creative commons

COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

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

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

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

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

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

교육학박사 학위논문

심층인공신경망 모형을 활용한 기초학력 미달 학생 예측연구

- 교육패널자료 및 불균형 자료 활용을 중심으로 -

2022년 8월

서울대학교 대학원

교육학과 교육학 전공

홍 유 정

심층인공신경망 모형을 활용한 기초학력 미달 학생 예측연구

- 교육패널자료 및 불균형 자료 활용을 중심으로 -

지도교수 박 현 정

이 논문을 교육학 박사 학위논문으로 제출함
2022년 6월

서울대학교 대학원
교육학과 교육학전공
홍 유 정

홍유정의 박사 학위논문을 인준함
2022년 8월

위 원 장 백 순 근 (인)

부위원장 이 현 숙 (인)

위 원 정 혜 원 (인)

위 원 함 은 혜 (인)

위 원 박 현 정 (인)

국 문 초 록

기초학력 미달(이하 기초미달) 학생의 학습 결손은 학년을 거듭할수록 심화되는 특성이 있으며, 누적된 학습 결손은 학업에 대한 부정적 정서를 초래하거나 학교생활 적응을 어렵게 하는 원인이 되기도 한다. 아울러 학습 결손이 나타나는 원인은 매우 다양하고 복잡하기 때문에, 기초미달 학생에게 실질적인 도움을 주기 위해서는 예방적 차원의 지원과 함께 개별화된 지원을 필요로 한다. 따라서 개별 학생의 향후 학업성취를 예측할 수 있는 시스템을 구축할 수 있다면 문제상황이 발생하기 이전 또는 초기 단계에 교육적 지원을 제공하는 데 도움이 될 것이라 예상할 수 있다.

하지만 학습부진 학생들의 학업성취도를 예측한 기존 선행연구를 살펴보면, 개별 학생의 학업성취도 예측보다는 교육 맥락 변인을 활용하여 학업성취도를 예측하는 모형을 설정하고, 이 예측모형에서 예측변인과 학업성취도 간 관계를 살펴보는 데 초점을 두는 경향이 있었다. 이는 교육연구에서는 개별학생 예측보다는 현 상황을 진단하고 교육적 지원을 제공하는 것이 중요하게 여겨졌고, 학업성취도에 영향을 미치는 주요 변인을 확인함으로써 교육과정 운영 및 정책적 시사점을 도출하고자 하는 데 연구의 목적이 있었기 때문이다.

또한 학생들의 학업성취를 예측한 연구 중 기초미달 학생에게 집중된 연구가 상대적으로 적은 모습을 확인할 수 있었다. 학교 또는 교육청 수준에서 기초미달 비율에 영향을 미치는 변인을 살펴보거나, 학업성취 하위 학생을 기초미달 수준의 학생과 통합하여 분석하는 모습이 나타났다. 이는 기초미달 학생이 전체 학생에 비해 매우 적은 비율에 속하는 학생들로서, 통계적으로는 불균형 자료(imbalanced data)에 해당되기 때문이다. 이로 인해 통계적 기본가정 충족이 쉽지 않고, 분석에 활용 가능한 사례 수를 확보하기 어렵다는 점에서 기초미달 학생에 초점을 둔 연구가 활발하게 이루어지지 못한 면이 있었다.

이에 기초미달 학생 예측모형을 구축하기 위한 연구방법으로 심층인공신경망(Deep neural network) 모형을 고려해볼 수 있다. 심층인공신경망 모형은 개별 학생을 정확하게 예측하는 모형을 설정하는 데 목적이 있는 연구방법이며, 자료의 특정 분포를 가정하지 않고 주어진 자료에서 확인할 수 있는 패턴을 학습하여 예측모형을 구축한다는 특징이 있다. 또한 예측모형을 훈련시키는 과정에서 불균형 정도에 따른 소수집단 정보가 누락되지 않도록 하는 여러 기법을 다양하게 적용할 수 있는 방법이기도 하다.

한편, 학업성취도 연구에서 많이 활용되는 교육패널자료는 학생들의 학습 환경과 관련된 다양한 변인에 대한 정보가 있으며, 매년 동일한 학생을 추적조사하여 변화에 대한 정보가 있는 자료이다. 교육분야에서는 개인정보 보호 등의 이유로 평가 결과나 상담 자료 또는 수업 상황에서의 음성, 문자 자료 등의 활용이 쉽지 않다. 상대적으로 패널자료는 국가 또는 시도 수준에서 학생들의 발달적 정보를 확인하고 교육 운영에 활용하기 위해 수년 간 구축된 자료이다. 따라서 교육패널자료를 활용하여 기초미달 학생 예측모형의 구축 가능성을 탐색해볼 필요가 있다.

이 연구에서는 수년에 걸쳐 체계적으로 수집되어온 교육패널 자료를 활용하여 소수집단에 속하는 기초미달 학생을 예측하는 심층인공신경망 모형을 탐색하는 데 목적을 두고 연구를 진행하였다. 분석에는 한국교육종단연구2013를 활용하였으며, 중학교 2학년 정보를 활용하여 중학교 3학년 시기의 국어, 영어, 수학 기초미달 여부를 예측하는 모형을 구현해 보고자 하였다.

실질적인 심층인공신경망 예측모형을 구축하기 이전에, 심층인공신경망이 많은 수의 변인을 한 모형에 활용할 수 있는 방법이긴 하나, 학업성취와 관련이 없는 변인의 활용은 오히려 예측모형의 성능을 떨어뜨릴 수 있다는 점을 고려하였다. 따라서 활용 가능한 361개 변인 중 학업성취와 관련이 높은 변인과 그 정도를 확인하였고, 이 결과를 바탕으로 예측변인의 개수를 늘려가며 모형 성능을 확인하고자 하였다.

이후 기초미달 학생 예측모형 구축에서는 예측모형 훈련 시 소수집단

정보를 다수집단만큼 반영해주는 방법과 심층인공신경망 종류에 따른 예측 성능을 비교하였다. 구체적으로 소수집단 정보 반영방법은 임의과대표집, Borderline-SMOTE, ADASYN, 비용민감방법을 적용하였으며, 심층인공신경망 종류는 다층퍼셉트론(Multi-layer perceptron, 이하 MLP)과 합성곱신경망(Convolutional neural network, 이하 CNN)으로 예측모형을 구현해보고자 하였다. 비교한 결과를 토대로 과목별 최종 예측모형을 확인하였으며, 기초미달 학생 예측모형의 성능을 향상시킬 수 있는 방안을 모색하였다.

주요 연구 결과를 요약하여 제시하면 다음과 같다.

우선, 다수의 변인 중 종속변인과 관련 있는 변인을 확인하는 동시에 각 변인의 계수 추정이 가능한 LASSO(Least Absolute Shrinkage and Selection Operator) 분석 방법을 활용하여 주요 설명변인을 확인한 결과, 총 361개 변인 중 국어에서는 91개, 영어에서는 84개, 수학에서는 83개 변인이 각 과목의 학업성취도와 관련이 있는 설명변인으로 나타났다. 하지만 선정된 변인의 계수를 확인하였을 때 여전히 종속변인과 관련성이 낮다고 판단되는 변인들이 포함된 모습이 나타났고, 이에 따라 계수의 절대값이 높은 상위 변인을 중심으로 예측에 활용하는 변인의 수를 늘려가며 예측모형을 구축하였다.

다음으로 기초미달 비율이 다른 국어, 영어, 수학 과목의 기초미달 학생을 예측하기 위하여, 예측변인의 개수, 소수집단 정보 반영방법, 심층인공신경망 종류에 따라 심층인공신경망 예측모형을 상세화하고 그 성능을 살펴본 결과는 다음과 같다. 첫째, 모든 과목에서 예측변인의 개수가 일정 수준 이상 늘어나면 예측모형의 성능이 낮아지는 경향이 나타났으며, 국어에서는 상대적으로 성능 감소가 큰 모습을 보였다. 이는 심층인공신경망 예측모형을 구축할 때 적절한 예측변인의 선정을 고민할 필요가 있음을 보여준다.

둘째, 불균형 정도를 조정하지 않은 Original 모형에서는 기초미달 학생 예측보다 기초이상 학생 예측에 적합한 모습이 나타났다. 특히 집단별 불균형 정도가 높은 국어에서는 대부분의 학생을 기초이상 학생으로

예측하고 있었다. 상대적으로 모형 훈련 시 소수집단 정보를 다수집단만큼 반영해준 경우에는 AUC, F2-score, G-mean, 민감도가 상승하는 경향을 보였다. 또한 소수집단 정보 반영방법 간 성능을 확인하였을 때에는 표집방법에 비해 비용민감방법을 적용한 경우의 모형 성능이 더 좋은 경향을 보였으며, 특히 예측변인이 49개 이상으로 커짐에 따라 차이가 커지는 모습이 모든 과목에서 유사하게 나타났다. 따라서 불균형 자료를 예측하는 경우에는 불균형 정도를 조정해주는 방법을 적용하는 것이 필요함을 확인하였으며, 이와 함께 불균형 자료 예측에서 비용민감방법의 적용을 제안하였다.

셋째, CNN 모형에 비해 MLP 모형의 예측 성능이 더 좋게 나타났으나, 그 차이가 크다고 판단하긴 어려웠다. 하지만 CNN 모형에서는 소수집단 정보 반영방법에 관계없이 예측변인 개수에 따른 모형 성능 감소폭이 적은 경향이 나타났다. 이를 바탕으로 많은 예측변인을 활용할 때에는 CNN 모형 적용을 고려해볼 수 있으나, 패널자료의 특성과 심층인공망 종류에 따른 모형 설정의 한계 및 자료 변환 등을 고려할 때 MLP 모형을 우선적으로 적용해볼 수 있다는 점을 논하였다.

최종적으로 모든 과목에서 예측변인 25개를 활용하고, 비용민감방법을 적용한 MLP 모형의 성능이 가장 좋게 나타났다. 하지만 정분류율이 0.704~0.768일 때 민감도 0.703~0.762 수준으로 예측모형을 실질적으로 활용할 수 있을 정도의 예측성능을 보이진 않았다. 따라서 이를 바탕으로 패널자료를 활용한 심층인공신경망 모형에서 고려해야할 사항과 함께 향후 심층인공신경망 모형의 활용 및 연구에 대한 제언을 하였다.

주요어 : 기초학력 미달, 학업성취 예측, 교육패널자료, 불균형 자료, 심층인공신경망

학 번 : 2018-37446

목 차

Ⅰ. 서 론 ····· 1
1. 연구의 필요성 ····· 1
2. 연구문제 ····· 7

Ⅱ. 이론적 배경 ····· 8
1. 교육연구에서 예측에 대한 논의 ····· 8
2. 기초학력 미달 학생 연구 ····· 13
가. 기초학력 미달 학생 연구의 개관 ····· 13
나. 기초학력 미달 학생의 특성 ····· 19
3. 심층인공신경망을 활용한 예측모형 ····· 27
가. 예측변인의 선정 ····· 29
나. 심층인공신경망 ····· 31
다. 불균형 자료와 심층인공신경망 ····· 49

Ⅲ. 연구방법 ····· 55
1. 연구 절차 ····· 55
2. 분석 자료 ····· 60
가. 연구대상 ····· 61
나. 주요변인 ····· 63
3. 연구모형 및 분석절차 ····· 69
가. 주요 설명변인 선정 모형 ····· 69
나. 기초미달 학생 예측모형 ····· 72

Ⅳ. 연구결과 ····· 89
1. 주요 설명변인 선정 결과 ····· 89
가. 과목별 설명변인 선정 결과 ····· 89
나. 주요 설명변인 선정에 따른 예측모형 설정 ····· 109
2. 기초미달 학생 예측모형 구현 ····· 111
가. 국어 ····· 111
나. 영어 ····· 121
다. 수학 ····· 131
라. 예측모형 성능 비교 결과 종합 ····· 141

Ⅴ. 결론 ····· 148
1. 요약 ····· 148
2. 논의 및 시사점 ····· 151

참고문헌 ····· 157
부 록 ····· 173
Abstract ····· 183

표 목 차

〈표 Ⅲ-1〉 심층인공신경망 예측모형 상세화 ·····57
〈표 Ⅲ-2〉 중2 과목별 기초능력검사 척도점수 기술통계 ·····64
〈표 Ⅲ-3〉 중3 과목별 성취수준에 따른 학생 수 및 비율 ·····64
〈표 Ⅲ-4〉 설명변인 선정 모형에서 활용한 독립변인 ·····67
〈표 Ⅲ-5〉 LASSO 분석을 위한 R 명령어 ·····71
〈표 Ⅲ-6〉 표집방법 적용을 위한 R 명령어 ·····75
〈표 Ⅲ-7〉 비용민감기법 적용을 위한 Python 명령어 ·····76
〈표 Ⅲ-8〉 다층퍼셉트론 모형 훈련을 위한 Python 명령어 ·····79
〈표 Ⅲ-9〉 합성곱신경망 모형 훈련을 위한 Python 명령어 ·····83
〈표 Ⅲ-10〉 기초미달 학생 실제 값과 예측 값에 대한 혼동행렬 ·····85
〈표 Ⅳ-1〉 중2 국어 학업성취도 주요 설명변인 선정 결과 ·····90
〈표 Ⅳ-2〉 중2 국어 학업성취도 주요 설명변인(상위 30개) ·····94
〈표 Ⅳ-3〉 중2 영어 학업성취도 주요 설명변인 선정 결과 ·····97
〈표 Ⅳ-4〉 중2 영어 학업성취도 주요 설명변인(상위 30개) ·····100
〈표 Ⅳ-5〉 중2 수학 학업성취도 주요 설명변인 선정 결과 ·····103
〈표 Ⅳ-6〉 중2 수학 학업성취도 주요 설명변인(상위 30개) ·····107
〈표 Ⅳ-7〉 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-Original) ·····111
〈표 Ⅳ-8〉 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-표집방법) ·····115
〈표 Ⅳ-9〉 표집방법에 따른 MLP, CNN 모형 성능 비교 (국어) ·····117
〈표 Ⅳ-10〉 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-비용민감방법) ·····118
〈표 Ⅳ-11〉 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (국어) ·····120
〈표 Ⅳ-12〉 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-Original)

