nam, Y., Lee, J. H., & Kim, U. K. (2023b). A study on teaching of convolution in engineering mathematics and artificial intelligence. *Communications of Mathematical Education*, 37(2), 277–297. <https://doi.org/10.7468/jksmee.2023.37.2.277>
- Lee, Y. J. (2023). An analysis of pre-service teachers' mathematics lesson design using ChatGPT. *Communications of Mathematical Education*, 37(3), 497–516. <https://doi.org/10.7468/jksmee.2023.37.3.497>
- Lee, Y. M., Han, C. L., & Lim, W. (2023c). Analysis of artificial intelligence mathematics textbooks: Vectors and matrices. *Communications of Mathematical Education*, 37(3), 443–465. <https://doi.org/10.7468/jksmee.2023.37.3.443>
- MacKay, R. J., & Oldford, W. (1994). *Statistics 231 Course Notes Full 1994*. University of Waterloo.
- Ministry of Education (2015). *Mathematics curriculum* (# 2015–74 supplement 8). Ministry of Education.
- Ministry of Education (2020). *Informatics curriculum* (# 2022–23 supplement 10). Ministry of Education.
- Ministry of Education (2022a). *The general overview of primary and secondary school curriculum* (# 2022–33 supplement 1). Ministry of Education.
- Ministry of Education (2022b). *Mathematics curriculum* (# 2022–33 supplement 8). Ministry of Education.
- Oh, S. J. (2023). Effective ChatGPT prompts in mathematical problem solving: Focusing on quadratic equations and quadratic functions. *Communications of Mathematical Education*, 37(3), 545–567. <https://doi.org/10.7468/jksmee.2023.37.3.545>
- Park, H. S. (2020). *Studying machine learning + Deep learning alone*. Hanbit.
- Park, H. Y., Son, B. E., & Ko, H. K. (2022). Study on the mathematics teaching and learning artificial intelligence platform analysis. *Communications of Mathematical Education*, 36(1), 1–21. <https://doi.org/10.7468/jksmee.2022.36.1.1>
- Park, J. S. (2023). Application of computational thinking based on moral machine for digital citizenship education. *Social Studies Education*, 62(1), 27–46. <https://doi.org/10.37561/sse.2023.3.62.1.27>
- Ryu, M. Y., & Han, S. K. (2022). Development and application of digital-based collaboration intelligence teaching and learning model. *The Journal of Education*, 42(2), 117–130. <https://doi.org/10.25020/je.2022.42.2.117>
- Sim, Y. H., Kim, J. H., & Kwon, M. S. (2023). Secondary mathematics teachers' perceptions on artificial intelligence (AI) for math and math for artificial intelligence (AI). *Communications of Mathematical Education*, 37(2), 159–181. <https://doi.org/10.7468/jksmee.2023.37.2.159>
- Watson, J. M., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, 2(2), 3–46.
- Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, 67(3), 223–248. <https://doi.org/10.1111/j.1751-5823.1999.tb00442.x>
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, 49(3), 33–35. <https://doi.org/10.1145/1118178.1118215>

고등학교 수학에서 딥러닝 예측을 이용한 통계교육 프로그램 연구

진혜성¹, 서보억^{2*}

¹대전둔산여자고등학교 교사

²충남대학교 교수

*교신저자: 서보억(eukeuk@cnu.ac.kr)

초 록

4차 산업혁명과 인공지능의 발전으로 교육 분야에서 많은 변화가 일어나고 있다. 특히, 인공지능을 기반으로 하는 교육의 중요성이 강조되고 있다. 이러한 흐름에 따라 본 연구에서는 고등학교 수학에서 딥러닝 예측을 이용한 통계교육 프로그램을 개발하고 이러한 통계적 문제해결 과정 중심의 통계교육 프로그램이 고등학생들의 통계적 소양 및 컴퓨팅 사고력에 미치는 영향을 고찰하고자 한다. 먼저, 본 연구에서는 고등학교 수학에 적용할 수 있는 딥러닝 예측을 이용한 통계교육 프로그램을 개발하였고, 이를 실제 수업상황에 적용하여 분석하였다. 분석 결과, 학생들은 자료가 어떤 맥락에서 생성되고 수집되었는지 경험함으로써 맥락에 대한 이해도가 향상되었으며, 다양한 데이터셋을 탐색하고 분석하는 과정에서 자료의 변이성에 대한 이해도가 높아졌고, 자료의 신뢰성을 검증하는 과정에서 자료를 비판적으로 분석하는 능력을 보였다. 통계교육 프로그램이 고등학생들의 컴퓨팅 사고력에 미치는 영향을 분석하고자 대응 표본 t-검정 시행하였고, 수업 전과 후의 컴퓨팅 사고력 ($t=-11.657, p<0.001$)은 통계적으로 유의한 차이가 있음을 확인하였다.

주요어 고등학교 수학교육, AI 융합교육, 통계교육, 딥러닝

