

컴퓨터교육학회 논문지 2026년 제29권 제7호
<https://doi.org/10.32431/kace.2026.29.7.012>



ICILS 2023 한국 중학생의 컴퓨팅 사고력 예측변수 탐색

Exploring Predictors of Computational Thinking among Korean Middle School Students in ICILS 2023

오준석[†] · 김승재^{††}

Junseok Oh[†] · Seungjae Kim^{††}

요약

본 연구는 ICILS 2023 한국 중학교 2학년 학생 3,723명의 자료를 활용하여 컴퓨팅 사고력을 예측하는 주요 변수를 탐색하였다. 학생 설문 123개 변수를 랜덤포레스트에 투입하고 SHAP을 적용하여 상위 20개 주요 예측변수를 선정하였다. 분석 결과, 수업 중 ICT 활용과 ICT 자기효능감 영역의 변수가 가장 큰 비중을 차지한 반면, 성별이나 부모 학력 등 배경 변수의 기여는 제한적이었다. SHAP 의존성 도표를 통해 변수별 경향성을 확인한 결과, 같은 영역 내에서도 활용 맥락과 인식 내용에 따라 컴퓨팅 사고력 예측에 대한 기여 방향이 정적 또는 부정적으로 상이하게 나타났다. 이는 컴퓨팅 사고력 함양을 위해 ICT 활용의 양적 확대보다 활용 맥락과 학습자의 자기효능감을 고려한 교육 설계가 필요함을 시사한다.

주제어 ICILS, 컴퓨팅사고력, 랜덤포레스트, SHAP, 예측변수

ABSTRACT

This study explored the key predictors of computational thinking (CT) among 3,723 Korean eighth-grade students using data from ICILS 2023. A total of 123 student questionnaire variables were entered into a random forest model, and SHAP was applied to identify the top 20 predictors. The results showed that variables related to ICT use in lessons and ICT self-efficacy accounted for the largest share, whereas background variables such as gender and parental education made limited contributions. SHAP dependence plots revealed that, even within the same domain, the direction of each variable's contribution to CT prediction differed depending on the context of use and the content of students' perceptions. These findings suggest that fostering CT requires instructional design that considers the context of ICT use and learners' self-efficacy rather than simply expanding the amount of ICT exposure.

Keywords ICILS, computational thinking, random forest, SHAP, Predictor variables

†정회원 영남대학교 대학원 교육학과 박사수로
 ††정회원 영남대학교 대학원 교육학과 박사수로
 (교신저자)

논문투고 2026년 05월 08일
 심사완료 2026년 06월 23일
 게재확정 2026년 06월 30일
 발행일자 2026년 07월 30일

1. 서론

현대 사회는 인공지능, 데이터, 소프트웨어 등의 디지털 기술의 발전으로 다양한 분야에서의 큰 변화를 계속해서 겪고 있다. 이러한 21세기 지식기반사회와 4차 산업혁명 시대의 도래는 기존 교육이 강조해 온 지식습득 중심 학습의 한계를 드러내며, 복잡한 문제를 새로운 방식으로 해결할 수 있는 창의적인 사고를 위한 교육이 강조되고 있다[1]. 이에 학교 교육에서도 단순히 디지털 도구를 활용하는 것에 멈추지 않고, 이를 통해 문제를 구조화하고 해결하는 고차적 사고 역량에 대한 관심이 커지고 있다. 이러한 맥락에서 복잡한 문제를 논리적이고 체계적으로 해결하는 사고 과정인 컴퓨팅 사고력(Computational Thinking, CT)은 디지털 시대 학습자가 갖추어야 할 핵심 역량 중 하나이다[2].

컴퓨팅 사고력은 컴퓨터 과학의 기본 개념을 활용하여 문제를 해결하고 시스템을 설계하며 인간 행동을 이해하는 사고 과정으로서[3], 추상화와 분해를 중심으로 복잡한 문제를 체계적으로 재구성·해결하는 역량을 의미하는 개념으로 논의되어 왔다. 국내 연구에서도 컴퓨팅 사고력을 단순한 프로그래밍 기술이 아니라 학습자가 다양한 상황에서 활용할 수 있는 보편적 사고 능력으로 규정하고 있으며[4], 이러한 컴퓨팅 사고력은 디지털 환경에서 복잡한 문제를 체계적으로 인식하고 해결 과정을 설계·검증하는 기반으로서, 학습자가 컴퓨팅 사고력을 어느 정도 갖추고 있는지 체계적으로 측정·비교하고 이를 함양할 수 있는 교육적 방안을 모색할 필요가 있다.

이러한 컴퓨팅 사고력을 국제적으로 측정·비교할 수 있는 대표적인 자료로는 국제교육성취도평가협회(International Association for the Evaluation of Educational Achievement; IEA)의 국제 컴퓨터·정보 소양 연구(International Computer and Information Literacy Study; ICILS)가 있다. ICILS는 학생들의 컴퓨터·정보 소양과 컴퓨팅 사고력을 국제적인 수준에서 객관적으로 측정하고, 관련 교육 현황을 분석하는 국제 비교 연구이며, 참여국 학생들의 컴퓨팅 사고력을 함양할 수 있도록 교육 정책 수립을 위한 데이터를 제공한다[5]. 특히 ICILS는 성취 측정에 그치지 않고 ICT 활용 경험, 디지털 태도, 학교 및 수업 맥락, 교육 경험 등 다양한 배경 변수를 함께 수집하기 때문에, 성취 수준과 관련된 교육적·맥락적 요인을 종합적으로 탐색할 수 있는 자료로서 의의를 지닌다. 우리나라에서는 ICILS 2023에 중학교 2학년 학생 3,723명이 참여하였으며, 컴퓨팅 사고력 평균은 537점으로 국제 평균 483점 대비 54점이 높았다. 그러나 이러한 평균 점수만으로는 어떤 학생 특성·경험·태도가 이러한 결과에 기여하였는지 설명하기 어려우므로, 학생 수준에서 컴퓨팅 사고력을 예측하는 변수를 체계적으로 탐색할 필요가 있다.

이를 탐색하는 것은 정책적 측면과 실천적 측면 모두에서 의의를 지닌다. 2015 개정 교육과정에서 정보 교과가 필수화되어 2018년부터 초·중학교에 단계적으로 적용된 이래

[6], ICILS 2023은 새 교육과정에서 소프트웨어 교육을 받은 학생들의 컴퓨팅 사고력을 측정한 첫 국제 비교 자료이다. 따라서 ICILS 2023 데이터를 분석하는 것을 통해 정보 교과 필수화 정책이 학생들의 컴퓨팅 사고력 성취로 어떻게 이어졌는지를 파악할 수 있다. 또한 학교 현장에서 컴퓨팅 사고력을 함양하기 위한 ICT 활동·태도·경험 중 어떤 요인이 실제로 성취와 관련되는지에 대한 실증적 근거는 충분히 축적되어 있지 않아, 다수 변수를 동시에 고려한 분석은 교사와 정책 입안자에게 실천적 시사점을 제공할 수 있다.

한국 중학생의 컴퓨팅 사고력에 영향을 미치는 변수를 탐색한 선행연구는 다양한 측면에서 이루어져 왔다. 학교에서의 교육 경험과 코딩 교육의 유형을 분석한 연구[4,5]는 이들이 컴퓨팅 사고력 성취와 관련된 변수임을 확인하였다. 또한 ICILS가 컴퓨팅 사고력과 함께 측정하는 컴퓨터·정보 소양은 ICT 관련 경험과 태도 등 공통된 변수의 영향을 받으므로, 컴퓨터·정보 소양을 대상으로 한 선행연구 역시 컴퓨팅 사고력을 이해하는 데 유의미한 시사점을 제공한다. ICT 자기효능감을 비롯한 심리적 변수를 다룬 연구[7]는 이를 핵심 예측 요인으로 제시하였으며, 한국 학생의 성취 맥락을 분석한 연구[8]는 높은 성취와 낮은 학교 내 ICT 활용 빈도라는 특수성에 대해 논의하였다. 이러한 선행연구는 한국 중학생의 컴퓨팅 사고력과 컴퓨터·정보 소양이 교육 경험, ICT 활용 경험, 심리적 변수 등과 연관되어 있음을 보여준다는 점에서 의의가 있으나, 대체로 이론적으로 선별된 소수의 변수를 중심으로 평균 차이 비교 또는 구조적 관계 검증에 초점을 두어, 다수의 잠재적 예측변수를 동시에 고려한 탐색은 이루어지지 못하였다. 또한 ICT 자기효능감이나 ICT 활용 경험이 컴퓨팅 사고력과 관련된다는 사실은 확인되었으나, 이러한 변수가 응답 수준에 따라 일관된 선형 관계를 보이는지, 혹은 특정 수준 이후에 관계의 방향이나 크기가 달라지는지에 대한 비선형적 경향성은 충분히 검토되지 않았다. 그리고 ICILS 자료를 활용한 국내 연구는 대부분 ICILS 2018에 기반하여 2015 개정 교육과정 적용 이후의 변화를 반영하지 못한다는 한계를 지닌다.

본 연구에서는 이러한 한계를 극복하기 위해 랜덤포레스트 기법을 활용하여 ICILS 2023의 한국 중학생 자료에 대한 설문 변수 전체를 동시에 고려하여 컴퓨팅 사고력을 예측하는 변수들이 전체 변수 체계 안에서 어떠한 상대적 우선순위를 갖는지 탐색하고, 이들의 응답 수준에 따라 컴퓨팅 사고력은 어떠한 기여 방향과 경향성을 보이는지 확인하고자 하였다. 랜덤포레스트는 변수 간의 비선형 관계나 상호작용을 탐지하기 유리한 머신러닝 기법으로, 복잡한 교육 데이터를 분석하는 데 효과적인 도구로 부상하고 있다[9]. 특히 랜덤포레스트는 변수 간 상관성이 높거나 변수가 많은 경우, 결측치가 있는 경우에도 안정적인 예측력을 제공하며, 각 변수의 상대적 중요도를 도출할 수 있다는 점에서 변수 선별과 예측력을 평가할 때 적합하다[10]. 이러한 특성은 전통적 통계 기법의 한계를 보완할 수 있어, 다수의 변수를 포함하는 ICILS 2023 자료를 분석하는 데 효과적이다.