·····122
<표 Ⅳ-13> 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-표집방법)
·····124
<표 Ⅳ-14> 표집방법에 따른 MLP, CNN 모형 성능 비교 (영어) ·····127
<표 Ⅳ-15> 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-비용민감방법) ·····128
<표 Ⅳ-16> 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (영어)
·····130
<표 Ⅳ-17> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-Original)
·····131
<표 Ⅳ-18> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-표집방법)
·····133
<표 Ⅳ-19> 표집방법에 따른 MLP, CNN 모형 성능 비교 (수학) ·····137
<표 Ⅳ-20> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-비용민감방법) ·····138
<표 Ⅳ-21> 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (수학)
·····140
<표 Ⅳ-22> 예측변인 개수 및 소수집단 정보 반영방법에 따른 과목별 기초미달 학생 예측모형 성능 (MLP) ·····143
<표 Ⅳ-23> 예측변인 개수 및 소수집단 정보 반영방법에 따른 과목별 기초미달 학생 예측모형 성능 (CNN) ·····145
<표 Ⅳ-24> 과목별 최종모형의 모형 성능과 혼동행렬 ·····147
<표 부록 1-1> 중2 국어 학업성취도 주요 설명변인(상위 31~91개) ·····173
<표 부록 1-2> 중2 영어 학업성취도 주요 설명변인(상위 31~84개) ·····177
<표 부록 1-3> 중2 수학 학업성취도 주요 설명변인(상위 31~83개) ·····180

그 림 목 차

[그림 Ⅱ-1] 인공신경망 관련 개념들 간 관계 ····· 28
[그림 Ⅱ-2] 단층퍼셉트론의 구조 ····· 32
[그림 Ⅱ-3] 은닉층 계층 수에 따른 표현의 복잡도 ····· 36
[그림 Ⅱ-4] 다층퍼셉트론의 구조의 예 ····· 37
[그림 Ⅱ-5] 2차원, 3차원에서 경사하강법의 적용 ····· 41
[그림 Ⅱ-6] 시그모이드, 쌍곡선탄젠트, ReLU 함수 그래프 ····· 43
[그림 Ⅱ-7] 깊은 다층퍼셉트론의 구조 ····· 46
[그림 Ⅱ-8] 합성곱신경망의 구조 ····· 47
[그림 Ⅱ-9] 합성곱 계층과 통합 계층의 적용 과정 ····· 48
[그림 Ⅲ-1] 연구절차 및 주요 연구내용 ····· 56
[그림 Ⅲ-2] 연구 과정별 분석자료의 활용 ····· 59
[그림 Ⅲ-3] 다층퍼셉트론 연구모형 ····· 78
[그림 Ⅲ-4] CNN 입력을 위한 자료 변환의 예 ····· 81
[그림 Ⅲ-5] 합성곱신경망 연구모형 ····· 83
[그림 Ⅳ-1] 국어 중2 학업성취도 제약모수의 20-fold 교차타당화 결과 90
[그림 Ⅳ-2] 설명변인 순위에 따른 LASSO 계수 (국어) ····· 92
[그림 Ⅳ-3] 영어 중2 학업성취도 제약모수의 20-fold 교차타당화 결과 96
[그림 Ⅳ-4] 설명변인 순위에 따른 LASSO 계수 (영어) ····· 98
[그림 Ⅳ-5] 수학 중2 학업성취도 제약모수의 20-fold 교차타당화 결과 103
[그림 Ⅳ-6] 설명변인 순위에 따른 LASSO 계수 (수학) ····· 105
[그림 Ⅳ-7] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (국어-Original) ····· 112
[그림 Ⅳ-8] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (국어-표집방법) ····· 116
[그림 Ⅳ-9] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (국어-비용민감방법) ····· 119

[그림 Ⅳ-10] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (영어-Original) ·····122
[그림 Ⅳ-11] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (영어-표집방법) ·····126
[그림 Ⅳ-12] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (영어-비용민감방법) ·····129
[그림 Ⅳ-13] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (수학-Original) ·····132
[그림 Ⅳ-14] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (수학-표집방법) ·····135
[그림 Ⅳ-15] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (수학-비용민감방법) ·····139
[그림 Ⅳ-16] 과목별 기초미달 학생 예측모형 성능 비교 (MLP) ·····144
[그림 Ⅳ-17] 과목별 기초미달 학생 예측모형 성능 비교 (CNN) ·····146

I. 서 론

1. 연구의 필요성

모든 학생이 사회의 일원으로 살아갈 수 있는 기본적인 능력을 키워주는 것은 교육의 본질이자 학교교육의 중요한 책무 중 하나이다. 따라서 국가적 차원에서는 기본적인 학습 수준에 도달하지 못하는 학습부진 학생들을 지속적으로 확인하는 동시에, 학력 향상을 위한 교육적 지원을 제공하고자 한다(교육부, 2021).

학습부진 학생들의 특징을 살펴본 연구에 따르면, 이들의 학습 결손은 특정 시기에 나타나는 현상이기보다는 학년이 올라감에 따라 점차 심화되는 특성이 있다고 보고하고 있다(김지은, 김지하, 2012; 유미경, 엄준용, 2016; 임현정, 2009). 또한 누적된 학습 결손은 학업적 무기력을 초래하거나 학교생활 적응을 어렵게 하는 원인이 되기도 하였다(송연주 외, 2014; 허유성 외, 2014). 이에 교육 관계자들은 학습 결손이 심화되는 것을 막고자 노력하는 동시에, 사후적 지원보다는 예방적 차원에서의 지원 방안을 모색하고 있다.

국가 차원의 학력 진단을 위하여 우리나라는 2003년부터 국가수준 학업성취도 평가 시스템을 구축하고, 학생들의 성취수준을 점검하고 있다. 성취수준은 우수학력, 보통학력, 기초학력, 기초학력 미달과 같이 4개 수준으로 구분하였으며, 여기서 기초학력이란 ‘평가 대상 학년급 학생들이 성취하기를 기대하는 기본 내용을 부분적으로 이해하는 수준’으로 정의하였다(김성숙 외, 2010: 21). 이에 따라 기초학력 미달(이하 기초미달) 학생은 기본 내용을 부분적 이해에도 못 미치는 수준이라 할 수 있으며, 적극적인 도움이 필요한 학생이다.

이러한 진단 결과를 토대로 하여 정부는 학력 향상을 위한 여러 정책들을 실시하였다. 특히 학력 향상의 중요성이 커진 2009년을 전후로 기

초학력 향상에 대한 단위학교의 책무성을 강화하고 다양한 정책적 지원을 제공하였다(교육과학기술부, 2009). 학력향상 중점학교 운영을 통해 학교역량을 강화하고자 하였으며, 이와 함께 기초미달 학생들을 위한 지도 자료 및 프로그램을 개발하여 보급하고, 학습부진학생 지도를 위한 인턴교사를 배치하는 등의 기초학력 향상 정책들을 시행하였다. 하지만 학습 결손이 나타나는 원인이 매우 다양하고 복잡함에도 국가, 시·도 교육청, 단위학교 차원에서의 지원 방안들과 정책들이 결과론적인 대응에 치중하고 있다는 문제점이 제기되었다. 이에 최근에는 '한 아이도 놓치지 않는 맞춤형 교육 실현'을 모토로 기초미달 학생 지원 정책의 흐름이 달라지는 경향을 보이고 있다(김태은 외, 2018).

선행연구에 따르면 기초학력 미달 수준으로 학습에 어려움을 겪는 학생들은 학업적으로는 일반 학생과 매우 다른 학습 전략을 사용하고, 정서적으로는 우울, 스트레스가 높으며, 학업정서 역시 부정적인 경향을 보이는 경향이 있었다(백병부, 2010; 전명남 2017; 전명남, 양명희, 2015). 그러나 개별 학생이 자신의 학습 장면에서 직면해 있는 어려움을 일반화시키는 것은 다소 어려운 면이 있었다(이화진 외, 2009). 이러한 이유로 학습 결손 해결을 위해서는 개별화된 도움이 필수적임을 강조하고 있었으며, 기초학력 보장 정책 또한 학생의 문제 상황을 파악하고, 가정·학교·지역사회 차원의 도움을 제공할 수 있는 방향으로 변화하고 있는 추세이다(김태은 외, 2018; 이대식, 2015).

이에 따라 교육 연구에서는 교육적 지원을 적기에 제공하기 위하여 학습 결손 발생 이전 또는 초기 단계에 도움이 필요한 학생들을 확인할 수 있는 연구의 필요성이 제기되고 있다.

학습부진 학생들의 학업성취도를 예측한 선행연구를 살펴보면, 다음의 특징을 확인할 수 있었다. 우선, 개별 학생을 정확하게 예측하는 모형을 탐색하는 연구보다는 다양한 교육 맥락 변인을 활용하여 학업성취도를 예측하는 통계적 모형을 설정한 후, 이를 토대로 예측변인과 학업성취도 간 관계를 살펴보는 연구들이 많았다. 김양분 외(2016) 연구와 강은영, 김진구(2018) 연구에서는 위계적 선형모형을 활용하여 전년도 정보

로 다음 해 학업성취도를 예측하고 이와 관련된 학생 및 학교 요인을 탐색하였다. 또한 임현정 외(2017) 연구에서는 초6-중3-고2 연계자료에 다층성장모형을 적용하여 개별 학생의 학업성취 변화와 유형을 파악하고, 이에 영향을 미치는 판별 요인을 살펴보는 모습을 확인할 수 있었다. 이러한 연구들에서는 학생들의 다양한 배경변인에 대한 정보를 포함하고 학생의 발달적 상황을 매년 추적 조사하는 대규모 교육패널자료를 주로 활용하고 있었고, 교육자료가 위계적 속성이 있다는 점에서 다층모형을 활용하는 경우가 많음을 확인할 수 있었다.

또한 기초미달 학생에게 집중된 연구가 상대적으로 적은 모습을 확인할 수 있었다. 기초미달 학생을 대상으로 하기보다는 학교 수준 또는 교육청 수준에서 기초미달 비율에 영향을 미치는 교육 맥락 변인을 살펴보는 경향이 있었으며(권재기, 2013; 이성은 외, 2012; 이효정 외, 2018; 장선희, 정제영, 2018 등), 개별 학생을 대상을 분석하는 경우에는 분석대상 학생 중 학업성취 하위 25~50% 학생을 통합하여 연구를 진행하여(박미현, 허균, 2021; 안영은, 2019; 허유성 외, 2014 등) 기초학력에 미치지 못하는 수준인 학생의 특성을 확인하기는 어려운 부분이 있었다.

이와 같이 선제적인 도움이 필요한 기초미달 학생을 개별적으로 예측하고 이를 바탕으로 맞춤형 지원을 제공하고자 하는 연구가 많지 않은 이유는 교육연구에서의 예측 목적이 학생들의 발달적 특징을 확인하고 학업성취도와 관련이 높은 변인들을 통계적으로 검증함으로써 교육과정 운영이나 정책적 시사점을 도출하는 데 있었기 때문이다. 기초미달 학생을 위한 교육적 지원을 고안하기 위해서는 현 상황에 대한 진단이 선행되어야 한다는 점에서 예측변인과 학업성취도 간 상호작용하는 양상을 확인하고자 하는 연구가 필수적으로 요구되었다. 이에 더해 예측연구에서 주로 활용하고 있는 대규모 교육패널자료는 모집단으로 상정한 학생 중 일부를 표집한 자료이기 때문에, 개별 사례의 예측 정확성을 확인하는 것보다는 연구 결과를 전체 모집단을 대상으로 일반화하는 통계적 검증 과정이 중요하게 다뤄졌을 수 있다.

다음으로, 기초미달 수준에 해당하는 학생들에게 집중된 연구가 활발

하게 이루어지지 않았던 이유로 기초미달 학생이 전체 학생에 비해 매우 적은 비율에 속하는 학생들이라는 점을 생각해볼 수 있다. 이는 통계적으로 불균형 자료(imbalanced data)의 형태로서 종속변인의 정규성 가정 또는 집단 간 등분산 가정 등 통계적으로 요구되는 기본 가정 충족을 어렵게 한다. 따라서 기초미달 학생 특성 확인 및 예측을 위한 모형을 구축할 때, 통계분석 모형을 적용하는 데 제한점이 많아지게 된다(Barros et al., 2019; He & Ma, 2013). 교육패널자료 예측연구에서 주로 활용되고 있는 다층분석은 기본적 가정을 충족하는 경우 개별 학생 예측에도 효과적이라 알려져 있으며(James et al, 2013), 심층인공신경망 예측모형과 비슷한 수준의 결과를 보이기도 하였다(이창묵 외, 2020). 그러나 기초미달 학생 예측에서는 다층분석의 기본적 가정을 충족하지 않거나 통계적 검증력이 낮아질 수 있음을 예상해 볼 수 있다(Cool et al, 2009).

불균형 자료의 형태는 통계적으로 분석하기에 까다로운 면이 많다는 것 외에도 분석할 수 있는 사례가 매우 적어진다는 문제도 발생한다. 국가수준 학업성취도 자료에 따르면 기초미달 학생의 비율은 2.6~7.2%[1] 내외이며, 국제학업성취도 평가에서도 기초적 수준에 도달하지 못하는 1수준 이하 학생의 비율은 7.6~15.4% 내외이다(교육부, 2016). 이는 5천 명의 학생을 조사하는 상황을 가정하였을 때, 기초미달 학생의 수는 약 135~770명 정도라 생각할 수 있다. 이로 인해 대안적인 방법으로 학교의 기초미달 비율을 종속변인으로 하여 기초미달 학생에제 미치는 교육 환경적 요인을 간접적으로 살펴보거나(김지은, 김지하, 2012; 권재기, 2013; 장선희, 정제영, 2018), 기초미달 수준은 아니지만 성취수준이 낮은 편에 속하는 학생들을 통합하여 분석에 활용하는 모습을 확인할 수 있었다(오지영, 이수정, 2008 조헌국, 2019).

따라서 소수집단에 속하는 기초미달 학생들에게 예방적 차원의 지원을 제공하기 위하여 다양한 분야에서 예측에 적극 활용되고 있는 심층인

1) 연구에서 활용하는 자료가 2015~2016년에 조사된 자료임을 고려하여, 해당 시기의 기초미달 학생 비율을 제시하였다. 최근 국가수준 학업성취도 결과에서 나타난 기초미달 학생 비율은 약 6.3~13.4%정도이다(교육부, 2020).