이에 본 연구는 ICILS 2023 한국 중학교 2학년 학생 자료를 활용하여, 학생 설문 변수 전체를 기반으로 컴퓨팅 사고력을 예측하는 주요 변수를 탐색하고, 이들의 응답 수준에 따른 컴퓨팅 사고력의 변화를 분석하고자 한다. 이를 통해 다수의 변수가 컴퓨팅 사고력에 미치는 상대적 기여도와 비선형적 변화 양상을 동시에 확인하고, 컴퓨팅 사고력 함양을 위한 구체적인 교육적 시사점을 도출하고자 한다. 이에 따른 연구문제는 다음과 같다. 첫째, ICILS 2023 학생 설문 변수 전체를 고려했을 때, 한국 중학생의 컴퓨팅 사고력을 예측하는 주요 변수는 무엇인가? 둘째, 주요 예측변수의 응답 수준에 따라 컴퓨팅 사고력은 어떠한 기여 방향과 경향성을 보이는가?

2. 이론적 배경

2.1 컴퓨팅 사고력

2.1.1 컴퓨팅 사고력의 개념

컴퓨팅 사고력(Computational Thinking, CT)은 Papert의 구성주의적 교육철학에서 비롯되었다. Papert는 학습자가 Logo 프로그래밍 언어를 활용한 컴퓨터와의 상호작용을 통해 논리적·절차적 사고를 발달시킬 수 있다고 보았으며, 이러한 사고 과정이 프로그래밍 자체에 그치지 않고 문제해결의 보편적 역량으로 전이될 수 있음을 주장하였다[11]. 이후 Wing이 이러한 사고를 '컴퓨터 과학의 기본 개념을 활용하여 문제를 해결하고, 시스템을 설계하며, 인간 행동을 이해하는 사고 과정'의 컴퓨팅 사고력이라는 명칭으로 제시하면서, 컴퓨팅 사고력은 컴퓨터 과학자만의 전유물이 아닌 모든 학습자에게 필요한 보편적 역량으로 본격적인 관심을 받기 시작하였다[3]. Wing은 이후 컴퓨팅 사고력에 대한 정의를 '문제를 해결책을 정보 처리 주체가 효과적으로 수행할 수 있는 형태로 표현하는 사고 과정'으로 수정하며, 추상화(abstraction)를 컴퓨팅 사고력의 가장 핵심적인 사고 과정으로 제시하였다[12].

이와 같은 Wing의 제안 이후 컴퓨팅 사고력의 구성 요소를 체계화하려는 시도가 다양하게 이루어졌다. 국제 컴퓨터과학교육 단체인 ISTE와 CSTA는 컴퓨터가 해결할 수 있는 형태로 문제를 형식화하고, 데이터를 논리적으로 조직·분석하며, 모델과 시뮬레이션을 통해 추상화하고, 알고리즘적 사고로 해결을 자동화하는 과정 등을 포함하는 것으로 컴퓨팅 사고력의 운영 정의(operational definition)를 제시하였다[13]. Grover와 Pea는 컴퓨팅 사고력의 핵심 요소로 추상화와 패턴 일반화, 체계적 정보 처리, 기호 체계와 표상, 알고리즘적 사고 등을 제안하였다[14]. 다음으로 Shute, Sun과 Asbell-Clarke는 컴퓨팅 사고력의 핵심 구성 요소로, 분해(decomposition), 추상화(abstraction), 알고리즘(algorithm), 디버깅(debugging), 반복(iteration), 일반화(generalization)의 여섯 가지를 도출하고, 컴퓨팅 사

고력을 '문제를 알고리즘적으로 해결하는 데 필요한 사고 방식이자 실행 과정'으로 정의하였다[15].

국내에서는 교육부가 컴퓨팅 사고력을 '컴퓨터과학의 기본 개념과 원리를 바탕으로 실생활 및 다양한 학문 분야의 문제를 창의적으로 해결하는 능력'으로 정의하였다[6]. 이를 기반으로 박찬정과 현정석은 컴퓨팅 사고력을 추상화, 자동화, 창의융합의 세 영역으로 분류하였으며, 이러한 분류가 기존의 컴퓨터교육을 단순한 도구 활용에서 사고 역량 중심의 교육으로 전환하는 계기가 되었음을 주장하였다[16].

본 연구가 활용한 ICILS 2023에서는 컴퓨팅 사고력을 '실세계 문제에서 컴퓨터적 형식화에 적합한 측면을 인식하고, 그 문제에 대한 알고리즘적 해결책을 평가·개발할 수 있는 개인의 능력'으로 정의하고 있다[17]. ICILS의 컴퓨팅 사고력 구인은 문제 개념화(conceptualizing problems)와 해결방안 운영(operationalizing solutions)의 두 영역으로 구성된다. 문제 개념화는 해결책 개발에 앞서 문제를 이해하고 알고리즘적·체계적 사고를 적용할 수 있는 형태로 구조화하는 과정으로, 디지털 시스템의 이해, 문제의 형식화 및 분석, 관련 데이터의 수집과 표상을 포함한다. 해결방안 운영은 알고리즘과 프로그램의 계획·작성, 해결방안의 평가 및 개선을 포함한다[17]. 이처럼 ICILS의 컴퓨팅 사고력 정의는 선행 문헌의 다양한 정의 중에서 문제를 컴퓨터와 상호작용할 수 있도록 형식화하고 알고리즘적 해결책을 구현·검증하는 ISTE와 CSTA의 정의와 연결되어 있다.

2.1.2 컴퓨팅 사고력 관련 선행연구

컴퓨팅 사고력 관련 연구 중 ICILS 자료를 활용한 연구는 국내외에서 활발히 이루어져 왔다. 국제적으로는 Fraillon 등이 ICILS 2018 국제보고서에서 성별, 사회경제적 배경, ICT 경험 등이 컴퓨팅 사고력 성취에 유의한 영향을 미침을 보고하였으며, 한국 학생이 컴퓨팅 사고력 평가 참여국 가운데 최상위 수준의 성취를 보였음을 제시하였다[18].

다음으로 ICILS 2023이 수행됨에 따라 이를 활용한 연구도 다수 이루어지고 있다. Guan 등이 수행한 연구에서는 LightGBM과 SHAP 기법을 적용하여 컴퓨팅 사고력의 주요 예측변수를 탐색하였으며, 그 결과, 교육 기대 수준, 가정의 문해 환경, SNS 활용, 맥락별 ICT 사용 등이 비선형적 영향 패턴을 보임을 확인하였다[19]. Hsu와 Chen은 ICILS 2023 대만 학생 5,211명을 대상으로 XGBoost, 랜덤포레스트 등 7개 머신러닝 모델을 비교 적용하여 사회인구학적 변수, 학생 변수, 학교 변수가 컴퓨팅 사고력에 미치는 영향을 분석하였다[20]. 이들 연구는 컴퓨팅 사고력이 다수의 변수에 의해 비선형적으로 예측되며, 이를 파악하기 위해 다수의 변수를 분석에 활용할 수 있는 머신러닝 기법이 유용함을 보여준다.

국내에서는 전성균과 박상욱(2022)이 코딩 경험이 컴퓨팅 사고력 성취와 관련이 있으며, 블록 기반 코딩과 텍스트 기반 코딩 경험 모두 성취 향상과 연관되어 있음을 밝혔다. 또한 코딩 경험이 전혀 없는 집단 대비 2년 이상 경험 집단

의 정당률이 유의하게 높게 나타남을 파악하였다[5]. 박지민과 길혜지는 구조방정식 모형을 적용하여 ICT 자기효능감이 컴퓨터·정보 소양에 가장 큰 직접 효과를 가지며, 여가 목적 ICT 활용이 자기효능감을 매개로 성취에 간접적으로 기여하는 반면, 수업 중 ICT 과업수행은 소폭의 부적 직접효과를 보이는 것을 확인하였다[7]. 김성원과 이영준은 텍스트 코딩과 블록 코딩을 모두 경험한 집단이 가장 높은 컴퓨팅 사고력을 보이며 남학생이 여학생보다 유의하게 높은 성취를 나타냄을 제시하였다[4].

위와 같은 선행연구를 종합하면, 컴퓨팅 사고력은 ICT 자기효능감, 코딩 경험, ICT 활용 빈도 및 맥락, 사회경제적 배경 등 다양한 변수와 관련되어 있으며, 이들 변수 간의 관계는 단순히 선형 관계로 설명하기 어려운 비선형적이고 복합적인 양상을 보일 수 있음을 알 수 있다. 그러나 대다수의 기존 국내 연구는 ICILS 2018 자료를 활용하여 연구를 수행하였기에 2015 개정 교육과정 적용 이후의 변화를 반영하지 못하였고, 전통 통계 기법을 활용하여 소수의 변수 간 관계를 검증하는 데 초점을 두어 여러 ICT 관련 변수 및 배경 변수를 동시에 고려했을 때 어떤 변수가 상대적으로 더 중요한지와 다수의 잠재적 예측변수와 컴퓨팅 사고력 간 관계를 파악하기 어려웠다. 또한 국제적으로 ICILS 2023 자료를 활용한 머신러닝 기반 연구가 등장하고 있으나, 한국 학생을 대상으로 한 단독 분석은 찾아보기 어려웠다. 이에 본 연구는 ICILS 2023 한국 자료를 활용하여 학생 설문 변수 전체를 동시에 투입한 랜덤포레스트 분석을 통해 기존 연구에서 중요성이 확인된 ICT 자기효능감과 ICT 활용 경험이 2015 개정 교육과정 적용 이후의 ICILS 2023 자료에서도 여전히 주요 예측변수로 나타나는지 확인하고 SHAP을 통해 각 변수의 기여 방향과 비선형적 경향성을 확인하고자 한다.