공신경망 모형의 적용을 고려해볼 수 있다. 심층인공신경망 모형은 통계적 가정을 기반으로 예측방정식을 설정하고 이를 토대로 학생들의 변화를 추정하는 기존의 분석 방법과 달리, 주어진 자료에 나타난 특징 및 변인 간 관계를 포착하여 예측모형을 구성한다. 변인 간 관계를 통계적으로 검증하거나 추론하지 않기 때문에 예측변인과 종속변인 간 관계를 살펴보긴 어렵지만, 변인 간 복잡한 관계를 스스로 학습하면서 포착할 수 있기 때문에 자료의 특정 분포를 가정하기 어렵거나, 연구자가 변인 간 비선형 관계나 상호작용을 확인하기 어려운 상황에서 적용할 수 있다는 장점이 있다. 또한 예측모형의 성능을 검증할 때 새로운 자료를 정확하게 예측하는지를 확인하기 때문에 향후 학업적 어려움을 겪게 될 학생을 예측하고 판별하는 상황에 적절한 모형이라 할 수 있다(이창묵 외, 2020).

또한 기초미달 학생과 같은 불균형 자료를 예측하고자 할 때, 불균형 자료에 맞는 특정 분포를 지정해주는 대신, 예측모형을 훈련시키는 과정에서 소수집단의 정보를 다수집단의 정보만큼 반영할 수 있는 기법을 다양하게 적용할 수 있어, 자료의 특성에 따른 유연한 분석이 가능한 방법이기도 하다. 이 외에도 심층인공신경망 모형은 다른 기계학습 방법에 비해 여러 시점의 종단적 예측을 모형화하거나 시점별 활용 변인을 다르게 지정하는 것이 상대적으로 용이하고, 다양한 자료를 예측변인으로 활용하는 데 강점을 지니고 있다. 교육분야에서 패널자료와 같이 정형화된 표형식의 자료 뿐만 아니라 수행평가 및 서술형 평가 결과와 같은 문자자료, 온라인 수업에서의 실시간 로그데이터, 수업 상황에서의 음성자료 등의 활용 가능성이 높아짐을 고려할 때, 다양한 자료의 접목을 기대할 수 있는 부분이기도 하다.

실제로 다양한 분야에서 심층인공신경망 모형을 활용한 소수집단 예측연구가 진행되고 있으며, 교육연구에서도 학교 중도탈락 예측(조헌국, 2018; Barros et al., 2019), 학습장애 판별(김동일, 홍성두, 2009), 온라인 수업 중도 이탈자 선별(이현진, 2018; Chaplot et al., 2015), 학업저성취 학생 예측(이창묵 외, 2020) 등의 연구가 진행되고 있다. 이들 모두 전체

학생에 비해 비율이 상당히 적으며, 초기 파악 후 지속적인 지원이 필요한 대상이다.

한편, 학업성취도 연구에서 많이 활용되는 교육패널자료는 국가 또는 시·도 수준에서 학생들의 발달적 정보를 확인하고 교육 운영에 활용하기 위해 수년 간 구축된 자료로, 학생들의 학습 환경과 관련된 다양한 변인에 대한 정보가 포함되어 있다(김양분 외, 2016). 또한 매년 동일한 학생을 추적조사하여 변화에 대한 정보가 있는 자료이다. 따라서 이를 기초미달 예측에 활용할 수 있는 방안을 탐색하고 교육현장에의 접목 가능성을 살펴볼 필요가 있다.

이와 같은 논의를 바탕으로 이 연구에서는 대규모 교육패널자료 중 하나인 「한국교육종단연구」(Korean Education Longitudinal Study)를 활용하여 향후 기초미달이 될 학생을 예측하는 예측모형을 구축하기 위하여 심층인공신경망 모형을 적용해보고자 하였다.

기초미달 학생 예측모형을 구축하기 위하여 먼저, 실제 예측이 이루지는 상황에서는 현재 가지고 있는 정보를 토대로 예측모형을 구축해야 함을 감안하여, 예측 이전 시점의 정보만으로 예측모형에 활용할 주요 설명변인을 확인하였다. 심층인공신경망에서는 예측에 활용할 수 있는 변인의 수에 대한 제한이 없기 때문에 학생들의 정보를 모두 예측변인으로 활용할 수도 있다. 하지만 종속변인과 관련없는 예측변인은 오히려 특징을 포착하기 어렵게 하여 예측을 어렵게 하기도 한다(García et al., 2016; Snieder et al., 2019). 따라서 심층인공신경망 예측모형을 상세화하는 과정에서는 학업성취도와 관련이 높은 변인을 중심으로 예측에 활용하고자 하였다.

또한 기초미달 학생이 소수집단임을 고려하여 예측모형을 학습시키는 과정에서 소수집단의 정보를 다수집단의 정보만큼 반영할 수 있는 다양한 방법을 적용하고 그 결과를 비교하고, 이러한 방법이 심층인공신경망 모형의 종류에 따라 예측 성능의 차이를 보이는지 확인하고자 하였다. 최종적으로 예측변인의 수, 소수집단 정보를 반영하는 방법, 심층인공신경망 모형 종류에 따른 기초미달 학생 예측성능을 확인하여 최종 예측모

형을 확인하였다. 이를 통해 교육패널자료를 활용하여 선제적인 도움이 필요한 기초미달 학생 예측을 위한 심층인공신경망 모형을 탐색해보고자 하였다.

2. 연구문제

이 연구에서는 교육패널자료를 활용하여 소수집단에 속하는 기초학력 미달 학생 예측을 위한 심층인공신경망 모형을 탐색하는 데 목적이 있다. 이를 수행하기 위하여 설정한 연구문제는 다음과 같다.

연구문제 1. 기초학력 미달 학생을 예측하기 위한 심층인공신경망 모형에서 예측변인으로 활용가능한 주요 변인은 무엇인가?

연구문제 2. 기초학력 미달 학생을 예측하기 위한 심층인공신경망 모형은 무엇이며, 그 성능은 어떠한가?

2-1. 예측변인의 수에 따라 예측모형의 성능은 어떻게 나타나는가.

2-2. 소수집단에 속하는 기초학력 미달 학생의 정보를 훈련 모형에 반영하는 방법에 따라 예측모형의 성능은 어떻게 나타나는가.

2-3. 심층인공신경망 종류에 따라 예측모형의 성능은 어떻게 나타나는가.

Ⅱ. 이론적 배경

1. 교육연구에서 예측에 대한 논의

교육연구에서는 학생의 교육적 성장을 도모하기 위하여 인지적·정의적 영역의 향상 또는 변화를 예측하고, 그 과정에 영향을 미치는 교육 맥락을 살펴보고자 한다. 이를 위하여 국가수준 학업성취도를 비롯하여 한국교육종단연구, 서울교육종단연구 등 국가 또는 시도 차원에서 교육 자료를 지속적으로 수집하고, 교육연구에 적극 활용한다. 이러한 대규모 교육패널자료에는 학생이 경험하는 학교 내외 교육활동 및 가정배경 등의 정보가 포괄적으로 포함되어 있으며, 매년 주기적으로 학생을 추적조사하기 때문에 학생의 변화 모습과 경향성을 확인할 수 있다(김양분 외, 2016). 이에 다수 연구에서 교육패널자료를 활용하여 학생의 학업성취도(강은영, 김진구, 2018; 김양분 외, 2016; 시기자 외, 2015; 임현정 외, 2017) 및 정의적 특성의 변화(김경희 외, 2013; 이현철 외, 2015)를 예측하고 이에 영향을 미치는 요인을 다양한 관점에서 살펴보는 모습을 볼 수 있다.

이와 같은 예측연구들은 위계적 선형모형(Hierarchical linear model), 다층성장모형(Multilevel growth model)과 같이 통계적 추론(statistical inference)에 기초한 통계방법을 적용하여 예측변인 값에 따른 종속변인의 변화를 살펴보는 특징이 있었다. 즉, 전통적 통계분석을 활용한 예측연구에서는 선행연구 연구결과를 바탕으로 주어진 자료를 간명하면서도 효과적으로 표현하는 예측방정식을 구한 후, 이 방정식에 예측변인의 값을 대입하여 종속변인의 값을 구하고자 하였다(한국교육평가학회, 1998).

한편, 이와는 다른 맥락에서 통계적 추론에 기반을 두지 않고 주어진 자료 내 패턴을 학습하여 예측모형을 상세화하며, 개별 학생 예측 정확도를 높이는 데 초점을 둔 예측연구도 이루어지고 있다. 주로 기계학습

(Machine learning) 방법을 활용하여 예측 정확도를 높이는 예측모형을 탐색하고자 하는 연구로서, 예측변인과 종속변인 간 관계를 통계적으로 검증하는 데에는 상대적으로 관심이 적은 모습을 볼 수 있다. 예를 들어, 이창묵 외(2020)의 연구에서는 국가수준 학업성취도자료에 인공신경망 종류 중 하나인 순환신경망 모형을 적용하여 학업저성취 학생을 예측하는 모형을 구축해보고자 하였다. 여기서의 예측은 주어진 자료를 활용하여 변인들 간 관계와 패턴을 모형화하고, 새로운 자료를 예측모형에 적용하였을 때 결과 값을 계산해내는 것을 의미하며, 통계적 검증을 통한 결과의 일반화보다는 개별 자료의 예측 결과에 더 초점을 둔다. 여러 기계학습 방법 중 심층인공신경망(Deep neural network) 모형은 모형을 유연하게 설정할 수 있고 다양한 데이터에 활용도 가능하다는 점에서 적용 가능성이 높으나, 그 과정이 블랙박스로 알려져 있으며 모형 내에서 예측변인 중요도를 확인할 수 있는 알고리즘이 없어 교육현장에서는 활발하게 활용되지 않는 모습이다.

이처럼 두 맥락의 예측연구는 연구의 초점이 다소 상이하며, 그에 따라 연구를 상세화하는 과정이나 논의의 방향도 다른 모습을 볼 수 있다. 따라서 심층인공신경망 모형을 활용한 기초학력 미달 학생 예측연구에 앞서, 전통적 통계방법을 적용한 예측연구와 기계학습 방법을 활용한 예측연구의 관점 및 특징을 비교 및 정리하고, 교육에서의 예측연구의 강조점과 예측의 상황, 예측에 활용할 수 있는 자료의 특징 등을 살펴보고자 한다.

전통적인 통계분석 방법을 활용한 예측연구와 심층인공신경망 모형과 같은 기계학습 방법을 적용한 예측연구 모두 다양한 예측변인을 활용하여 학생의 인지적・정의적 영역의 변화를 예측하지만 분석에서의 강조점에서 차이가 있다(유진은, 2019; 이창묵 외, 2020). 전통적인 통계분석에서는 종속변인을 간명하고 효과적으로 설명하는 교육 변인들이 무엇인지 탐색함으로써 학생의 성장을 위해 어떠한 면에 중점을 두어야 하는지 살펴보고, 이를 교육과정 운영이나 교육정책 수립 결정에 참고하고자 한다. 따라서 예측모형을 설정할 때에는 탄탄한 이론적 배경을 기반으로

한 관심변인을 중심으로 종속변인 변화에 미치는 영향을 통계적으로 검증하며, 개별 학생에게 주목하기보다는 주어진 자료에 나타난 전반적인 경향성을 확인하고자 한다(안영은, 2019; 정혜경, 2013; 허유성 외, 2014 등). 전통적인 통계분석이 기본적인 가정을 충족하면 예측에도 효과적으로 기능하는 연구방법임에도 정확한 예측보다는 그러한 예측을 이끌어내는 변인에 대한 연구가 많은 이유이기도 하다.

반면, 기계학습 방법에서는 종속변인과 예측변인 간 관계보다는 개별 사례를 정확히 예측하는 데 초점을 두며, 이에 따라 종속변인을 정확하게 예측하기 위한 모형을 탐색하는 접근을 취한다. 전통적인 통계분석에서 통계적 가정과 이론적 배경을 바탕으로 예측변인과 종속변인 간 관계를 잘 설명하는 예측모형을 설정하는 것과는 달리, 입력한 자료를 바탕으로 예측 알고리즘을 학습시킨 후 이 예측모형을 새로운 사례에 적용하였을 때 제대로 기능하는지를 확인한다. 이때 예측변인과 종속변인의 관계를 살펴보고자 하지 않으므로 예측모형의 이론적 틀이 중요하게 다뤄지지 않는 경우가 많고(유진은, 2019), 상대적으로 복잡하고 비효율적인 예측모형을 허용하는 경향도 확인할 수 있다(이창묵 외, 2020).

이에 따라 교육 분야에 실질적으로 연구를 활용하는 측면을 고려하였을 때, 교육연구에서는 기계학습 맥락의 예측연구보다는 전통적인 통계분석 맥락의 예측연구가 더 유용했을 수 있다. 교육 외 여러 분야에서는 질병의 발병, 고객의 선호도 및 서비스 이탈 여부, 상품의 수요, 기업의 부도 등을 예측하는 예측모형을 구축한 후, 새로운 환자, 고객 또는 기업의 정보를 토대로 미래의 결과를 예측하는 것이 실질적으로 필요하다. 하지만 상대적으로 교육 분야에서는 예측모형을 구축하고 새로운 자료를 예측해야 하는 필요성이 크지 않은 면이 있다. 학생의 교육적 변화 속도는 수년에 걸쳐 나타나고, 개인적, 가정적, 환경적 변인이 복합적으로 영향을 미친다는 특징이 있기 때문에, 예측보다는 현재 상황이 나타난 맥락과 원인을 진단하고 이를 바탕으로 한 교육적 지원을 모색하는 것이 요구되었다. 따라서 교육연구에서의 예측은 예측의 정확성 확인이나 예측률을 높이기 위한 모형의 탐색보다는 변인 간 관계를 확인할 수 있는

관점의 예측연구가 더 활발하게 이루어진 모습을 볼 수 있다.

따라서 학생들을 정확하게 예측하고 그 결과에 따라 문제 상황이 발생하기 이전에 개별적 지원을 제공할 수 있도록 하기 위한 연구는 학습부진 학생, 중도이탈, 정서적 문제 행동의 발병과 같이 사후적 접근보다 예방적 차원의 접근이 필요한 학생을 대상으로한 일부 연구에서 다뤄지고 있다(이창묵 외, 2020; 이현진, 2018; 조헌국, 2018; Barros et al., 2019; Chaplot et al., 2015). 하지만 예측 과정에 대한 정보를 제공해주지 못하기 때문에 예측 결과를 활용할 때 불안감을 해소하지 못한다는 점에서(최재식, 2019), 예측시스템을 구축하고 예측 결과를 바탕으로 교육 지원을 하는 등 실질적으로 활용하는 데에는 적극적으로 적용되지 못하고 있는 모습이다.