2.2 랜덤포레스트

랜덤포레스트(Random Forest)는 Breiman이 제안한 의사결정나무 기반의 앙상블 학습 기법이다[9]. 구체적으로, 전체 데이터에서 부트스트랩 표본을 반복하여 추출한 뒤 다수의 의사결정나무를 독립적으로 생성하고, 개별 의사결정나무의 예측값의 평균을 계산하여 최종 결과를 산출한다. 개별 나무를 생성할 때 나무의 분할 지점에서 전체 변수 중 무작위로 선택된 일부 변수만을 후보로 고려하므로 나무 간 상관성이 감소하여 앙상블의 일반화 성능이 향상된다[9,10].

랜덤포레스트는 전통적인 회귀분석에 비해 변수 간 비선형 관계나 상호작용을 별도의 모형 설정 없이 포착할 수 있고, 다중공선성에 영향을 덜 받기 때문에 상관성이 높은 변수가 다수 포함되어 있어도 안정적으로 작동한다[10]. 또한 대량의 변수를 이론적 배경에 따른 선별 없이 동시에 투입할 수 있으며, 변수 중요도(variable importance) 지표를 통해 각 변수의 상대적 기여도를 산출할 수 있어, 다수의 후보 변수 가운데 주요 예측변수를 탐색하는 데 적합하다[9,10]. 다만, 랜덤포레스트는 다수의 나무를 결합하여 예측을 수행하는 특성상 개별 변수가 예측에 미치는 영향의 경향성을 직

접 파악하기 어렵다는 한계를 지니고 있다.

2.3 SHAP

SHAP(SHapley Additive exPlanations)은 Lundberg와 Lee가 협력적 게임이론의 Shapley 값을 머신러닝 모형 해석에 적용하여 제안한 기법이다[21]. Shapley 값은 본래 협력적 게임에서 각 참여자의 기여도에 따라 보상을 공정하게 배분하기 위해 고안된 개념으로, SHAP은 이를 예측 모형에 적용하여 각 변수를 게임의 참여자로, 예측값을 보상으로 간주한다. 구체적으로, 특정 변수의 기여도를 산출하기 위해 해당 변수를 제외한 모든 변수들이 모형에 들어가는 모든 경우의 수를 따져본 뒤, 각 경우별로 특정 변수가 모형에 포함될 때와 포함되지 않을 때의 예측값 차이를 평균함으로써 해당 변수의 기여도를 산출한다. 이를 통해 변수 간 상호작용이 있는 경우에도 각 변수의 기여가 공정하게 배분된다[21].

이렇게 산출된 개별 관측치별 SHAP 값을 전체 데이터에 대해 변수별로 절댓값의 평균을 구하면 변수의 상대적 중요도를 산출할 수 있으며, 이는 앞서 언급한 랜덤포레스트의 변수 중요도가 제시하지 못하는 기여의 방향 정보를 함께 제공한다. SHAP 요약 도표(summary plot)는 변수별 중요도 순위와 기여 방향을 동시에 제시하여 주요 변수의 전반적인 영향 패턴을 파악하는 데 활용되며, SHAP 의존성 도표(dependence plot)는 특정 변수의 값에 따라 예측 기여분이 어떻게 달라지는지를 시각화하여 변수와 종속변수 간의 비선형적 관계나 임계점을 파악하는 데 유용하다[21].

랜덤포레스트와 SHAP의 결합은 교육 분야에서 다수의 변수를 동시에 고려하여 주요 예측변수를 탐색하는 연구에 다수 적용되어 왔다. 김진희와 김준엽은 중·고등학생의 교사-학생관계에 영향을 미치는 변수를 랜덤포레스트로 탐색하고 SHAP을 통해 해석하였으며[22], 장은아와 정혜원은 고등학생의 학업무기력 예측변수를 랜덤포레스트와 SHAP으로 분석하였다[23]. 백예은과 정혜원은 혼합효과 랜덤포레스트에 SHAP을 결합하여 청소년기 진로성숙도 발달의 주요 예측변수를 탐색하였다[24]. 랜덤포레스트와 SHAP의 결합은 다수의 변수가 포함된 교육 데이터에서 변수의 상대적 중요도를 파악하고 그 영향의 경향성을 해석하는 데 유용한 접근으로 자리 잡고 있다.

3. 연구 방법

3.1 연구대상

본 연구는 대한민국 중학생들의 컴퓨팅 사고력에 영향을 미치는 요인을 탐색하기 위해 IEA(국제교육성취도평가협회)가 실시한 ICILS(International Computer and Information Literacy Study) 2023의 한국 데이터를 활용하였다. ICILS는 중학교 2학년 학생을 대상으로 컴퓨터·정보 소양(CIL)과 컴퓨팅 사고력(CT)을 측정하는 국제 비교

연구로, 2013년에 처음 시행된 이래 5년 주기로 조사가 이루어지고 있다. 2018년 2주기 조사에서 처음으로 컴퓨팅 사고력 영역이 포함되었으며, ICILS 2023은 세 번째 주기에 해당한다. ICILS 2023에는 전 세계 34개 교육 체제가 참여하였으며, 이 중 24개 교육 체제가 컴퓨팅 사고력 평가에 참여하였다. ICILS의 표집 방식은 학교를 먼저 선정한 후 해당 학교 내에서 학급 및 학생을 추출하는 2단계 층화집락 표집 방식으로 이루어졌다. 본 연구의 분석 대상은 ICILS 2023에 참여한 한국 중학교 2학년 학생 3,723명이다.

3.2 측정 도구

3.2.1 종속변수

본 연구의 종속변수는 ICILS 2023에서 측정한 컴퓨팅 사고력(CT) 점수이다. ICILS 2023에서 컴퓨팅 사고력을 ‘실세계 문제에서 컴퓨터적 형식화에 적합한 측면을 인식하고, 그 문제에 대한 알고리즘적 해결책을 평가·개발할 수 있는 개인의 능력’으로 정의하고 있으며[17], 문제 개념화(conceptualizing problems)와 해결방안 운영화(operationalizing solutions)의 두 영역으로 구성하고 있다.

ICILS 2023에서는 학생의 컴퓨팅 사고력을 단일 점수가 아닌 5개의 추정값(plausible values)으로 제공한다[25]. 이는 제한된 수의 문항으로 학생의 능력을 측정할 때 발생하는 불확실성을 반영하기 위한 것으로, 각 추정값은 해당 학생이 가질 수 있는 능력 추정 분포에서 무작위로 추출된 값이다. 분석 대상의 컴퓨팅 사고력은 평균 535.4점($SD=121.3$)이었으며, 본 연구에서는 5개의 추정값을 평균하여 분석에 활용하였다.

3.2.2 예측변수

본 연구의 예측변수는 ICILS 2023 학생 설문지의 원문항 125개에 IEA가 제공하는 파생변수 4개(학생 나이, 이민 배경, 부모 최고 학력, 부모 최고 직업지위)를 더한 총 129개를 후보로 하였다. ICILS 학생 설문지는 학생의 인구통계적 배경, 가정환경, ICT 사용 빈도, ICT 학습 경험, ICT 자기효능감, ICT에 대한 인식 및 태도, CT 관련 학습 경험 등을 포괄하며, 응답 형식은 명목형, 서열형, 리커트 척도 등으로 구성된다. 이 중 부모 1과 부모 2의 정보를 각각 측정한 6개 문항(부모 1·2의 출신 국가, 최고 학력, 최고 직업지위)은 응답자별로 부모 1과 2가 가리키는 대상이 일관되지 않아 분석에서 제외하고, 부모 중 더 높은 수치를 나타내는 IEA 파생변수(이민 배경, 부모 최고 학력, 부모 최고 직업지위)로 대체하였다. 그 외에 컴퓨터 사용 경력, CT 학습 경험 등 개별 문항의 의미가 명확히 구분되는 문항은 문항 자체를 하나의 변수로 활용하여 분석하였다. 이를 통해 최종적으로 123개의 변수가 분석에 활용되었으며, 영역별 구성은 Table 1과 같다.

Table 1. Predictor Variables in the Random Forest Model

Domain	Variables
Personal Characteristics (5)	Sex, Expected highest education level, Country of birth, Language spoken at home, Age
Home Environment (11)	Highest parental education (ISCED), Highest parental occupational status (ISEI), Immigration status, Number of books at home, Home internet quality (2 items), Number of ICT devices at home (2 items), Access to ICT devices for schoolwork (3 items)
ICT Use Frequency (9)	Computer experience, ICT use frequency by context (6 items), Parental screen time limits (2 items)
ICT Learning Experience (24)	Internet-related tasks learned at school (6 items) and outside school (6 items), ICT skills learned at school (6 items) and outside school (6 items)
ICT Safety & Digital Citizenship (5)	Topics learned at school about responsible ICT use (5 items)
Out-of-School ICT Activities (6)	Out-of-school activities such as social media, Internet search, and video/music (6 items)
ICT Use in Lessons (22)	ICT use by subject (9 items), ICT tools used in lessons (13 items)
ICT Self-Efficacy (13)	Confidence in ICT tasks such as editing, searching, programming, and installing (13 items)
ICT Perceptions & Attitudes (18)	Perceptions of ICT and society (8 items), perceptions of ICT and learning/career (10 items)
CT Learning Experience (9)	CT-related learning tasks such as decomposing problems, detecting patterns, and testing for bugs (9 items)
ICT Course Enrollment (1)	Whether studying information technology/computer science

3.3 분석 절차

본 연구는 다음과 같은 절차로 분석을 수행하였다. 먼저, 결측치는 Stekhoven과 Bühlmann이 제안한 MissForest 기법을 활용하여 대체하였다[26]. MissForest는 랜덤포레스트를 기반으로 변수 간 관계를 반영하여 결측치를 반복적으로 추정하는 방법으로, 본 연구에서 수행하고자 하는 랜덤포레스트와 동일한 알고리즘에 기반하여 결측치를 처리할 수 있다는 장점을 지닌다. 분석에 투입된 123개 변수(서열형 121개, 연속형 2개)의 결측률은 평균 2.5%였으며, 부모 최고 직업지위(20.3%)를 제외한 모든 변수에서 10% 미만이었다. 서열형 변수의 대체값은 정수로 반올림하였으며, 대체 전후 변수별 평균과 표준편차의 차이는 각각 평균 .009, .018로 나타나, 결측치 대체로 인한 분포의 왜곡은 나타나지 않았다. 다음으로, SHAP 의존성 도표의 해석 방향을 통일하기 위해 응답 수준이 높을수록 부정적 응답을 나타내는 변수(ICT 자기효능감, ICT 인식-태도, ICT 학습 경험, CT 관련 학습 영역 69개 문항)에 대해 역코딩을 실시하여 높은 값이 긍정적 응답을 나타내도록 변환하였다. 이후, 표집 설계에 따른 학생별 표집 확률의 차이를 보정하기 위해

ICILS에서 제공하는 학생 총 가중치(TOTWGTs)를 랜덤포레스트 모형 학습과 교차검증 과정에 표본 가중치(sample_weight)로 활용하였으며, SHAP 값은 가중치가 반영된 최종 모형으로부터 산출하였다.

다음으로, 전체 데이터를 훈련용(70%)과 검증용(30%)으로 분할한 후, 훈련 데이터를 활용하여 모형을 구축하였다. 랜덤포레스트는 개별 의사결정나무 생성 시 각 분할 지점에서 전체 변수가 아닌 무작위로 선택된 일부 변수만을 후보로 고려하는데, 이때 후보 변수의 수(max_features)는 Breiman의 제안에 따라 전체 독립변수 수의 1/3인 41개로 설정하였다[9]. 이외에 의사결정나무 수(n_estimators), 최대 깊이(max_depth) 등의 하이퍼파라미터는 GridSearchCV를 활용하여 5-fold 교차검증 기반으로 최적값을 탐색하였으며, 그 결과 의사결정나무 수(n_estimators) 500개, 최대 깊이(max_depth) 15, 리프 노드 최소 표본 수(min_samples_leaf) 10으로 설정하였을 때 평균제곱근오차(Root Mean Square Error)가 가장 낮게 도출되는 것을 확인하여, 이를 최종 모형으로 선정하였다. 최종 모형의 검증용 자료에 대한 결정계수(R^2)는 .381, 평균 제곱근오차(RMSE)는 90.1로 나타났다. OOB(out-of-bag) 결정계수도 .374로 검증용 결정계수와 거의 일치하여 모형의 일반화 안정성이 확인되었다. 그러나 이러한 설명력은 예측의 정확성 자체를 극대화하기에 높은 수준이라고 보기는 어려우나, 이는 점수의 예측보다 주요 예측변수의 탐색에 목적을 둔 본 연구의 분석 설계와 학생 설문 문항만을 예측변수로 활용한 자료의 특성을 반영하는 것이다. 선행연구에서는 전체 투입 변수의 약 5%~20%를 주요 변수로 제시한 바 있다[22,23]. 이를 참고하여 본 연구에서는 전체 123개 변수 중 SHAP 중요도 지수 상위 20개(약 16%)를 주요 예측변수로 선정하였으며, 선정된 20개 변수의 SHAP 중요도 합은 전체 변수 중요도 합의 63.3%에 해당하였다. 이후, 도출된 주요 예측변수에 대하여 SHAP 의존성 도표를 통해 주요 예측변수의 응답 수준에 따른 컴퓨팅 사고력의 경향성을 살펴보았다.

앞서 기술한 바와 같이 본 연구에서는 5개 추정값의 평균을 종속변수로 사용하였다. 이는 각 추정값을 개별적으로 분석한 후 결과를 합산하는 IEA의 권고[27]와 다른 방식이나, 평균 처리의 편향이 집단 수준의 분산과 표준오차의 추정에서만 나타난다는 점에서[25] 모집단 수준의 표준오차 추정이나 통계적 유의성 검정이 아닌, 변수 중요도 순위의 탐색을 목적으로 하는 본 연구에 적용 가능하다고 판단하였다. ICILS 2023 데이터에 머신러닝을 적용한 선행연구에서도 5개 추정값의 평균이 종속변수로 사용된 바 있다[19]. 다만 이러한 판단을 실증적으로 검토하기 위하여 동일한 연구 절차로 종속변수를 5개의 추정값으로 바꾸어 가며 5회의 분석을 수행하였다. 본 분석에서 선정된 상위 20개 변수 중 17~19개는 개별 분석에서도 상위 20위 이내에 포함되었다. 또한 5회의 분석에서 도출된 중요도를 합산한 순위 기준으로 상위 20개 변수의 구성이 본 분석과 완전히 일치하

였으며, 두 순위 간 Spearman 순위상관계수는 .99로 나타났다(Table 2). 개별 추정값의 순위상관은 모두 .97~.98 수준으로 나타나 평균값 기준 순위와 큰 차이를 보이지 않았다. 가장 중요한 변수인 '개념맵 소프트웨어 사용'은 모든 추정값에서 1위를 유지하였으며, 나머지 변수들도 대체로 4~6위 정도의 순위 변동에 그쳤다. 또한 변수 중요도 값 역시 약 15% 안팎의 변동을 보였다. 따라서 본 연구에서 사용한 추정값의 평균을 종속변수로 활용하는 것이 주요 예측변수 선정에 미치는 영향이 크지 않음을 확인하였다.

Table 2. Stability of the top 20 predictors across the five plausible-value analyses
(○ = within the top 20 × = not within the top 20)

Top 20 Variable	PV1	PV2	PV3	PV4	PV5	Combined
Concept mapping SW	○	○	○	○	○	○
Image insertion	○	○	○	○	○	○
Computer-based info. resources	○	○	○	○	○	○
Non-school day, non-academic	○	○	○	○	○	○
Graphic design SW	○	○	○	○	○	○
SNS posting/viewing	○	○	○	○	○	○
Fewer jobs due to ICT	○	○	○	○	○	○
Program/app installation	○	○	○	○	○	○
Expected education level	○	○	○	○	○	○
IT subject ICT use	○	○	○	○	○	○
Importance of ICT changes	×	○	○	○	○	○
Internet search	○	○	×	○	○	○
Bug testing	○	○	○	○	○	○
Parental occupational status	○	○	○	○	○	○
Video conferencing	○	×	○	○	○	○
Internet search (outside school)	○	○	○	○	○	○
Text-based programming (at school)	○	○	○	○	×	○
School day, non-academic	×	○	○	○	○	○
Visual coding	×	×	○	×	○	○
Simulation/modelling SW	○	○	○	×	○	○

3.4 분석 방법

3.4.1 랜덤포레스트 회귀

본 연구에서는 국내 중학생의 컴퓨팅 사고력을 예측하는 주요 예측변수를 탐색하기 위해 랜덤포레스트 회귀분석을 실시하였다.

선행연구에서 ICILS 2023 데이터를 다양한 머신러닝 모형으로 비교분석한 결과 트리 기반 앙상블 모형들은 로지스틱 회귀 대비 성능이 우수하며 모형 간 성능 차이는 크지 않은 것으로 확인되었는데[20], 본 연구는 예측 정확도의 극대화보다 변수 탐색과 해석을 목적으로 한다. 이에 트리 기반 앙상블 모형인 랜덤포레스트, XGBoost, LightGBM의 예측 성능과 주요 예측변수의 안정성을 비교하였다(Table 3). 검증용 결정계수(Test R^2)는 부스팅 모형이 다소 높았으나 Train R^2 와의 격차가 랜덤포레스트보다 크게 나타나 상대적으로 과적합 경향을 보였으며, SHAP 중요도 상위 20개 변수는 모형 간 16개 이상이 일치하였다. 따라서 과적합에 보수적이고 변수 중요도 산출이 안정적인 랜덤포레스트를 채택하였다.

Table 3. Predictive performance of tree-based ensemble models

Model	Train R^2	Test R^2	Train - Test gap
Random Forest	.603	.381	.22
XGBoost	.853	.424	.43
LightGBM	.962	.412	.55

분석은 Python 3.9.13 환경에서 scikit-learn 라이브러리의 RandomForestRegressor와 shap 라이브러리를 활용하였다.

3.4.2 SHAP

본 연구에서는 랜덤포레스트 회귀로 도출한 주요 예측변수가 컴퓨팅 사고력 예측에 기여하는 방향과 크기를 파악하기 위해 SHAP 기법을 활용하였다[21]. 구체적으로, 트리 기반 앙상블 모형에 적합한 TreeExplainer를 사용하여 검증용 자료의 각 학생에 대한 SHAP 값을 산출하였으며, 주요 예측변수에 대해서는 SHAP 의존성 도표를 작성하여 응답 수준에 따른 컴퓨팅 사고력의 기여 방향과 비선형적 경향성을 분석하였다.

4. 연구결과

4.1 컴퓨팅 사고력을 예측하는 주요 변수

본 연구에서는 랜덤포레스트 분석을 통해 컴퓨팅 사고력에 대한 예측변수의 중요도 지수를 산출하였으며, 상위 20개 주요 예측변수를 Table 4에 제시하였다. 이를 구체적으로 살펴보면, 우선 수업 중 ICT 활용 영역에서는 개념맵 소프트웨어, 컴퓨터 기반 정보자료, 화상회의 시스템, 그래픽 디자인 소프트웨어, 시뮬레이션/모델링 소프트웨어 등 수업에서 활용하는 다양한 소프트웨어의 사용 빈도와 정보(컴퓨터) 교과 수업에서의 ICT 사용 빈도가 주요 예측변수로 나타나, 상위 20개 변수 중 가장 많은 6개 변수가 이 영역에 포함되었다.