예측연구 결과를 교육현장에서 활용하는 측면에 더하여, 예측연구에 활용할 수 있는 분석자료의 종류에서도 교육연구에서의 예측연구 특징을 살펴볼 수 있다. 교육 분야에 예측연구를 살펴보면, 교육패널자료를 활용한 연구를 다수 확인할 수 있었다. 이는 교육패널자료에는 학생이 학습 경험과 관련된 다양한 교육 변인의 정보가 있으며, 학생의 변화를 확인할 수 있는 자료이기 때문이다. 대부분의 교육패널자료는 모집단으로 상정한 학생들 중 일부를 표집하고 이 학생들의 정보를 토대로 전체 학생의 발달상황을 확인하고자 하는데, 이에 따라 교육패널자료를 활용한 예측연구에서는 통계적 검증을 통해 자료에 나타난 특징을 모집단 수준에서도 타당하게 해석할 수 있는지 확인하고자 하는 모습이 나타났다. 또한 교육패널자료가 학생들의 발달 수준을 확인하고 이에 따른 교육적 시사점을 도출하는 데 목적을 두고 수집된 자료라는 점에서(임현정 외, 2017), 예측력이 높은 모형을 구축한 후 새로운 사례에 예측모형을 적용하기 보다는 변인 간 관계를 통계적으로 검증하고 모집단 학생들에게서 나타나는 특징이나 현상을 살펴보고자 하는 연구가 주로 이루어졌다.

상대적으로 기계학습은 변인 간 관계에 대한 이해를 토대로 예측모형을 설정하기 어려운 상황에서 활용성이 높은 방법이다. 교육 분야에서 변인 간 관계를 연구자가 확인하기 어려운 비정형 자료로는 수행평가 및

서술형 평가 자료, 온라인 수업의 로그자료, 수업 상황에서의 음성자료 등을 생각해볼 수 있으나, 현재 실질적으로 연구에 활용할 수 있을 정도로 축적되거나 정제되지 않은 상황이다. 이로 인해 기계학습을 활용한 예측연구는 다소 활발하게 이루어지지 않은 면이 있다. 하지만 교육패널 자료를 활용하는 경우에도 예측변인과 종속변인 간 선형성을 가정하기 어렵거나, 예측변인 간 관계가 독립적이지 않거나, 다중공선성 문제가 발생할 수 있는 경우 등 전통적인 통계분석 방법을 적용하기 어려운 상황에서 특정한 가정 없이 주어진 자료 자체의 패턴을 파악하는 기계학습 방법이 적용되는 모습이 나타나고 있다(손윤희 외, 2020; 이창묵 외, 2020; 홍유정, 박현정, 2021).

이상의 내용을 종합하면, 교육연구에서는 학생의 변화에 영향을 미치는 맥락 변인을 통계적으로 추론 및 검증하고 이를 토대로 학생들을 이해함으로써 교육과정 및 정책운영 결정에 참고하고자 하는 예측연구가 주로 이루어지고 있으며, 상대적으로 개별 학생을 정확하게 예측하는 관점에 연구는 예방적 지원이 필요한 일부 분야에서 소수의 연구가 진행되고 있음을 확인할 수 있었다. 이는 예측결과를 실제 교육현장에 활용할 때 학생들의 발달적 특징을 파악할 수 있는 전자의 예측연구 관점이 더 유용했을 수 있으며, 예측연구에 활용할 수 있는 교육자료 또한 변인 간 관계를 확인하는 데에 더 적절하였기 때문이다.

하지만 교육자료를 체계적으로 수집하는 목적이 학생들의 특징을 파악하는 것과 함께 학교 교육에서 학생들이 겪는 어려움을 파악하고 적기에 필요한 처방을 제공하는 것임을 고려할 때, 현재 가용할 수 있는 정보를 활용하여 학생들의 향후 변화를 정확하게 예측하는 연구 역시 매우 중요하다고 할 수 있다. 특히 기초적인 학업 수준에도 미치지 못하는 기초학력 미달 학생이나 학교 적응에 심각한 어려움을 겪는 학생들 같은 경우에는 학년을 거듭할수록 그 정도가 심화된다는 점을 고려할 때 개별 학생을 예측하는 시스템이 구축된다면 효과적인 교육적 지원을 제공할 수 있을 것이다.

2. 기초학력 미달 학생 연구

우리나라는 2003년부터 국가수준 학업성취도 평가를 실시하여 국가수준의 학력을 진단하는 시스템을 구축하고, 성취수준을 우수학력, 보통학력, 기초학력, 기초학력 미달과 같이 4개 수준으로 구분하여 제시하였다. 여기서 기초학력은 '평가 대상 학년급 학생들이 성취하기를 기대하는 기본 내용을 부분적으로 이해하는 수준'으로 정의된다(김성숙 외, 2010, p21). 이에 따라 부분적 이해에도 못 미치는 수준의 학생을 기초학력 미달(이하 기초미달) 학생이라 정의할 수 있다.

교육 연구 및 학교 현장에서는 기초미달 학생과 학습부진 학생을 유사한 의미로 사용하고 있다(김태은 외, 2021, p.506). 이는 학습부진이 기대되는 성취에 비해 실제 성취가 현저하게 낮은 상태(김정아, 2021; Reis, 2007, p.74 재인용)를 의미하기 때문일 것이다. 공통적으로는 학업적으로 낮은 성취를 보이는 학생 중에서도 전체 학생의 평균 수준이 아닌, 절대적으로 요구되는 기초적 수준 이하인 학생들을 가리킨다고 할 수 있다. 따라서 기초미달 학생을 기초학력에 못 미치는 수준의 학생이라는 의미에 더해, 학습부진 학생을 포함한 개념으로 사용하고자 한다.

이 절에서는 먼저, 국가 또는 지역 수준에서 수집된 대규모 데이터를 분석자료로 하여 양적연구 방법을 적용한 연구들을 중심으로, 기초미달 학생 연구의 주요 흐름을 살펴보고자 하였으며, 다음 소절에서는 기초미달 학생의 특성을 살펴본 연구를 검토하였다.

가. 기초학력 미달 학생 연구의 개관

기초학력 미달(이하 기초미달) 학생 연구는 분석 목적에 따라 기초미달 학생들을 대상으로 시행하는 정책의 효과를 확인하는 연구, 기초미달 학생의 학업성취 또는 정의적 영역에 영향을 미치는 요인을 살펴보는 연

구, 학업적 성취를 예측하고자 하는 연구로 나눌 수 있다.

첫 번째로 기초미달 학생을 대상으로 실시된 정책의 효과를 살펴본 연구들은 주로 각 정책이 시행되기 시작한 시기를 중심으로 이루어졌다. 분석 자료로는 대규모 패널데이터를 활용하여 정책 실시 여부에 따른 차이를 비교하는 모습을 확인할 수 있었다.

2000년대 초 학습부진 학생의 수준을 고려한 수준별 수업이 일선 학교에 적용되면서 수준별 수업의 효과를 살펴보기 위한 연구가 이루어졌다. 허유성, 남창우(2009)는 학교의 학습부진 프로그램과 수준별 이동수업이 저성취 중학생의 학업성취도와 자기조절학습능력 변화에 미치는 영향을 살펴보았다. 이를 위해 한국교육종단연구 1차에서 4차 자료를 활용하였으며, 4차년도의 점수를 기준으로 평균에서 1 표준편차 낮은 점수의 학생들을 저성취 집단, 1 표준편차 높은 집단을 고성취 집단으로 구분하여 집단별 효과를 비교하였다. 분석결과 학습부진프로그램은 오히려 저성취학생에 부적인 영향을 주고 있었으며, 수준별 이동수업은 유의미한 영향을 미치지 못하고 있음을 확인하였다. 이는 같은 자료의 다른 시기의 학생들을 대상으로 분석한 백병부(2011)의 연구결과와도 일치하는 부분이었다.

방과후학교 정책과 관련하여 이성은 외(2012)는 방과후학교 지원비의 기초학력향상 효과를, 정혜경(2013a)은 방과후학교 프로그램이 기초학력 미달 고등학생의 학업성취도에 미치는 효과를 연구하였다. 이성은 외(2012)의 연구에서는 전국의 초중고등학교 주요공시정보를 활용하여 방과후학교 지원비가 각 학교의 기초학력 미달 비율 감소에 효과가 있는지 살펴보았다. 방과후학교 지원비의 결과라 볼 수 있는 방과후학교 강좌수와 학생의 방과후학교 참여율을 독립변수로 하여 중다회귀분석을 실시한 결과, 학교의 방과후학교 강좌 수가 많을수록 학생의 방과후학교 참여율이 높을수록 학교의 기초학력 미달 비율이 감소하는 모습이 중학교는 국어와 영어, 고등학교는 모든 과목에서 나타났다. 정혜경(2013a) 연구에서는 국가수준 학업성취도 자료의 기초학력 미달 학생만 분석 집단으로 하여 군집 내 매칭 후 위계적회귀모형을 적용하여 방과후학교 효과

를 살펴보고자 하였다. 그 결과, 국어, 영어, 수학 모든 과목에서 방과후 학교에 참여한 학생이 그렇지 않은 학생에 비해 평균적으로 7~10점 이상의 높은 성취를 보이고 있음을 확인하였다.

또한 2009부터 시행되기 시작한 학력향상 중점학교 사업의 기초학력 효과를 살펴본 연구도 있었다. 김지은, 김지하(2012)는 전국 초중고등학교 주요공시정보, 국가수준 학업성취도 자료, 학력향상 중점 학교 사업비 집행내역 자료를 활용하여 학력향상 중점 학교 운영 여부와 사업비 규모가 학교의 기초학력 미달 비율에 미치는 영향을 확인하였다. 로지스틱 회귀분석과 중다회귀분석을 순차적으로 적용하여 사업 효과를 확인한 결과, 수학과 영어 과목에서 학력향상 중점학교 사업을 운영한 학교가 일반학교에 비해 기초학력 미달 비율을 낮출 확률이 더 높았으며, 1인당 사업비 규모가 클수록 미달 비율을 낮추는 정도가 더 크게 나타났다.

이 외에도 정혜경(2013b)는 학습부진학생 지도를 위한 전담교사 배치 정책의 효과를 살펴보고자 하였다. 이 연구에서는 국가수준 학업성취도 자료의 성취수준을 기준으로 기초학력 미달에 해당하는 학생만을 대상으로 정책효과를 확인하고자 하였다. 학생들의 중학교 사전 특성을 경향점수 매칭 방법으로 통제한 후, 위계적회귀분석으로 고등학교 시기의 정책효과를 확인한 결과, 전담교사 배치가 기초학력 미달 학생들의 학업성취도에 통계적으로 유의미한 영향을 미치지 않았다. 그러나 평균적인 효과는 유의미하지 않았으나 임의효과가 유의미한 점을 고려했을 때, 특정 학교에서는 전담교사 배치의 효과가 나타나고 있음을 추측하기도 하였다.

두 번째로 기초학력 미달 학생의 학업성취 또는 정의적 영역에 영향을 미치는 요인을 확인하고자 하는 연구들은 그 영향요인을 교육지원청, 학교, 교사, 학생 수준으로 다양하게 살펴보고 있었다. 일반적인 학업성취도 영향요인 관련 연구들은 학생 수준의 배경 변인 요인도 다방면으로 살펴본 것에 비해, 기초미달 학생을 대상으로 한 연구에서는 학교 및 교육지원청 수준에서의 영향요인을 탐색한 연구의 비중이 상대적으로 높았다. 이는 기초미달 학생을 위한 학교 및 교육지원청 수준의 제도적 지원

방안을 모색하려는 모습들이 반영된 것이라 짐작해 볼 수 있으며, 전체 학생에서 기초미달 학생의 사례가 매우 적어, ‘기초미달 학생 비율’과 같은 학교수준 변인을 활용했을 수 있다.

우선, 교육지원청 수준에서의 영향요인을 살펴본 연구를 살펴보면, 천세영, 이성은(2013)은 지리적 구분의 17개 시・도 교육청이 아닌 일정 수의 학교를 위한 교육 지원을 기준으로 구분된 178개의 교육지원청 수준에서 분석을 하였다. 국가수준 학업성취도의 보통학력 이상 도달비율 평균과 기초학력 미달이 없는 학교의 비율을 종속변인으로 하여 교육지원청의 교육력 효과를 확인하고자 하였다. 동시에 학교특성, 교사특성, 교육과정, 재정 특성에서 기초학력 향상에 영향을 미치는 변인들을 확인하였다.

다음으로 학교를 대상으로 학교 수준의 영향요인을 살펴본 연구이다. 권재기(2013) 연구에서는 성장혼합모형(Growth Mixture Model)을 활용하여 학교수준에서 종단적으로 기초학력 미달 비율의 변화를 탐색하였다. 학교 유형은 기초학력 미달 감소, 기초학력 미달 정체, 기초학력 미달 제로로 구분되었으며, 유형별 영향요인을 비교하였다. 그 결과 유형에 따라 영향요인이 다르게 나타났으며, 이를 토대로 기초학력 제고를 위한 정책적 제언을 하였다.

또한 이효정 외(2018)와 장선희, 정제영(2018)은 국가수준 학업성취도 자료를 활용한 위계적선형모형(Hierarchical Linear Model)을 적용하여 각각 중학교와 고등학교의 학교 관련 영향요인을 탐색하였다. 이효정 외(2018)은 학교 환경과 교장공모제의 영향력에 초점을 두고 연구를 진행하였으며, 분석 결과 기초학력 미달률에 학교특성 요인들은 대부분 유의하게 나타났고, 학교 변화에 영향을 줄 수 있는 교장공모제 또한 국어와 영어 기초학력 미달 비율에 긍정적인 영향을 미치고 있음을 확인하였다. 장선희, 정제영(2018)은 일반계고 기초미달 비율에 미치는 학교 교육활동 관련 요인을 탐색하였으며, 그 결과를 바탕으로 독서활동 장려, 학생 동아리 활동의 내실있는 운영, 방과후학교 운영 활성화 등의 정책적 제안을 하기도 하였다.

마지막으로 학생 수준의 자료를 활용하여 학생, 교사, 학교 수준의 영향요인을 살펴본 연구이다. 연구는 주로 위계적선형모형을 적용한 연구가 많았으며, 최근에는 기계학습 방법 중 하나인 랜덤포레스트(Random Forest) 방법을 활용한 연구도 확인할 수 있었다.