다음으로, ICT 자기효능감 영역에서는 문서에 이미지를 삽입하거나 프로그램 혹은 앱을 설치하고, 학교 과제를 수행하는 데 있어 인터넷 정보를 검색하는 데 느끼는 자신감과 비주얼코딩에 대한 자신감 변수가 주요 예측변수로 도출되었다. ICT 인식·태도 영역에서는 ICT의 사회적 영향에 대한 인식을 측정하는 "ICT로 인해 일자리가 감소할 것"이라는 인식과 ICT 변화에 대한 적극성을 측정하는 "ICT 변화를 따라가는 것이 중요하다"는 인식이 주요 예측변수로 나타났다.

또한, ICT 사용 빈도 영역에서는 비수업일 학교 밖에서의 비학업 목적 ICT 사용 빈도와 수업일 학교 내에서의 비학업 목적 ICT 사용 빈도가 주요 예측변수로 나타나, 학업과 직접 관련되지 않는 ICT 사용이 사용 맥락에 따라 컴퓨팅 사고력 예측에 기여하는 것으로 확인되었다. 학교 밖 ICT 활동 영역에서는 SNS에 글을 게시하거나 열람하는 빈도가 주요 예측변수로 도출되었다. ICT 학습 경험 영역에서는 학교 밖에서의 인터넷 정보검색 학습 경험과 학교에서의 텍스트 프로그래밍 학습 경험이 주요 예측변수로 나타났다. 컴퓨팅 사고력 관련 학습 영역에서는 프로그래밍의 검증과정에 해당하는 프로그램 버그/오류의 체계적 테스트 경험이 주요 예측변수로 나타났다.

이 외에 개인특성 영역에서는 학생이 기대하는 최종학력이, 가정환경 영역에서는 부모의 최고 직업지위가 각각 주요 예측변수로 도출되었다. 한편, 성별이나 부모 학력과 같은 인구통계 및 사회경제적 배경 변수는 상위 20개에 거의 포함되지 않았으며, 부모 최고 직업지위만이 14위에 위치하여 컴퓨팅 사고력 예측에서 사회경제적 배경의 상대적 기여가 제한적임이 나타났다.

4.2 주요 예측변수의 응답 수준에 따른 컴퓨팅 사고력의 경향성

주요 예측변수의 응답 수준에 따른 컴퓨팅 사고력의 경향성을 파악하기 위하여 상위 20개 변수의 SHAP 지수의 분포를 SHAP 의존성 도표를 통해 확인하였다(Fig 1, 2).

구체적으로 SHAP 의존성 도표의 Y축은 SHAP 값을 나타내며 X축은 해당 변수의 응답 수준을 나타낸다. 예컨대, '이미지삽입 자신감' 변수의 경우 응답 수준이 3 이하일 때 SHAP 지수가 음수로 나타나고, 3보다 높을 때 SHAP 값이 양수로 나타나 응답수준 4일 때 SHAP 지수가 높게 나타났다. 이는 '이미지삽입 자신감' 변수에 대한 응답수준이 3 이하일 때는 컴퓨팅 사고력 예측에 부적 기여를 하는 반면, 응답수준 4일 때는 컴퓨팅 사고력 예측에 정적 기여를 하여 예측값이 평균 대비 더 높게 나타남을 의미하여, 해당 변수의 응답 수준이 높을수록 예측에 정적으로 기여하는 것으로 해석할 수 있다. 반면에 '개념맵 소프트웨어 사용' 변수의 경우 앞선 경우와 반대로 응답수준이 1일 때 SHAP 값이 양수이고 2이상일 때는 음수 형태를 보이고 있다. 이는 해당 변수가 컴퓨팅 사고력 예측에 음의 방향으로 기여한다는 것으로 해석할 수 있다.

Table 4. Key Predictors of Computational Thinking

	Variable Name	Variable Description	Importance	Rank
ICT Use in Lessons	Concept mapping SW	Frequency of concept mapping software use in lessons	18.61	1
	Computer-based info. resources	Frequency of computer-based information resources use in lessons	6.43	3
	Graphic design SW	Frequency of drawing and graphic design software use in lessons	5.50	5
	IT subject ICT use	Frequency of ICT use in information technology lessons	3.59	10
	Video conferencing	Frequency of video conferencing system use in lessons	2.35	15
	Simulation/modelling SW	Frequency of simulations and modelling software use in lesson	2.04	20
ICT Self-Efficacy	Image insertion	Confidence in inserting an image into a document or message	10.23	2
	Program/app installation	Confidence in installing a program or app	4.74	8
	Internet search	Confidence in searching for relevant information for a school project on the Internet	2.82	12
	Visual coding	Confidence in using visual coding	2.12	19
ICT Perceptions & Attitudes	Fewer jobs due to ICT	Agreement with "With more ICT there will be fewer jobs"	4.99	7
	Importance of ICT changes	Agreement with "It is important to keep up with changes in ICT"	3.08	11
ICT Use Frequency	Non-school day, non-academic	Frequency of ICT use outside school for other purposes on non-school days	5.55	4
	School day, non-academic	Frequency of ICT use at school for other purposes on school days	2.17	18
Out-of-School ICT Activities	SNS posting/viewing	Frequency of using social media to post or view content outside school	5.46	6
ICT Learning Experience	Internet search (outside school)	Extent of learning to use the Internet to find information outside school	2.33	16
	Text-based programming (at school)	Extent of learning to write computer programs using a text-based programming language at school	2.28	17
CT Learning Experience	Bug testing	Extent of learning to systematically test computer programs to find bugs and errors	2.73	13
Personal Characteristics	Expected education level	Highest level of education expected to complete	4.21	9
Home Environment	Parental occupational status	Highest International Socio-Economic Index (ISEI) of parents	2.58	14

각 영역별 결과는 다음과 같다. 우선 상위 20개 변수 중 가장 많은 6개가 포함된 ‘수업 중 ICT 활용’ 영역을 살펴보면, ‘개념맵 소프트웨어 사용’(1위), ‘그래픽디자인 소프트웨어 사용’(5위), ‘화상회의 시스템 사용’(15위), ‘시뮬레이션 소프트웨어 사용’(20위)은 컴퓨팅 사고력 예측에 부적으로 기여하는 것으로 확인되었다. 반면, ‘정보자료 사용’(3위)과 ‘정보교과 ICT 사용’(10위)은 컴퓨팅 사고력 예측에 정적으로 기여하고 있음이 나타났는데, 이를 통해 수업 중 ICT 활용 영역에서는 특정 소프트웨어의 개별적 사용과 교과 수업 맥락에서의 ICT 활용 간에 컴퓨팅 사고력에 대한 기여 방향이 상이하게 나타남을 확인할 수 있었다.

다음으로, ‘ICT 자기효능감’ 영역에서는 ‘이미지삽입 자신감’(2위), ‘프로그램설치 자신감’(8위), ‘정보검색 자신감’(12위), ‘비주얼코딩 자신감’(19위) 모두 응답 수준이 높아질수록 컴퓨팅 사고력 예측에 정적으로 기여하였다. 특히 ‘이미지삽입 자신감’은 응답 수준 1에서 약 -17, 응답 수준 4에서 약 +17로 주요 변수 중 가장 큰 변화폭을 보여 ICT에 대한 자기효능감이 컴퓨팅 사고력 예측에 중요하게 기여하는 것으로 나타났다.

‘ICT 인식·태도’ 영역에서는 ‘일자리감소 인식’(7위)과

‘ICT 변화 중요도 인식’(11위)이 주요 변수로 도출되었다. ICT의 사회적 영향에 대해 “일자리가 감소할 것”이라고 인식하는 것은 컴퓨팅 사고력 예측에 부적으로 기여하였으며, ICT 변화의 중요성에 대한 강한 인식은 컴퓨팅 사고력 예측에 정적으로 기여하여, 같은 ICT 인식·태도 영역 내에서도 인식의 내용에 따라 컴퓨팅 사고력에 대한 기여 방향이 상이하게 나타남을 확인할 수 있었다.

‘ICT 사용 빈도’ 영역에서는 ‘비수업일 비학업 ICT’ 빈도(4위)는 사용 빈도가 높은 구간에서 SHAP 값이 양의 방향으로 기여하는 반면, ‘수업일 비학업 ICT’ 빈도(18위)는 사용 빈도가 높아질수록 SHAP 값이 음의 방향으로 기여하였다. 이와 같이 동일한 비학업 목적 ICT 사용임에도 불구하고 학교 밖에서의 자발적 사용은 정적으로, 학교 내에서의 사용은 부적으로 기여하는 대조적 양상이 확인되었다.

이 외에 ‘학교 밖 ICT 활동’ 영역의 하위 요인인 ‘SNS 게시/열람’ 빈도(6위)는 컴퓨팅 사고력 예측에 부적으로 기여하였으며, ‘ICT 학습 경험’ 영역에서는 ‘학교 밖 정보검색 학습’(16위)이 컴퓨팅 사고력 예측에 정적으로 기여하는 반면, ‘학교에서의 텍스트 프로그래밍 학습 경험’(17위)은 컴퓨팅 사고력 예측에 부적으로 기여하여, 학습 내용에 따라 컴퓨

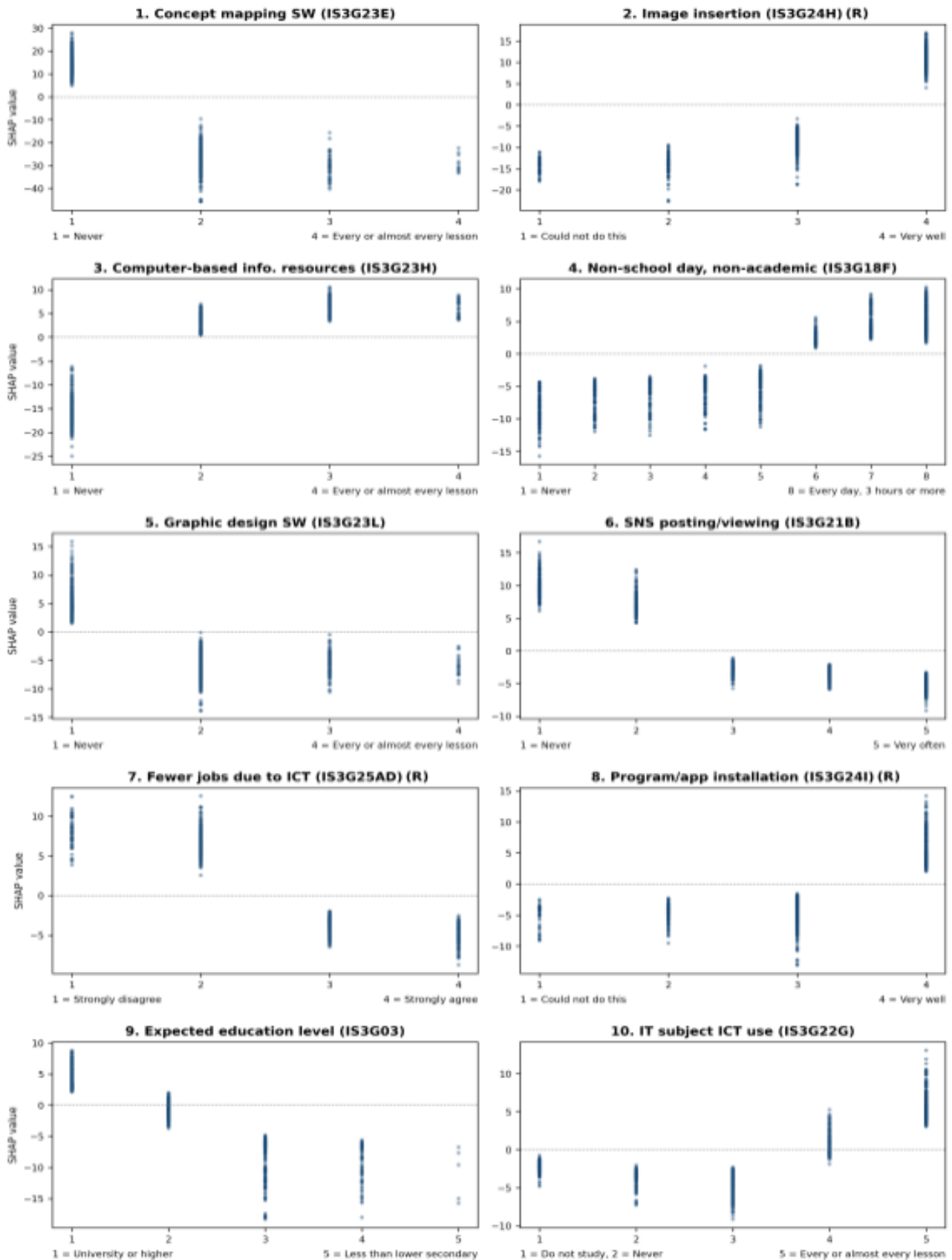


Figure 1. SHAP dependence plots for major predictor variables
(R) indicates reverse-coded items.

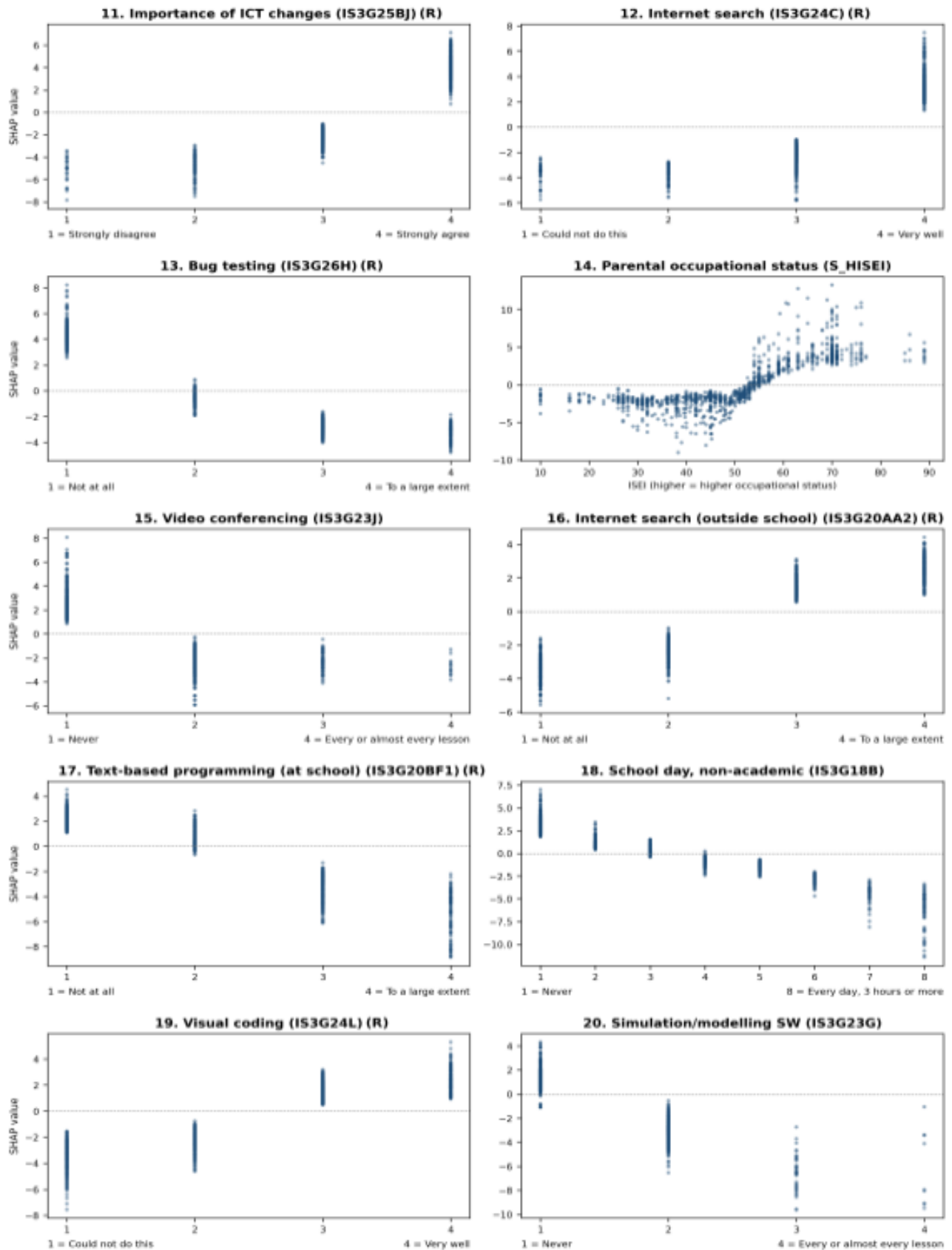


Figure 2. SHAP dependence plots for major predictor variables
(R) indicates reverse-coded items.

팅 사고력 예측에 대한 기여 방향이 상이하게 나타났다. ‘CT 관련 학습’ 영역의 ‘버그 테스트 경험’(13위)은 경험이 많아 질수록 SHAP 값이 음의 방향으로 기여하였다. ‘개인특성’ 영역의 ‘기대 최종학력’(9위)은 대학교 이상의 높은 학력을 기대할수록 SHAP 값이 높게 나타나, 높은 교육 포부를 지닌 학생일수록 컴퓨팅 사고력 예측에 정적으로 기여하였으며, ‘가정환경’ 영역의 ‘부모 최고 직업지위’(14위)는 ISEI 점수가 55 이상인 구간에서 SHAP 값이 급격히 상승하여 컴퓨팅 사고력 예측에 정적으로 기여하였다. 이상의 결과를 종합하면, 컴퓨팅 사고력 예측에는 ICT 활용 경험이나 자기효능감과 같은 학생의 ICT 관련 특성이 성별이나 사회경제적 배경보다 상대적으로 기여하는 바가 높은 것으로 나타났다.

5. 논의

본 연구의 목적은 ICILS 2023 한국 중학교 2학년 학생의 설문 변수 전체 자료를 활용하여 컴퓨팅 사고력을 예측하는 주요 변수를 탐색하고, 주요 예측변수와 컴퓨팅 사고력 간의 응답 수준에 따른 경향성을 분석하는 것에 있다. 분석 결과를 바탕으로 논의하면 다음과 같다.

첫째, 한국 중학생의 컴퓨팅 사고력 예측에는 학생의 ICT 관련 경험과 ICT 자기효능감이 핵심적으로 기여하는 것으로 나타났다. 주요 예측변수 상위 20개 가운데 수업 중 ICT 활용 영역이 가장 큰 비중을 차지하였고, ICT 자기효능감 관련 변수 역시 다수 포함되었다. 반면 성별, 나이와 같은 전통적 배경 변수는 거의 포함되지 않았으며, 부모 최고 직업지위 정도만이 제한적으로 포함되었다. 이는 한국 중학생의 컴퓨팅 사고력을 설명할 때 단순한 인구통계학적 배경보다 학생이 ICT를 실제로 어떻게 활용하고, 관련하여 스스로 수행할 수 있다고 인식하는지가 더 직접적으로 관련되어 있음을 의미한다. ICT 자기효능감이 컴퓨터·정보 소양에 가장 큰 영향을 미치고, ICT 활용과 수업에서의 ICT 과업 수행 역시 자기효능감을 매개로 성취에 영향을 미친다고 보고한 박지민과 길혜지[7]의 결과와도 부합한다. 즉, 컴퓨팅 사고력은 외부 환경에 의한 단순 노출보다는, 기술을 활용하는 과정에서 형성되는 수행 가능감과 실제 활용 경험의 질에 의해 더 크게 설명된다고 볼 수 있다.