허유성 외(2014) 연구에서는 한국 아동・청소년패널조사를 활용하여 위계적선형모형을 적용하였으며, 중학생의 정서행동 문제와 학교환경 변인의 영향 정도를 성취수준 하위 25% 학생과 일반학생을 비교하였다. 그 결과 저성취 학생에게는 부정적 정서행동이 큰 영향을 미치고 있었으며, 수업참여 정도나 학교에 대한 가치, 교우관계 등이 긍정적인 영향을 주는 모습을 확인하였다.

안영은(2019)는 일반계고 학생의 최소 기준학력 도달에 미치는 학교 수준의 영향요인을 살펴보기 위하여, 기초이하(기초학력+기초학력 미달) 집단과 보통이상(보통학력+우수학력) 집단을 비교하였다. 학교의 교수학습 연수비율, 방과후 학교 참여율 등이 기준학력 도달에 긍정적이 영향을 주는 모습을 확인하였다. 이를 위해 국가수준 학업성취도 자료를 활용하였으며, 학생-시점-학교의 3수준 위계적선형모형을 적용하였다. 강은영, 김진구(2018)와 연구에서도 학교의 학습풍토가 저성취 학생들에게 영향을 주고 있는 모습을 확인하였다.

박미현, 허균(2021) 연구에서는 랜덤포레스트 방법을 대구교육종단연구 자료에 적용하여 기초학력 주요 예측 변인을 확인하였다. 4단계 성취수준을 기준으로 기초학력과 기초학력 미달 학생을 1로 보통학력, 우수학력을 0으로 재범주화 하였으며, 랜덤포레스트 예측모형의 변인 중요도를 확인하였다. 이와 유사한 연구로, 남나라(2021) 연구에서는 랜덤포레스트를 한국교육종단연구2013에 적용하여 주요 예측변인을 살펴보았으며, 문항단위의 변인을 활용하여 국어, 영어, 수학 과목별 예측요인을 상세하게 살펴보았다.

세 번째 연구 목적으로 기초미달 학생을 사전에 파악하기 위하여 학업성취 예측모형을 구축하는 데 목적을 두는 연구들이 있다. 학업성취 예측 연구는 통계적 추론(inference)에 근거한 분석방법을 적용한 연구와

통계적 추론 없이 자료의 특징을 학습하는 방법을 적용한 연구로 구분하여 살펴볼 수 있다.

전자의 연구에서는 대규모 교육자료가 학생이 학교에 내재(nested)하고 있음을 고려하여, 학업성취 예측을 위해 위계적선형모형을 활용하는 연구들이 많았다. 또한 학업성취도에 영향을 미치는 교육 변인들을 확인하는 과정에서 간접적으로 예측 연구를 시행하고 있는 모습이었다. 따라서 정책 효과성 검증 및 영향요인 연구의 분석 맥락과 매우 유사하다. 앞서 다루었던 연구 중 독립변인과 종속변인의 시점 차이를 두어 학업성취가 나타나기 이전의 정보를 독립변인으로 활용한 연구들이 이에 포함된다고 할 수 있다(강은영, 김진구, 2018; 남창우, 전훈, 2012; 백병부, 2010; 허유성 외, 2014).

후자의 연구는 기계학습(machine learning) 모형이 등장하면서 교육분야에 적용되기 시작하였으며, 여러 변인 간의 관계를 탐구하기보다는 학업적 어려움을 겪는 학생을 미리 선별할 수 있는 모형을 찾는 데 초점을 두는 모습이었다(이창묵 외, 2020). 이로 인해 전자의 연구에 비해 교육적 지원을 결정해야 하는 주체에게 교육맥락에 대한 정보를 제공하기에는 다소 어려운 면이 있었다.

기계학습을 활용하여 중등학생을 대상으로 학업성취 예측을 시도한 연구는 이창묵 외(2020) 연구가 있었다. 이창묵 외(2020)의 연구에서는 2011년에서 2016년 국가수준 학업성취도 세 시점 연계자료를 활용하여 개별 학생의 학업성취도 수준을 종단적으로 예측하고자 하였다. 학생의 성취수준은 국가수준 학업성취도 자료의 4단계 성취수준을 기준으로 기초학력과 기초학력 미달 학생을 1로 우수학력과 보통학력을 0으로 재범주화하였다. 분석방법으로는 전통적인 통계분석방법 중 종속변수가 이분변수인 경우에 적합한 위계적일반화선형모형(Hierarchical Generalized Linear Model)과 심층인공신경망 모형 중 하나인 순환신경망(Recurrent Neural Network) 모형을 각각 적용한 결과를 비교하였다. 그 결과 학업저성취 학생을 미리 예측하고 선제적 지원을 줄 수 있는 가능성을 확인하였다.

중등학생을 대상으로 한 연구는 아니지만 초등학생을 대상으로 회귀분석, 의사결정분석, 신경망 분석 기법들을 활용하여 효과적인 학습장애 진단 기법을 탐색하고자 한 연구나(김동일, 홍성두, 2009), 대학생을 대상으로 기계학습 분석 방법 중 하나인 XGBoost 방법을 활용하여 학습부진 위험학생을 조기에 예측하고자 한 연구(신종호, 최재원, 2019), k-최근접 이웃 알고리즘, 서포트벡터머신, 의사결정 나무, 랜덤포레스트, 그래디언트 부스팅, 인공신경망 6가지 방법으로 학습 성취수준을 예측하고자 한 연구들이 있었다(조헌국, 2018).

이상의 내용을 종합하면, 기초미달 학생 연구는 정책 효과를 검증하고 영향요인을 확인하는 연구가 중점적으로 이루어져 왔으며, 학업성취를 예측하는 목적의 연구는 상대적으로 많은 연구가 축적되지 않았다고 할 수 있다.

구체적으로 정책 효과 검증 및 영향요인 확인 연구는 학교 단위의 연구의 비중이 높음을 확인할 수 있었다. 이는 기초학력 정책 적극 시행되면서 정책적 판단을 위한 연구 자료가 필요했기 때문일 수 있으며, 동시에 학생들을 위한 지원을 학교 단위, 교육청 단위로 제공하고자 하는 관점이 반영되었기 때문일 수 있다. 또한 분석 자료 자체로는 기초미달 학생이 적은 비율 집단이기 때문에 분석 대상으로 삼기에는 자료가 부족했을 수 있으며, 통계적 기본 가정을 충족하기 어려워 통계적 검증을 진행하는 데 어려움이 있었을 것으로 보인다. 이로 인해 대부분 기초미달 개별 학생을 대상으로 하는 연구는 전수조사로 실시된 국가수준 학업성취도 자료를 활용한 연구가 많음을 확인할 수 있었다.

나. 기초학력 미달 학생의 특성

기초학력 미달(이하 기초미달) 학생을 지원하고자 하는 노력은 정책의 수립, 교수·학습 방법의 개발, 심리상담 기회 제공 등 다방면으로 이루

어져 왔다. 이는 공교육 과정을 거친 모든 학생들이 사회에서 요구하는 기초 수준의 학력을 갖추도록 하는 데 일차적 목표를 두고 있기 때문이다(김성숙 외, 2010). 하지만 기초미달 학생들은 인지적·정의적 특징 및 주어진 상황과 상호작용하는 방식이 일반학생과는 다른 특성을 보이고 있어, 이들을 일반학생과 구분하여 살펴보려는 시도가 이어지고 있다.

기초미달 학생 연구에서는 이들의 학습 결손이 누적되고 심화되는 특성이 있으며, 학생 스스로의 노력으로 학습 결손을 해결하기 어려울 수 있음을 이야기하는 경향이 있었다(김태은 외, 2017; 이화진 외, 2009; 허은영, 유현실, 2012). 또한 학습부진이 단순히 성적의 하락만을 초래하는 것이 아니라, 학습흥미를 저하시키고, 학교생활 전반에 대한 거부감을 강화하며, 부정적 자기개념을 형성하는 정서적 문제를 심화시킬 수 있다고 보고한다(김동일, 2001; 이대식, 황매향, 2011; 이화진 외, 2009; Butkowsky, 1980). 따라서 학습 결손을 초기에 발견할 수 있는 점검 체제와 학습 결손 원인을 파악하여 누적을 방지하고자 하는 접근이 필요함을 논하고 있었다(김지은, 김지하, 2012; 유미경, 엄준용, 2016; 임현정, 2009).

이와 함께 여러 기초미달 학생 연구들은 학습부진의 원인은 매우 다양하고 복합적이라는 점을 강조하고 있었다(이대식 외, 2010; 이화진 외, 2009). 4년에 걸쳐 학습부진 극복을 위한 핵심적 요인을 탐색한 질적 종단연구에서는, 개별 학생이 학습부진이 되는 개인적, 환경적, 상황적 여건에 대한 파악이 우선되어야 한다고 이야기하였다(김태은 외, 2017; 김태은 외, 2018). 학습성취가 낮다는 공통점을 제외하고는 개별 학생마다 학습에서 어려움을 겪는 이유와 특성, 성취수준이 모두 다르나, 이에 맞는 정확한 점검 및 개별 맞춤형 지원에 대한 분석이 체계적으로 이루어지지 않았다는 점을 토로하며, 학생들이 학습부진을 겪지 않기 위한 예방적 관점의 지원의 필요성을 논하였다.

전술한 바와 같이 개별 학생이 처한 문제 상황에 따라 정도의 차이는 존재하며 기초미달 학생에게 나타나는 일반적 특성이 개별 학생 모두에게 나타나는 것은 아니다. 하지만 기초미달 학생의 전반적 이해를 위하

여 이들의 일반화된 ‘경향성’을 살펴보고자 한다. 이를 위해 개인적 측면, 가정적 측면, 학교생활 및 학교 상황적 측면의 특성으로 나누어 살펴보고자 하며, 기초미달 학생의 특성과 함께 각 측면과 관련된 학업 성취 영향요인도 확인하고자 한다.

1) 개인적 특성

개인적 측면에서의 기초미달 학생 특성은 다시 인지적 특성과 정서적 특성으로 구분할 수 있다. ‘인지적 특성’은 학습 내용을 이해하고 기억하는 능력, 자신의 수준 및 상태에 대한 지각, 학습상황에서의 인지적 조절과정 전반을 포괄적으로 의미하는 개념이다(이대식, 황매향, 2011; 조난심 외, 2009). 기초미달 학생들은 학습능력이 일반학생에 비해 상대적으로 낮으며, 이에 따라 학습 속도가 매우 느린 특성을 보인다(김태은 외, 2017). 이로 인해 기초미달 학생들은 교과과정을 학습하고 이해하는데 많은 시간이 소요되며, 일반적인 수업을 따라가지 못하여 학습부진으로 이어지게 된다(백병부, 2010; 이대식, 황매향, 2011).

기초미달 학생의 학습전략 인식 및 활용을 연구한 선행연구들은 기초미달 학생들이 학업 고성취 집단에 비해 다소 비효율적인 학습 방법을 활용하고 있다고 보고하고 있다(전명남, 양명희, 2015; 최효식 외, 2013). 학업성취가 높은 학생들은 기초미달 학생보다 다양한 학습전략을 사용하고 있으며, 많은 경우에는 2배 이상의 전략을 사용하는 모습을 보이기도 하였다(Zimmerman &Martinez-Pons, 1986, 1988). 일부 연구에서는 기초학력 미달 수준의 중학생들이 성실도, 직업가치, 진로성숙도, 진로 효능감에서는 일반 중학생과 비슷한 수준이나, 자기조절 학습전략이 전반적으로 낮게 나타나는 모습도 나타났다(전명남, 양명희, 2015).

기초학력 미달 학생들은 학습 상황에서의 인지 능력 뿐만 아니라 진로에 대한 이해도 및 사회 구성원으로서의 책임의식 등도 일반 학생에

비해 낮게 나타났다. 전명남(2017) 연구에서는 국가수준 학업성취도평가에서 기초학력 미달로 분류된 19명의 중학생을 대상으로 질적연구를 진행하였다. 기초미달 학생들은 진로에 대한 생각은 있으나, 자신의 능력에 비해 터무니없이 높거나 낮은 포부수준을 보이는 모습이 나타나, 막연한 진로신념과 계획을 지니고 있음을 확인할 수 있었다.

다음으로, '정서적 특성'은 개인이 가지고 있는 긍정적 또는 부정적인 마음의 상태를 의미하며, 학생의 자신감, 성취감, 정서조절 능력, 스트레스에 대한 반응 등을 포괄하는 개념이다(김동일, 2001, 김태은 외, 2017). 기초학력 미달 학생과 같이 학습에 어려움을 겪고 있는 학생들은 학업을 수행하는 과정에서 좌절, 실패 등 부정적 경험을 지속적으로 겪어왔기 때문에, 부정적 자아개념이 형성된 경우가 많다. 이에 따라 자존감 또한 일반학생에 비해 낮은 것으로 보고된다(백병부, 2010; 조난심 외, 2009). 또한 일상에서도 의기소침하거나 심리적으로 위축되어 있는 모습을 보이기도 하였다(이대식, 황매향, 2011).

전명남, 양명희(2015) 연구에서는 기초미달 중학생은 일반 중학생에 비해 학업정서에서 긍정정서는 낮은 반면, 부정정서가 높음을 확인하였고, 시험불안과 무기력 점수가 통계적으로 유의하게 높게 나타났다. 이는 기초미달 중학생을 대상으로 3개월에 걸쳐 심층면접을 진행한 전명남(2017) 연구에서도 유사하게 나타났다. 이러한 정서적 특성은 특히 학업과 관련된 영역에서 더 두드러지게 나타나고 있으며(백병부, 2010; 임현정, 2009), 일부 학생들은 학습 영역에서 자신의 수준에서 충분히 해결가능한 과제조차 타인의 도움을 구하는 과잉 의존성을 나타내는 모습도 확인되었다(김선 외, 2001)

또한 기초미달 학생들은 학습 상황이 아닌 일상생활에서의 우울, 스트레스가 일반학생에 비해 높은 반면 행복감은 상대적으로 낮게 나타나기도 하였다(송연주 외, 2014; 허유성 외, 2011). 성취수준 하위 25%인 중학생의 특징을 연구한 허유성 외(2014) 연구에서는 학업 저성취 학생들이 심각한 정서・행동문제를 드러내지는 않더라도 일반학생들에 비해 정서・행동 문제를 경험할 가능성이 높다는 점을 지적하기도 하였다.

다음으로 개인적 특성과 관련된 학업성취 영향요인 연구를 살펴보면, 성별, 독서능력, 학습성향과 관련된 변인 및 학업과 관련된 정서적 요인들이 있었다. 남학생이 여학생에 비해 기초미달로 분류될 가능성이 높은 경향을 보였으며(김태은 외, 2012; 임현정, 2009), 독서능력이 기초학력 미달 확률에 영향을 미치고 있었다(임현정, 2009). 이는 독서능력은 기초적인 학습능력과 밀접한 관련이 있기 때문이며, 수학과 과학 기초미달 학생에게도 통계적으로 유의한 영향을 미치는 것으로 나타났다.

학습 성향 및 정서적 요인으로는 백병부(2010) 연구에서 교과목 효능감 및 학업자아개념이 기초미달 학생의 학업 성취 향상에 유의미한 영향을 미치는 모습을 확인하였다. Bryan 외(2004)와 Mercer(2010)의 연구에서는 학습할 때 느끼는 지루함, 답답함, 실망, 좌절 등의 부정적인 정서는 그 자체로도 문제가 되지만, 기억력이나 정보처리와 같은 인지적 사고력 수행을 어렵게 할 수 있음을 강조하며 정서적 영역이 인지적 영역에 미치는 영향을 보여주었다. 또한 김동일(2005), 신종호(2002), Swanson(2011)의 연구에서는 인지적 영역에 초점을 맞춘 학습전략을 습득함으로써 학습 동기와 같은 정의적 영역의 변화를 촉진할 수 있음을 보여주기도 하였다.

2) 가정적 특성

기초미달 학생들은 가정적 측면에서도 일반학생과 상이한 특성이 있었다. 기초미달 학생들은 구조적 결손이 있는 가정에서 생활하거나, 경제적으로 어려움을 겪고 있는 경우가 많다고 보고되었다. 이와 관련하여 김경근 외(2007)는 서울 시내의 6개 초등학교에 재학 중인 학습부진학생 11명을 심층면담한 결과 이들 대부분이 한부모 가족 또는 그에 상응하는 가족구조에 노출되어 있다고 보고하였으며, 임현정(2009)은 전국의 초등학교 6학년을 모집단으로 하여 무선표집한 데이터를 분석한 결과, 기초

미달 집단은 기초미달 이상의 집단에 비해 상대적으로 가계소득이 적었다고 보고하였다. 또한 기초미달 학생들은 일반 학생들에 비해 부모와의 문화활동이나 교육적 지원, 부모와의 상호작용이 더 적은 경향이 있었으며(임현정, 2009), 이는 학교 현장에서 기초미달 학생을 지도하는 교사들의 인식과도 일치하는 부분이었다(조난심 외, 2009).

가정적 측면과 관련된 영향요인 또한 앞서 언급한 가정적 특성들과 관련이 높았다. 김태은 외(2012) 연구에서는 초등학생을 대상으로 기초학력 미달이 될 확률에 영향을 미치는 변인들을 탐색한 결과, 결손가정의 학생일수록, 가정에서의 돌봄이 부족할수록 기초학력 미달이 될 확률이 유의하게 높게 나타났다. 백병부(2010) 연구에서도 부모의 학업지원 정도가 중학생의 기초학력 미달 여부를 결정하는 주요 요인이며, 수학과 영어 교과에서 유의미한 영향을 미치고 있음을 확인하였다.

그러나 일부 연구에서는 일반적인 학업성취도 연구와는 달리 부모의 사회·경제적 지위가 기초학력에 도달하는 데 미치는 영향이 크지 않은 모습도 확인되었다. 국가수준 학업성취도 자료를 활용하여 기초미달 학생의 영향요인을 살펴본 임현정(2009) 연구에서는 부모의 사회·경제적 지위가 높을수록 기초미달 확률이 낮아지는 경향이 읽기에서만 약하게 나타났고, 수학과 영어에서는 유의미한 영향이 없었다고 보고하였다. 상대적으로 문화적 활동, 부모와의 상호작용, 부모의 자녀에 대한 교육포부가 모든 과목에서 유의미한 영향요인으로 확인되었다. 이에 비추어 기초미달 학생에게 영향을 미치는 가정 요인의 양상이 일반적인 학업 저성취 학생과도 다를 수 있는 가능성을 제시하기도 하였다.

3) 학교 생활 및 학교 특성

기초미달 학생들은 학교 생활 모습에서도 일반학생과 차이를 보인다고 보고된다. 수업상황에서 참여도가 낮으며, 수업준비가 상대적으로 미

흡하고, 적극적이고 주도적인 모습이 부족한 면이 있다. 일부 학생들은 반항적 행동이나 잠을 자는 것과 같은 수업 거부 행동을 보이며 부적절한 학습행동을 보이기도 한다(김태은 외, 2018; 이대식, 황매향, 2011).

또한 기초미달 학생들의 비율이 높은 학교에서도 특징적인 모습이 나타났다. 기초학력 미달 학생 비율로 선정되는 학력향상중점학교의 특성을 확인한 이화진 외(2010) 연구를 살펴보면, 학력향상중점학교로 지정된 학교들이 일반학교보다 기초생활수급자, 다문화 가정 학생, 중식지원학생, 결손학생 비율이 높은 것으로 보고되고 있었다. 3년에 걸쳐 기초미달 비율의 변화를 학교수준에서 살펴본 권재기(2013) 연구에서도 비슷한 양상이 나타났음을 확인할 수 있었다.

기초미달 학생에게 영향을 미치는 학교생활 및 학교수준 변인과 관련된 영향요인 또한 기초미달 학생 및 기초미달 비율이 높은 학교의 특성과 연관성이 높았다. 기초미달 학생이 수업분위기를 좋게 인식하고, 수업태도가 좋을수록 기초학력 미달에 속하지 않을 확률이 높게 나타났다(김태은 외, 2012; 이화진 외, 2009; 임현정, 2009). 또한 수업참여도 및 학교 가치에 대해 긍정적으로 인식할수록 학습을 저해하는 정서적・행동적 문제가 유의미하게 낮아졌으며, 주의집중 문제도 함께 개선되었다(허유성 외, 2014). 학생이 인식한 학교의 학습문화가 협력적인 경우 기초학력 미달에 속할 확률을 낮추는 반면, 경쟁적 학습문화인 경우에는 그 반대로 나타나는 모습도 확인되었다(임현정, 2009).

학교 수준에서는 교사의 학력지도에 대한 직무연수를 받을수록, 보조교사가 많이 투입될수록 학교의 기초학력 미달 확률이 감소하는 모습이 나타났다(이화진 외, 2011). 또한 공립학교에 비해 사립학교, 읍면지역 학교에 비해 대도시에 소재하는 학교가 학교의 기초학력 미달률이 낮았고, 학교 크기가 클수록 미달률이 낮아지는 특징을 보였다(이효정 외, 2018).

그러나 일부 연구에서는 학교환경 요인이 기초미달 학생에게 통계적으로 유의미한 영향을 미치지 않거나, 영향 정도가 매우 적게 나타나기도 하였다(강은영, 김진구, 2018; 곽현주 외, 2016; 허유성 외, 2014). 곽현주 외(2016) 연구에서는 국가수준 학업성취도 중학교 자료를 활용하여

학교의 우수학력 비율, 기초미달 비율에 영향을 미치는 학교 변인들을 탐색하고자 하였다. 그러나 학교의 저성취 수준에 유의하였던 영향요인으로는 학교의 특성이라기 보다는 자기주도학습, 방과후학교 참여, 학부모 지원과 같은 학생 및 학부모 기반의 변인들이었으며, 학교 변인으로는 맞춤형 개별화 프로그램 운영이 영향을 미치고 있었다. 개별 학생을 대상으로 학습풍토가 수학 학업성취에 미치는 영향을 발펴본 강은영, 김진구(2018)의 연구에서도 학교의 학습풍토가 학업성취에 미치는 영향은 학력이 낮을수록 적어지는 모습을 확인하였고, 이를 토대로 바람직한 학습풍토만으로는 학업성취의 향상을 도모하기 어려울 수 있음을 논하기도 하였다.

3. 심층인공신경망을 활용한 예측모형

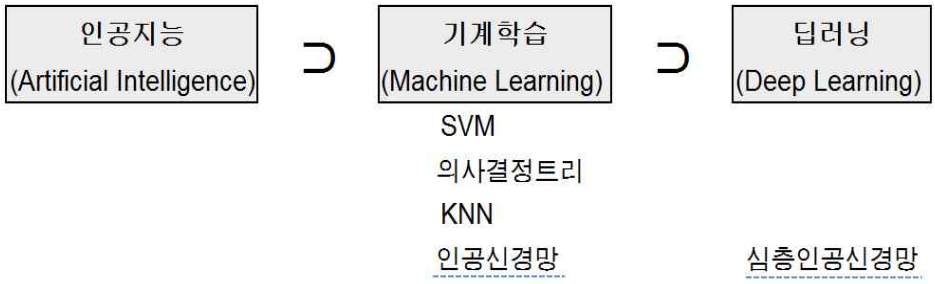
인공신경망(Neural network)은 인공지능 및 기계학습의 발전과 맥을 같이하며, 2010년대를 전후로 심층인공신경망(Deep neural network)으로 확장되었다(Hinton et al., 2006). 심층인공신경망은 사용하는 맥락에 따라 인공신경망, 심층인공신경망, 딥러닝 등 다양한 용어로 사용되고 있으며, 이 외에도 인공신경망, 인공지능, 기계학습 등의 용어와 혼재되어 사용되고 있다. 따라서 관련 용어 간의 포함관계와 함께 각 용어의 의미를 살펴보며 심층인공신경망의 의미를 명확하게 정의한 후, 심층인공신경망을 활용한 예측모형을 살펴보고자 한다.

인공지능(Artificial intelligence)은 사람의 지능을 모방하여 사람이 하는 것과 같이 복잡한 일을 할 수 있는 기계를 만드는 것을 말한다. 인간의 지적능력을 컴퓨터로 구현하고자 하는 기술을 포괄하는 개념이며, 이러한 인공지능을 구현하는 방법 중 중요한 방법이 기계학습이다.

기계학습(Machine learning)은 "컴퓨터가 명시적으로 프로그램되지 않고도 학습할 수 있도록 하는 연구 분야"로 정의된다(Samuel, 1959). 여기서 학습이란 "만약 어떤 작업(task)에서 경험(experience)을 통해 성능(performance)이 향상되는 것"을 의미하며, 이것이 기계학습의 핵심이다(Mitchell, 1999). 입력값과 함께 결과값(lable)을 같이 주고 학습시키는 방법으로, 서포트벡터머신(Support Vector Machine), 의사결정 트리(Decision Tree), 회귀모델 등이 있으며, 결과 값 또는 정답이 없이 입력값만을 이용하여 학습을 시키는 K-최근접이웃(k-nearest neighbor)도 기계학습에 포함된다.

인공신경망(Artificial neural network)은 신경세포(neuron)의 작동 원리를 가상으로 구현한 모형으로, 기계학습에 속하는 여러 방법 중 하나이다. 일반적으로 은닉층(hidden layer)이 있는 다층퍼셉트론(Multi-layer perceptron) 모형을 인공신경망이라 칭하는데, 이 다층퍼셉트론을 기반으로 하는 인공신경망에서 은닉층의 수를 늘린 모형을 심층인공신경망

(Deep neural network)이라 한다. 여기서 심층(deep)의 의미는 신경망의 층이 많고 각 층마다 고려되는 변수가 많다는 것을 의미하며, 딥러닝(Deep learning)이라고도 불린다(김의중, 2016; Hinton et al., 2006). 은닉층이 여러 개인 심층인공신경망을 구분하여 지칭하는 이유는 층의 깊이가 깊어지면서 훨씬 복잡한 모형을 구현할 수 있게 되었기 때문이다. 이는 자료의 비선형적이고 복잡한 패턴을 파악하는 데 용이함을 의미한다.



[그림 II-1] 인공신경망 관련 개념들 간 관계

이와 같이 심층인공신경망 모형은 자료 자체의 패턴을 포착하여 자료를 분류하거나 예측하는 데 강점이 있는 모형이다. 또한 자료의 특성에 따라 층의 개수와 구성을 달리 하여 다양한 형태로 발전하고 있다. 따라서 이 절에서는 심층인공신경망 예측모형을 구축하기 위해 선행되어야 하는 예측변인 선정과정을 살펴보고, 심층인공신경망 모형의 기본적 개념과 작동원리를 정리해보고자 한다. 아울러 심층인공신경망의 대표적인 모형인 다층퍼셉트론 모형과 합성곱신경망 모형에 대해서도 살펴볼 것이다.

가. 예측변인의 선정

심층인공신경망 분석 방법은 기존 통계분석 방법에 비해 기본적인 통계적 가정 및 변수 간 다중공선성(multicollinearity) 문제로부터 자유로운 편이기 때문에, 많은 변인를 예측변인으로 활용할 수 있다는 특징이 있다. 하지만 종속변인과 직접적인 관련이 없는 예측변인이 많아지면, 일정 수준 이상의 모형을 구축하기 위하여 상당히 많은 훈련자료(training data)가 필요하거나, 오히려 모형 성능이 낮아지는 문제가 발생하기도 한다(García et al., 2016; Snieder et al., 2019). 이에 심층인공신경망 모형을 구축하는 과정에서는 모형에 활용할 수 있는 변인 중 종속변인과 관련이 높은 변인을 선정하고자 하며, 이를 변인선택(Variable selection) 또는 특징선택(Feature selection)이라 한다.

변인선택의 초점은 모형 성능은 우수한 수준을 유지하면서, 입력자료를 효율적으로 설명할 수 있고 잡음(nois) 또는 관련 없는 변인의 영향을 줄이는 데 있다(Guyon & Elisseeff, 2003). 교육패널자료에는 수백 개의 변인이 있으며, 변인 중에는 서로 상관관계가 상당히 높은 변인들이 존재한다. 예를 들어, 진로성숙도와 관련된 5개의 문항이 있는 경우, 4개 문항은 상당히 유사한 속성을 나타내는 반면, 나머지 하나는 상대적으로 다른 속성에 대한 정보를 제공할 수 있다. 이때 4개 문항 중 종속변인과 가장 관련이 높은 한 문항만으로도 데이터를 설명하기에 충분할 수 있으므로 진로성숙도 5개 문항 중 2개 문항을 활용하여 예측모형을 구축할 수 있을 것이다. 이처럼 종속변인과 관련이 높은 문항을 선정하려면 종속변인과의 관련성을 측정할 수 있는 기준이 필요하게 된다.