둘째, ICT 활용과 ICT 인식·태도 모두에서 사용 맥락이나 인식 내용에 따라 컴퓨팅 사고력 예측에 대한 기여 방향이 달라지는 양상이 나타났다. 수업 중 ICT 활용에서는 컴퓨터 기반 정보자료 활용과 정보(컴퓨터) 교과 수업에서의 ICT 활용은 컴퓨팅 사고력 예측에 정적으로 기여하였으며, 개념맵·그래픽디자인·화상회의·시뮬레이션 소프트웨어 사용은 부적으로 기여하였으며, 비학업 목적 ICT 사용에서도 비수업일 학교 밖에서의 사용은 정적으로 기여, 수업일 학교 내에서의 사용과 SNS 게시·열람은 부적으로 기여하였다. 또한 ICT 인식·태도 영역에서 ICT 변화에 대한 적극적 인식은 정적으로 기여, ICT의 부정적 영향에 대한 인식은 부적

으로 기여하여, 인식의 내용에 따라 컴퓨팅 사고력 예측에 대한 기여 방향이 상반되게 나타났다.

이는 ICT를 다루는 빈도 자체가 아니라 학습 목표와 결합된 활용 경험과 학습자의 능동적 인식이 컴퓨팅 사고력 예측과 더 밀접하게 연관된다는 것으로 해석할 수 있다. 수업 중 ICT 활용에서 학습 정보 탐색이나 정보 교과처럼 학습 맥락과 통합된 활용은 컴퓨팅 사고력 예측에 양의 방향으로 기여하지만, 특정 소프트웨어를 단편적으로 다루는 경험은 같은 양상을 보이지 않을 수 있다. 또한 같은 비학업 ICT 사용이라도 자율적 시간에 다양한 디지털 환경에 능동적으로 노출되는 경험은 디지털 친숙도와 자기효능감 형성을 통해 컴퓨팅 사고력 예측에 정적으로 기여할 수 있는 반면, 수업 시간 내 학습 외적 ICT 사용이나 소비 중심의 SNS 활동은 학습 몰입을 저해하는 회피적 양상을 반영할 수 있다. ICT 인식 측면에서도 ICT를 미래 사회에서 따라가야 할 변화로 받아들이는 학습자는 디지털 환경에 능동적으로 임할 가능성이 높은 반면, ICT를 위협적 변화로 인식하는 학습자는 회피적 태도를 형성하여 컴퓨팅 사고력과 관련된 경험을 충분히 축적하지 못할 수 있다.

이는 ICT 활용의 효과가 활용 목적과 자기효능감, 과업 맥락에 따라 달라질 수 있음을 보여준 선행 연구와도 부합한다[7,28]. 따라서 학교 현장에서는 단순히 ICT 활용을 확대하는 것이 아니라, 어떠한 활동과 인식이 컴퓨팅 사고력과 직접 연결되는지를 고려하여 수업을 설계할 필요가 있다.

셋째, 학습 경험이 컴퓨팅 사고력 예측에 기여하는 양상은 단순한 학습활동 참여 여부보다 그 활동이 어떻게 조직되고, 학습자가 그것을 어떻게 받아들이는가에 따라 다르게 나타났다. 선행 연구에서는 코딩·프로그래밍 교육 경험, 정보 과목 이수 경험, 소프트웨어 교육 관련 학교 경험이 많을수록 컴퓨팅 사고력 성취가 높게 나타났으며, 코딩 교육 경험 기간이 길수록 컴퓨팅 사고력 점수가 높아지는 경향이 있음이 보고되었다[5,29]. 그러나 본 연구의 결과, 학교 밖 정보검색 경험은 컴퓨팅 사고력 예측에 정적으로 기여하는 반면, 학교에서의 텍스트 프로그래밍 학습 경험과 프로그램 버그·오류의 체계적 테스트 경험은 부적으로 기여하였다. 이 결과는 표면적으로는 기존 연구와 상반되어 보일 수 있으나, 실제로는 분석 방식과 해석 수준의 차이로 이해할 필요가 있다. 기존 연구가 주로 경험 여부나 이수 시간에 따른 평균 차이를 다뤘다면[30], 본 연구는 다수의 변수를 동시에 고려한 상태에서 각 변수의 상대적 기여 방향을 본 것이다. 따라서 프로그래밍 학습 경험이나 디버깅 경험이 부적으로 나타난 것은 해당 경험 자체의 부정적 효과라기보다, 어려운 과업에 대한 부담, 충분히 구조화되지 않은 교수·학습 경험, 혹은 아직 숙달되지 않은 상태에서의 어려움이 반영된 결과일 가능성이 있다. 이는 컴퓨팅 사고력 교육에서 단순히 프로그래밍 경험을 늘리는 것만으로는 충분하지 않으며, 학습자가 문제를 분해하고 해결 과정을 검토하며 오류를 수정하는 과정을 의미 있게 경험할 수 있도록 교수·학습의 질을 함께 보장해야 함을 의미한다.

넷째, 컴퓨팅 사고력 예측에는 사회경제적 배경보다 교육적으로 개입이 가능한 변수가 상대적으로 더 크게 기여하는 것으로 나타났다. 부모 직업지위가 주요 예측변수에 포함되긴 하였으나 상대적으로 낮은 순위에 머물렀고, 성별이나 부모 학력은 기여하는 정도가 미미하였다. 또한 부모 직업지위의 기여는 일정 수준 이상에서 두드러지는 비선형적 양상으로 나타나, 사회경제적 배경이 모든 학생에게 균등하게 작용하기보다 특정 임계점 이상에서 나타나는 특성을 보였다.

즉, 한국 중학생의 컴퓨팅 사고력에는 사회경제적 배경보다 학교 수업 맥락, ICT 활용 방식, ICT 자기효능감, 교육 포부와 같은 교육적으로 조정 가능한 요인이 더 직접적으로 작용한다고 볼 수 있다. 특히 학생의 교육 포부가 컴퓨팅 사고력과 정적으로 연결된 점은, 학생의 자기 전망과 미래 지향이 컴퓨팅 사고력 예측과 밀접하게 관련됨을 보여준다. 즉, 컴퓨팅 사고력 향상을 위한 교육적 개입은 배경 요인의 결핍을 보완하는 방안 이외에, 학생이 ICT를 활용하여 과제를 해결할 수 있다는 인식과 미래 학업에 대한 기대를 형성하도록 지원하는 방향으로 설계될 필요가 있다.

이러한 논의를 종합하면, 한국 중학생의 컴퓨팅 사고력은 단순한 개인 배경 변수보다 ICT 활용 경험, 자기효능감, 수업 맥락과 같은 교육적으로 형성이 가능한 요인들과 더 밀접하게 연결되어 있다고 볼 수 있다. 동시에 ICT 활용이 컴퓨팅 사고력 예측에 기여하는 방향이 언제나 정적으로 기여하는 것은 아니며, 이는 ICT의 활용 목적, 과업의 성격, 학습 경험의 질에 따라 달라질 수 있음을 확인하였다. 이는 컴퓨팅 사고력을 위한 교육이 단순한 디지털 도구 사용 확대나 코딩 경험 제공에 머물러서는 안 되며, 문제 해결 중심의 과업 설계, 정보 탐색과 판단 활동의 강화, 그리고 학생의 ICT 자기효능감을 높이는 방향으로 정교하게 조직될 필요가 있음을 시사한다.

본 연구는 ICILS 2023 한국 자료를 활용하여 학생 설문 전체 변수를 동시에 투입한 랜덤포레스트 분석과 SHAP 분석을 통해 한국 중학생의 컴퓨팅 사고력을 예측하는 주요 변수와 응답 수준에 따른 경향성을 탐색하였다. 특히 기존 연구가 일부 변수에 대한 평균 차이 또는 선형적 구조관계를 중심으로 논의한 것과 달리, 본 연구는 다양한 변수의 상대적 중요도와 비선형적 경향을 함께 제시하였다는 점에서 의의가 있다. 또한 성별이나 부모 학력보다 ICT 관련 경험과 자기효능감, 수업 맥락과 같은 교육적으로 개입이 가능한 요인이 더 중요한 예측변수로 나타났다는 점은 학교 현장에서의 교육 설계에 실질적인 시사점을 제공한다.

본 연구는 ICILS에서 제공하는 척도 변수 대신 123개의 개별 문항을 분석 단위로 사용하였다. 척도 변수를 활용한 분석은 영역 수준의 결론에 국한되나, 문항을 활용한 분석은 동일 영역 안에서도 어떤 문항의 내용들이 컴퓨팅 사고력 예측에 기여하는지를 식별할 수 있으므로 구체적인 예측 기여를 확인할 수 있기 때문이다. 실제로 본 연구에서 같은 영역에 속한 문항들이 서로 다른 방향의 기여를 보인 결과는

문항 단위 분석을 통해서만 포착할 수 있는 것이었다. 그러나 유사한 개념을 측정하는 문항들이 함께 투입되면 중요도가 여러 문항에 분산되어, 개별 문항의 중요도가 해당 구성 개념 전체의 중요성을 충분히 반영하지 못할 가능성이 존재한다. 따라서 본 연구의 중요도 순위는 영역 간 상대적 중요성보다는 문항 수준의 기여로 해석하는 것이 적절하며, 후속 연구에서는 척도 변수를 활용한 분석과의 비교, 그리고 주요 예측변수 간 관계 구조에 대한 분석을 통해 본 연구의 결과를 보완할 필요가 있다.