변인은 선정하는 방법은 Wrapper, Filter, Embedded 방법으로 구분할 수 있으며, 이중 Wrapper 방법은 예측모형에 의존하여 변인을 선정하는 반면, Frilter와 Embedded 방법은 예측모형의 종류에 상관없이 변인을 선정한다는 특징이 있다(Cai et al., 2018; Li te al, 2017).

Wrapper 방법은 주어진 변인의 모든 조합을 활용하여 모형 성능을

확인한 후, 그중 모형 성능이 가장 높은 변인 집합을 선택하는 방식이다. 따라서 $2^{(\text{변인의 수})}-1$개 집합을 확인해야 하기 때문에 계산시간이 상당히 많이 걸린다. 하지만 모형 성능을 확인할 최적의 모형이 정해지지 않은 경우에는 변인 선정이 어려울 수 있으며, 예측모형을 구축하는 상황에서는 종속변인을 알지 못하는 상황을 가정하므로 Wrapper 방법을 적용하는 데에는 한계가 있을 수 있다.

Filter 방법은 변인과 종속변인 간 통계적인 분석을 통해 관련성 정도를 측정하고 관련성이 높은 변인을 선정하는 방법이다. 상관계수, χ^2 검정 통계량에 기반하여 순위를 부여하는 방법이 있으며, 각 변인이 가지는 고유한 상호정보량(Mutual information)을 등을 바탕으로 변인을 선정하기도 한다. FIlter 방법은 계산 과정이 간단하다는 장점이 있으나, 각 변인과 종속변인 간 관계를 확인하기 때문에 변인 간 종속성이 높은 경우를 반영하기 어려운 점이 있다. 또한 변인들의 중요 순위를 부여하는 기준에 따라 결과가 달라지는 모습이 나타나기도 하였다(한광희 외, 2010; Bolón-Canedo et al., 2013).

Embedded 방법은 변인 간 상호작용을 고려하는 동시에 계산의 효율성을 높이기 위해, 규제화(regularization) 또는 변인 중요도(feature importance)를 활용하여 변인을 선정하는 방법이다. 이는 변인 선정을 위한 모형을 구현하고 모형 내 변인을 선정하는 알고리즘이 있는 것이라 생각할 수 있다. 규제화를 활용하는 대표적인 방법으로는 LASSO(Least Absolute Shrinkage and Selection Operator) 회귀모형이 있으며 종속변인과 상관없는 변인을 제거하는 접근을 취한다. 이와 달리 변인 중요도를 활용하는 랜덤 포레스트(Random forest) 모형에서는 Filter 방법과 유사하게 중요한 변인의 순위를 확인하여 변인을 선정한다.

변인선택 방법은 그 특징에 따라 강점이 다르기 때문에 분석자료의 특징이나 연구자의 연구 방향에 따라 적절한 방법을 선택하여 사용하게 된다. 변인선택 방법에 따른 특징을 살펴본 연구에 따르면, Filter 방법은 변인 중요도를 계산하는 방법에 따라 선정되는 변인이 달라지는 모습이 나타나기 때문에 다양한 지표에 근거한 변인 중요도 결과를 종합하여 공

통적으로 선택하는 경향을 보였으며, Wrapper 방법은 예측변인의 개수가 많지 않고, 예측모형이 상당히 구체화된 상황에서 적용되는 경우가 많았다. 상대적으로 Embedded 방법은 예측모형에 종속되지 않고 독립적으로 주요 변인을 확인할 수 있으며, 다양한 종류의 변인을 한번에 다룰 수 있다는 점에서 많은 연구에서 활용되는 모습이었다. 그중에서도 LASSO 방법은 종속변인 및 설명변인의 종류 및 특징에 따른 변화가 크지 않는 모습을 확인할 수 있었으며(Fu et al., 2019; Zhao & Yu, 2006), 심층인공신경망 예측모형에서 좋은 결과를 보이기도 하였다(Sun et al. 2016; Zhang et al. 2019).

교육연구에서도 Embedded 방법을 활용하여 예측변인을 선정하거나(홍유정, 박현정, 2021), 중요 변인을 확인하는 연구(유진은, 2016; 정혜원 외, 2022)를 확인할 수 있었다. 홍유정, 박현정(2021) 연구에서는 심층인공신경망 예측모형의 예측변인을 선정하기 위하여 Sparse Group LASSO를 적용하였으며, 회귀계수가 0이 아닌 변인들을 예측변인으로 활용하고 예측 성능을 살펴보았다. 한편, 유진은(2016) 및 정혜원 외(2022) 연구에서는 예측변인 선정을 위한 접근은 아니지만 종속변인과 관련이 높은 주요 변인을 확인하기 위하여 LASSO방법과 혼합효과 랜덤 포레스트(Mixed-effects random forest)를 활용하기도 하였다. 공통적으로 학생의 인지적・정의적・환경적 특성을 보여주는 수백 개 변인 중 종속변인과 관련이 높은 변인을 확인하고자 하였다.

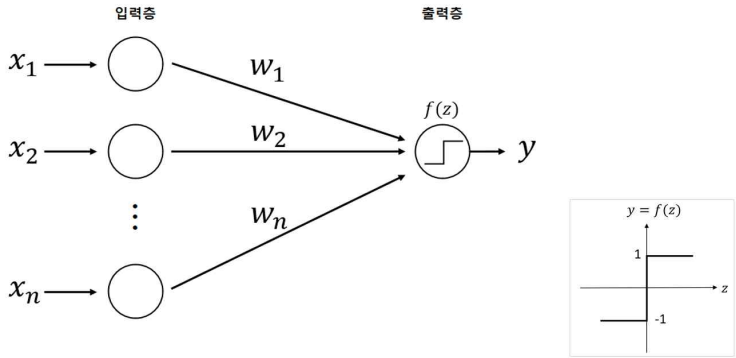
나. 심층인공신경망

1) 인공신경망 모형과 작동원리

인공신경망의 구조는 생물학적 신경망의 작동원리를 수학적으로 구현

한 모형이다(김의중, 2016; Hebb, 1949; McClelland & Rumelhart, 1988). 하나의 신경세포(neuron)에서 이루어지는 자료 처리 과정이 자료를 바탕으로 모수를 산출하는 퍼셉트론의 알고리즘으로 구현되었고, 이후 심층인공신경망으로 발전되었다. 심층인공신경망은 퍼셉트론 구조가 복잡해지고 깊어진 경우를 말한다. 따라서 이 소절에서는 단층퍼셉트론과 은닉층이 1개 있는 다층퍼셉트론의 연산과 작동원리를 살펴봄으로써, 신경망 모형의 기본 체계를 이해하고자 한다.

퍼셉트론(Perceptron)은 인공신경망이론을 설명하는 가장 기본적인 알고리즘이며 단층퍼셉트론(Single-layer perceptron)이라고도 한다. 단층퍼셉트론은 [그림 Ⅱ-2]와 같이 여러 개의 노드로 구성된 입력층과 한 개의 노드로 구성된 출력층으로 이루어진 모형이다(Rosenblatt, 1958). 입력층에서 입력된 값들은 노드의 활성화 함수를 거쳐 그 결과에 따라 특정 범주로 분류된다.



[그림 Ⅱ-2] 단층퍼셉트론의 구조(Rosenblatt, 1958 참고)

여기서 입력층(input layer)은 입력 자료를 받아들이며, 출력층(output layer)은 일련의 과정을 거쳐 처리된 결과를 출력한다. 또한 [그림 Ⅱ-2]에서 원으로 표현된 노드(node)는 입력값에 활성화 함수(activation function)를 적용하여 출력하는 역할을 하며[2], 각 층의 노드 간 관계는

2) 입력층의 노드에서는 연산 과정 없이 입력 자료를 다음 층으로 전달해주는

가중치 또는 연결강도(weight)로 정의된다. 인공신경망에서는 이 가중치 (w_1, w_2, ..., w_n)가 변수가 되어 반복적인 학습 과정을 통해 최적의 가중치를 찾고자 하는 알고리즘이라 할 수 있다.

[그림 Ⅱ-2]에서의 x_n은 n번째 변인의 입력자료를 의미한다. 따라서 입력층에는 변인의 수와 동일한 n개의 노드를 가진다. w_n는 n번째 입력 노드와 출력 노드 사이의 가중치를 말하며, 입력자료의 정보를 출력층에 전달해주는 정도를 결정한다. 출력층에는 y가 출력되는 하나의 노드가 있으며, 출력층 노드에서는 활성화 함수에 입력자료와 가중치를 곱한 후 모두 더한 값을 넣어 y를 출력하게 된다.

단층퍼셉트론은 식 (Ⅱ-1)과 같이 나타낼 수 있다(오일석, 2017의 표기를 설명에 맞게 변형하였음).

$$\hat{y} = f(z)$$
$$z = w_0 + \sum_{i=1}^{n} w_i x_i, \qquad f(z) = \begin{cases} 1 & z \geq 0 \\ -1 & z < 0 \end{cases} \qquad \text{식 (Ⅱ-1)}$$

출력되는 값은 실제 값이 아닌 모형에 의해 예측된 값이기 때문에 $\hat{y}$으로 표현하고, w_0는 절편을 의미한다. 이는 $y = f(W^T X)$와 같은 행렬식으로도 나타낼 수 있으며, $X = (1, x_1, x_2, \cdots, x_n)^T$, $W = (w_0, w_1, \cdots, w_n)^T$이다.

만약 교육 프로그램에 참여하는 학생의 중도 탈락을 판별하기 위한 단층퍼셉트론 모형을 구현하고자 한다면, 학생의 출석 횟수, 성별, 과제 제출 여부, 중간고사 성적을 판별을 위한 기초 자료로 사용할 수 있을 것이다. 이때 입력층의 노드 수는 4개(출석 횟수, 성별, 과제 제출 여부, 중간고사 성적)가 되며, 출력층에서는 각 학생의 중도 탈락 여부가 된다.

역할만 하므로 층의 개수를 세지 않는다. 퍼셉트론은 연산이 이루어지는 한 개의 출력층만 있으므로 단층퍼셉트론이라고 부른다(오일석, 2017).

이때 출력값은 각 학생의 4개 변인에 해당하는 값과 각 가중치의 곱을 모두 더한 후(z) 이 값을 활성화 함수에 투입하여 산출한 값이다. 단층퍼셉트론에서의 활성화 함수는 z값이 0 이상이면 1을, 0 미만이면 -1을 산출하는 계단함수를 사용하는데, 이후 다층퍼셉트론에서는 다양한 활성화 함수를 적용한다.

하지만 여기서 계산된 출력 값과 실제 값의 차이를 최소로 하는 적절한 가중치를 알지 못한다는 문제가 발생한다. 따라서 적절한 가중치를 찾기 위해 손실함수(loss function)를 최소로 하는 값을 찾아 반복적으로 가중치를 갱신하는 최적화(optimization)[3] 과정을 거치게 된다. 손실함수는 실제값과 모형에서 예측한 값의 차이를 표현하는 함수이며, 모형이 주어진 자료를 잘 표현하고 있는지 살펴볼 수 있는 지표이기도 하다.

단층퍼셉트론의 가중치를 $W = (w_1, w_2, \cdots, w_n)^T$이라 한다면 손실함수를 식 (Ⅱ-2)와 같이 설정할 수 있다. 단층퍼셉트론에서는 계단함수(step function)를 활성화 함수로 사용하였기 때문에 결과값($\hat{y}$)이 1 또는 -1이 출력되므로, 다음과 같이 실제값과 예측한 값이 서로 다른 정도를 나타낸다(오일석, 2017).

$$E(W) = -\sum_{X_k \in Y} y_k (W^T X_k) \qquad \text{식 (Ⅱ-2)}$$

기계학습에서는 미분을 통해 손실함수가 작아지는 방향을 찾아내며, 이는 심층인공신경망 작동원리의 핵심인 경사하강법(Gradient Descent)의 주 원리이기도 하다. 퍼셉트론을 포함한 심층인공신경망은 많은 변수를 필수적으로 다루기 때문에 편미분을 사용하여 미분값이 이루는 벡터인 그레디언트(gradient) 값을 찾은 후, 이에 따라 최저점을 찾아간다. 경사하강법과 심층인공신경망의 최적화 문제는 이후 심층인공신경망 상세화 소절에서 자세히 다룰 예정이다.

---

3) 최적화(optimization)란 통계 모형에서 자료를 통해 가중치 등의 모수 값을 추정하는 것을 말한다.

따라서 식 (Ⅱ-2)를 편미분한 후, 경사하강법의 학습 방법인 델타 규칙(delta rule)에 적용하게 되며, 최종적으로 가중치 w_i의 값은 손실함수가 감소하는 방향으로 갱신된다. 식 (Ⅱ-4)의 ρ는 학습률(learning rate)를 의미한다(오일석, 2017의 표기를 설명에 맞게 변형하였음).

$$\frac{\partial E(W)}{\partial w_i} = \sum_{X_k \in Y} \frac{\partial(-y_k(w_0 x_{k0} + w_1 x_{k1} + \cdots + w_i x_{ki} + \cdots + w_n x_{kn}))}{\partial w_i}$$

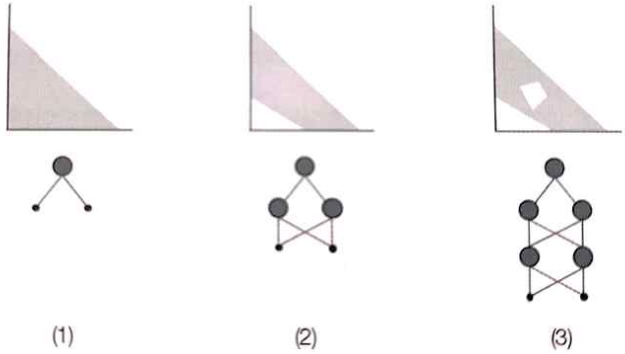
$$= -\sum_{X_k \in Y} y_k x_{ki}, \quad i = 0, 1, \cdots, n \qquad \text{식 (Ⅱ-3)}$$

$$w_i = w_i - (-\rho \sum_{X_k \in Y} y_k x_{ki}), \quad i = 0, 1, 2, \cdots, n \qquad \text{식 (Ⅱ-4)}$$

하지만 이와 같이 입력층과 출력층으로만 구성되어 있는 단층 퍼셉트론은 비선형 문제를 해결하기 어려웠으며(Minsky & Papert, 2017), 이를 해결하기 위해 입력층과 출력층 사이에 은닉층(hidden layer)을 추가한 다층퍼셉트론(Multi-Layer Perceptron)이 등장하였다. 단층퍼셉트론을 표현한 식 (Ⅱ-2)를 풀어보면 $w_0 + w_1 x_1 + w_2 x_2 + \cdots + w_n x_n$과 같이 하나의 직선 방정식이라는 것을 알 수 있다. 따라서 주어진 자료를 선형적으로만 분류할 수 있고, 간단한 문제라 하더라도 비선형 문제의 판별은 불가능한 것이다. [그림 Ⅱ-3]는 은닉층의 계층 수에 따라 표현할 수 있는 복잡도를 보여준다.