본 연구에서 사용한 모형의 설명력은 $R^2=.381$ 로 높은 편은 아니다. 다만 본 연구의 목적은 개별 학생의 점수를 정확히 예측하는 것이 아니라, 여러 설문 변수 가운데 컴퓨팅 사고력과 밀접한 주요 예측변수를 탐색하는 데 있다. 이때 중요한 것은 예측의 정확도보다 변수 중요도 순위의 안정성이다. OOB 결정계수(.374)는 검증용 결정계수(.381)와 거의 같았고, Table 2에 나타나듯 종속변수를 5개 추정값으로 바꾸어도 상위 예측변수의 구성이 거의 그대로 유지되어 순위의 신뢰성을 확인할 수 있었다. 설명력이 이 수준에 그친 것은 모형보다 자료의 특성과 관련이 깊다고 볼 수 있는데, 컴퓨팅 사고력은 학생 설문으로 측정되지 않은 여러 요인의 영향도 함께 받기 때문이다. 따라서 본 연구에서 사용한 모형은 예측 모형으로 쓰기에는 충분하지 않지만 연구의 목적인 주요 예측변수를 탐색하는 데에는 무리가 없다고 판단하였다.

다만 본 연구는 ICILS 2023의 한국 중학교 2학년 자료만을 활용하였으므로 다른 학교급이나 국가로 일반화하는 데 한계가 있으며, 학생의 자기기입식 응답을 기반으로 하였다는 점에서 실제 수행 수준이나 교실 맥락을 충분히 반영하지 못했을 가능성이 있다. 또한 예측 모형에 기반한 분석이라는 점에서 특정 변수가 컴퓨팅 사고력에 인과적으로 영향을 미친다고 단정하기는 어렵다. 후속 연구에서는 국가 간 또는 학교급 간 비교를 통해 주요 예측변수의 구조가 어떻게 달라지는지 확인할 필요가 있으며, 교실 관찰, 수행 자료, 면담 자료 등을 결합하여 실제 학습 경험의 질을 입체적으로 분석할 필요가 있다. 아울러 랜덤포레스트 외의 다양한 머신러닝 기법과 종단적 설계를 적용하여 ICT 자기효능감, 수업 맥락, 프로그래밍 경험이 컴퓨팅 사고력 발달에 어떠한 영향을 미치는지를 보다 정교하게 검증할 필요가 있다. 나아가 교사와 수업의 질, 학교 수준의 교육 여건, 학생의 인지적 특성 등 본 연구가 다루지 못한 변수를 함께 고려한다면 컴퓨팅 사고력에 대한 설명력을 한층 높일 수 있을 것이다.

참고문헌

- [1] Jeong, M. (2026). An Integrative Meta-Analysis of Creativity Research in Korea Over the Last 25 Years (2001–2025). *Korean Journal of Thinking Development*, 22, 17-35.
- [2] Jang, E., Oh, K. (2025). Analysis of the Impact of Generative AI Utilization on Computational Thinking and Problem-Solving Ability. *The Journal of Korean Association of Computer Education*, 28(7), 83-98. <https://doi.org/10.32431/kace.2025.28.7.008>
- [3] Wing, J. M. (2006). *Computational thinking*. *Communications of the ACM*, 49(3), 33-35. <https://doi.org/10.1145/1118178.1118215>
- [4] Kim, S., Lee, Y. (2020). Computational Thinking of Middle School Students in Korea. *Journal of The Korea Society of Computer and Information*, 25(5), 229-241. <https://doi.org/10.9708/jksoci.2020.25.05.229>
- [5] Jeon, S., Park, S. (2022). Analysis of Computational Thinking Characteristics of Korean Middle School Students Based on ICILS 2018 Results. *The Journal of Korean Association of Computer Education*, 25(6), 1-11. <https://doi.org/10.32431/kace.2022.25.6.001>
- [6] Ministry of Education. (2015). The revised national curriculum 2015 for elementary and secondary schools. Ministry of Education.
- [7] Park, J., Kil, H. (2021). Structural Relationships among Factors Affecting Computer and Information Literacy of Middle School Students : Focusing on ICILS 2018. *Journal of Education & Culture*, 27(3), 307-328. <https://doi.org/10.24159/joec.2021.27.3.307>
- [8] Jeon, S., Kim, K. (2021). Analysis of the correct-answer rate of Computer and Information Literacy items of Korean students in ICILS 2018. *The Journal of Korean Association of Computer Education*, 24(5), 27-35. <https://doi.org/10.32431/kace.2021.24.5.003>
- [9] Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32. <https://doi.org/10.1023/A:1010933404324>
- [10] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning. Data mining, inference, and prediction (2nd ed.)*. Springer. <https://doi.org/10.1007/978-0-387-84858-7>
- [11] Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.
- [12] Wing, J. M. (2011, March 6). *Research notebook: Computational thinking—What and why?* Carnegie Mellon University, School of Computer Science. <https://openlab.bmcc.cuny.edu/edu-211-b181-spring-2024-j-longley/wp-content/uploads/sites/3085/2023/06/CT-What-And-Why-copy.pdf>
- [13] Computer Science Teachers Association, & International Society for Technology in Education (2011). *Operational definition of computational thinking for K-12 education*. ISTE & CSTA.
- [14] Grover, S., & Pea, R. (2013). Computational thinking in K-12: A review of the state of the field. *Educational Researcher*, 42(1), 38-43. <https://doi.org/10.3102/0013189X12463051>
- [15] Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*, 22, 142-158. <https://doi.org/10.1016/j.edurev.2017.09.003>
- [16] Park, C., Hyun, J. (2018). Gender and Abstract Thinking Disposition Difference Analyses of Visual Diagram Structuring for Computational Thinking Ability. *The Journal of Korean Association of Computer Education*, 21(3), 11-22. <https://doi.org/10.32431/kace.2018.21.3.002>
- [17] Duckworth, D., & Fraillon, J. (2024). Computational thinking framework. In J. Fraillon & M. Rožman (Eds.), *IEA International Computer and Information Literacy Study 2023 assessment framework* (pp. 35-43). Springer. https://doi.org/10.1007/978-3-031-61194-0_3
- [18] Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 international report*. Springer. <https://doi.org/10.1007/978-3-030-38781-5>
- [19] Guan, T., Zhang, L., Ji, X., He, Y., & Zheng, Y. (2025). Student characteristics and ICT usage as predictors of computational thinking: An explainable AI approach. *Journal of Intelligence*, 13(11), 145. <https://doi.org/10.3390/jintelligence13110145>
- [20] Hsu, T.-C., & Chen, M.-S. (2025). Using machine learning approaches to predict Taiwanese eighth graders' computational thinking performance in ICILS 2023 study. *Computers in Human Behavior Reports*, 17, 100896. <https://doi.org/10.1016/j.chbr.2025.100896>
- [21] Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 30, 4765-4774.
- [22] Kim, J., Kim, J. (2022). Exploring the Influential Factors of Teacher-student Relationship based on Random Forest Regression and Interpretation using SHAP. *Journal of Educational Evaluation*, 35(3), 409-437. <https://doi.org/10.31158/JEEV.2022.35.3.409>
- [23] Jang, E., Chung, H. (2023). Exploring the Major Predictors Affecting Academic Helplessness of High School Students: Application of Random Forest and SHAP. *Korean Journal of Educational Research*, 61(1), 31-63. <https://doi.org/10.30916/KERA.61.1.31>
- [24] Baek, Y., Chung, H. (2023). Exploring the Predictive Variables on Adolescents Career Maturity Development Based on Mixed Effect Random Forest: Interpretation using SHAP. *Journal of Educational Evaluation*, 36(3), 473-498. <https://doi.org/10.31158/JEEV.2023.36.3.473>
- [25] von Davier, M., Gonzalez, E., & Mislevy, R. J. (2009). What are plausible values and why are they useful? IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 2, 9-36.
- [26] Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, 28(1), 112-118. <https://doi.org/10.1093/bioinformatics/btr597>
- [27] Meyer, S., Daniel, T., & Tieck, S. (Eds.). (2025). ICILS 2023

user guide for the international database. International Association for the Evaluation of Educational Achievement (IEA).

- [28] Lee, Y., Koo, D., & Lim, H. (2017). Effects of Student- and School-level ICT-related Factors on Computer-based Problem Solving: Focusing on Korea and Japan. *JOURNAL OF The Korean Association of information Education*, 21(4), 425-435. <https://doi.org/10.14352/jkaie.2017.21.4.425>
- [29] Jeon, S., Son, Y., & Park, S. (2020). Analysis of the ICILS 2018 Results by Korean Students' Educational Experience in Computer and Information Literacy and Computational Thinking. *The Journal of Korean Association of Computer Education*, 23(3), 1-8. <https://doi.org/10.32431/kace.2020.23.3.001>
- [30] Na, C., Joo, H., Lee, J., & Kim, D. (2018). Inducing Computational Thinking in Korean SW Education: Synthesizing Standardized Mean Changes through Meta-analysis. *Journal of Educational Technology*, 34(3), 775-815. <https://doi.org/10.17232/KSET.34.3.775>



오준석

- 2020년 영남대학교 교육학과(교육학사)
- 2022년 영남대학교 교육학과 교육공학전공(교육학 석사)
- 2024년 영남대학교 교육학과 교육공학전공 박사 수료
- 2022년~현재 영남대학교 교육개발센터 연구원

✚ 관심분야: 교육공학, 교수설계, 교수방법, 강의평가
✉ sgj741@naver.com



김승재

- 2024년 영남대학교 교육학과 교육공학전공(교육학 석사)
- 2026년 영남대학교 교육학과 교육공학전공 박사 수료

✚ 관심분야: AI기반 학습환경 설계, AI 에이전트, 생성형 AI, 머신러닝
✉ ican0407@gmail.com