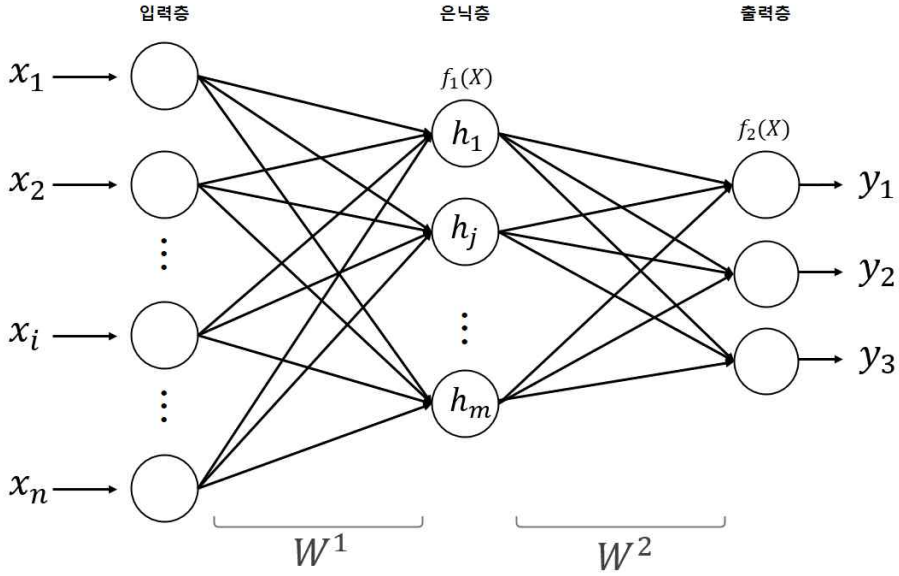
다음 [그림 Ⅱ-3]은 다층퍼셉트론 중 은닉층(hidden layer)이 하나이며, 출력 노드가 2개인 다층퍼셉트론의 예이다. 은닉층은 입력층과 출력층 사이에 위치하여 외부로 나타나지 않는 층이다. 은닉층은 각 입력과 가중치 곱의 합을 입력으로 하며, 이 값을 은닉층 활성화 함수를 적용시켜 출력을 계산한다. 일반적으로 은닉층 및 노드의 개수가 늘어날수록 복잡

한 문제를 잘 해결하는 경향이 있으나, 입력 자료의 크기와 형태에 따라 적절한 층의 깊이와 각 층의 노드 수는 다른 것으로 알려져 있다.



[그림 II-3] 은닉층 계층 수에 따른 표현의 복잡도 (김의중, 2016, p. 255)

다층퍼셉트론의 작동원리는 단층퍼셉트론의 작동원리의 확장형이라 할 수 있다. 여기서 출력층 노드의 수는 주어진 사례를 몇 개의 그룹으로 분류하는지에 따라 개수가 정해진다. 단층퍼셉트론의 예에서는 중도탈락 여부를 구분하는 문제상황이었기 때문에 출력층의 노드가 1개였으나, 중학교 3학년 학생들의 고교 진학을 특목고, 일반고, 특성화고로 예측하는 상황이라면 [그림 II-4]와 같이 출력층의 노드는 3개가 될 것이다.



[그림 Ⅱ-4] 다층퍼셉트론의 구조의 예

[그림 Ⅱ-4]에 나타난 x_n은 입력자료이며, W^1는 입력층과 은닉층을 연결하는 가중치 행렬, W^2는 은닉층과 출력층을 연결하는 가중치를 의미한다. h_j은 은닉층의 j번째 노드이며, y_1, y_2, y_3은 출력층의 3개 노드를 지칭한다. [그림 Ⅱ-4]의 다층퍼셉트론의 연산 과정을 식으로 표현하면 다음과 같다.

가중치 행렬 W^1와 W^2는 식 (Ⅱ-5)와 같이 표현할 수 있으며, 행렬의 각 항 w_{ji}는 i번째 노드와 j번째 노드 사이의 가중치를 의미한다.

$$W^1 = \begin{pmatrix} w^1_{10} & w^1_{11} & \cdots & w^1_{1n} \\ w^1_{20} & w^1_{21} & \cdots & w^1_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w^1_{m0} & w^1_{m1} & \cdots & w^1_{mn} \end{pmatrix} \quad W^2 = \begin{pmatrix} w^2_{10} & w^2_{11} & \cdots & w^2_{1m} \\ w^2_{20} & w^2_{21} & \cdots & w^2_{2m} \\ w^2_{30} & w^2_{31} & \cdots & w^2_{3m} \end{pmatrix} \qquad \text{식 (Ⅱ-5)}$$

우선, 자료가 입력층에 입력되면 W^1과 곱해지게 되며, 이는 각 은닉

층 노드마다 입력자료와 가중치 곱의 합이 활성화 함수를 거쳐 출력 됨을 의미한다[4]. 즉, 은닉층의 j번째 노드에서는 입력자료 X와 가중치 W_j^1의 행렬곱이 은닉층의 활성화 함수인 $f_1()$을 통해 결과값을 출력한다. 따라서 식 (Ⅱ-6)의 s_j는 활성화 함수를 적용한 후 출력되는 은닉층의 출력값이며, 여기서 W_j^1는 j번째 은닉층 노드에 연결된 가중치 벡터(vector)를 의미한다(오일석, 2017의 표기를 설명에 맞게 변형하였음).

$$s_j = f_1(W_j^1 X),\ j = 1, 2, \cdots, m$$
$$\text{여기서, } W_j^1 = (w_{j0}^1, w_{j1}^1, \cdots, w_{jn}^1),\ X = (1, x_1, \cdots, x_n)^T \qquad \text{식 (Ⅱ-6)}$$

이후 최종 출력값 y_l은 식 (Ⅱ-7)과 같이 나타날 수 있으며, W_l^2는 l번째 출력층 노드에 연결된 가중치 벡터를 말한다. 즉 다층퍼셉트론은 각 층의 가중치를 함수로 하는 합성함수의 연산이라 할 수 있다. 이는 추후 손실함수를 최소화하기 위해 그레디언트를 계산할 때 연쇄법칙(chain rule)을 적용할 수 있도록 하는 부분이기도 하다.

$$y_l = f_2(W_l^2 S),\quad l = 1, 2, 3$$
$$\text{여기서, } W_l^2 = (w_{l0}^2, w_{l1}^2, \cdots, w_{lm}^2),\ S = (1, s_1, s_2, \cdots, s_m)^T \qquad \text{식 (Ⅱ-7)}$$

$$\hat{Y} = f_2(W^2 f_1(W^1 X)) \qquad \text{식 (Ⅱ-8)}$$

이후 모형에서 계산한 $\hat{Y}$와 실제 Y의 값을 손실함수를 통해 확인한 후, 손실함수가 최소가 될 때까지 이 과정을 반복하게 된다.

---

4) 다층퍼셉트론에서는 은닉층 노드에서도 활성화 함수가 적용되며, 활성화 함수의 종류에 따라서 은닉층 출력 값을 조정할 수 있다. 활성화 함수와 관련된 구체적인 내용은 이후 소절에서 자세하게 다룰 예정이다.

이와 같이 은닉층의 개수를 증가시킴으로써 입력층에서 출력층으로 이어지는 연산 과정이 퍼셉트론의 변수인 가중치의 n제곱 형태가 됨에 따라 입력과 출력 사이의 복잡한 관계를 표현할 수 있는 모델을 만들 수 있게 된다.

2) 심층인공신경망 모형 상세화

심층인공신경망 은닉층을 추가하는 방법으로 복잡한 모형을 구현하여 자료의 패턴 인식 성능을 향상시킨 결과이다. 앞서 살펴보았듯이, 인공신경망 모형에서는 노드 간의 가중치(weight)를 최적화할 때, 각 노드마다 활성화 함수에 의한 계산이 이루어지며 이를 반복하게 된다. 그러나 심층인공신경망에서는 기존 인공신경망에 비해 은닉층의 개수, 노드의 개수가 늘어나면 계산량이 급격하게 늘어나게 되는 단점이 있다. 계산량의 증가는 컴퓨터의 발전으로 해결이 되어가고 있는 추세이나, 활성화 함수 적용 과정과 최적화 과정에서 '기울기 상실' 문제와 최적화 과정에서의 '지역최솟값', '과적합' 문제가 발생하기도 한다(오일석, 2017; Priddy & Keller, 2005.

심층인공신경망 연구에서는 이를 보완할 수 있는 방법을 지속적으로 연구하면서 발전시켜 나가고 있다. 기울기 상실과 관련하여 기존 문제를 보완한 활성화 함수(e.g. ReLU 함수)를 사용하거나 사전훈련(pretraining) 방법을 적용하려는 모습을 볼 수 있으며, 최적화 문제와 관련해서는 지역최솟값(local minimum)에서 쉽게 탈출할 수 있는 경사하강법(e.g. 확률적 경사하강법, 미니배치 경사하강법) 또는 중요하지 않은 정보를 배제하고 분석할 수 있는 최적화 기법(e.g. 드롭아웃)을 활용하여 해결하려는 모습을 볼 수 있다. 이를 구체적으로 정리하면 다음과 같다.

가) 역전파 알고리즘과 최적화 문제

통계 모형에서 자료를 통해 가중치 등의 모수 값을 추정하는 것을 최적화(Optimization)라고 한다. 이와 같은 맥락에서 인공신경망에서는 원하는 결과를 출력하기 위해서 각 층 간 정보 전달 과정에서의 적당한 가중치를 알아내는 최적화 작업을 하게 되며, 구체적으로는 모든 훈련 데이터에 대해서 출력층의 결과 값과 실제 값이 허용 오차 이내가 될 때까지 가중치를 업데이트하는 것을 반복하는 것을 의미한다.

단층퍼셉트론에서는 입력층과 출력층만 존재하기 때문에 출력층의 결과값을 실제값과 비교하여 오차가 최소가 되도록 가중치를 갱신하면 되었으나, 다층퍼셉트론에서는 입력층과 출력층 사이에 여러 은닉층이 존재하고, 각 은닉층과 연결되어 있는 가중치가 모두 연결되어 있어 가중치 갱신이 어려운 문제가 있었다.

이러한 문제는 역전파(Back propagation) 알고리즘 적용하여 해결가능하며, 심층인공신경망은 이를 기본적인 학습 알고리즘으로 활용한다(Rumelhart, Hinton, Williams, 1986). 역전파라는 용어는 예측된 결과와 실제값의 차이인 오차를 출력층에서 입력층 방향으로 반영주는 것을 뜻하며, 역전파 알고리즘은 크게 세 단계로 설명할 수 있다. 앞에서 다루었던 다층퍼셉트론의 간단한 예인 [그림 Ⅱ-4]의 경우로 살펴보고자 한다.

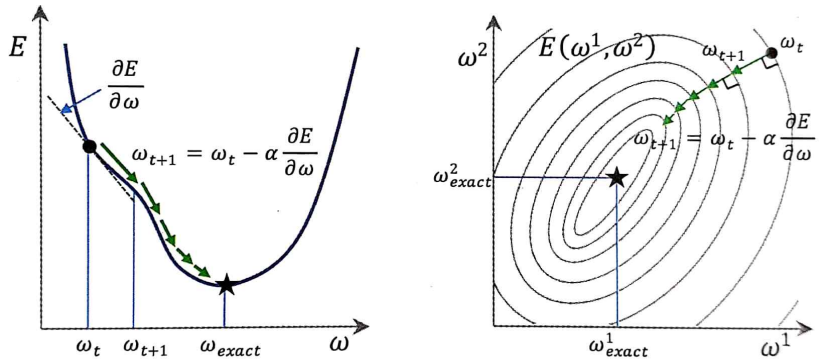
첫 번째는 입력층에서 출력층 방향으로 연산을 하여 출력을 구하며, 이때 가중치는 임의의 작은 값으로 초기화한다. 두 번째는 결과값과 실제값의 차이인 손실함수(Loss function)를 구한 후, 손실함수에서 W^2 가중치 행렬의 각 w_{lm}^2에 대한 미분값을 먼저 계산한 후, W^1 가중치 행렬의 각 w_{mn}^1에 대한 미분값을 구한다. 마지막 단계에서는 두 번째 단계에서 계산한 각 층의 가중치를 조정한다. 구체적으로는 앞 단계에서 계산된 미분값을 식 (Ⅱ-4)와 같은 델타 규칙(delta rule)에 적용하여 각 사이의 가중치를 손실함수가 줄어드는 방향으로 수정한다. 그리고 이 과정을

오차가 허용된 범위 내에 도달할 때까지 이 과정을 반복하게 된다.

이처럼 역전파 알고리즘에서는 각 층을 지날 때마다 가중치를 갱신하게 되는데, 손실함수가 최소가 되도록 가중치를 조정하기 위해 사용되는 방법이 경사하강법(Gradient descent)이다. 경사하강법에 사용되는 그레디언트(Gradient)는 기하학적으로는 함수 $f(x)$의 가장 가파른 접선, 즉 기울기를 말한다.

[그림 Ⅱ-5]의 2차원 손실함수의 그래프에서 알 수 있듯이 손실함수 E의 최솟값에서 접선의 기울기가 가장 완만하다는 것을 알 수 있다. 따라서 경사하강법에서는 매 반복마다 손실함수의 접선의 기울기의 반대 방향으로 가중치 조정하며 손실함수를 가장 작게 하는 값을 찾아 나간다.

그러나 역전파 알고리즘으로 심층인공신경망 모형의 모형 최적화과 원활하게 이루어지는 것은 아니며, 실제 이를 적용하는 과정에서는 경사하강법을 통해 도달한 최적값이 실제로 가장 좋은 값이 아닐 수 있다는 점과, 과적합 문제를 고려하는 것이 필요하다[5].



[그림 Ⅱ-5] 2차원(왼쪽), 3차원(오른쪽)에서 경사하강법의 적용
(김의중, 2016, p. 263)

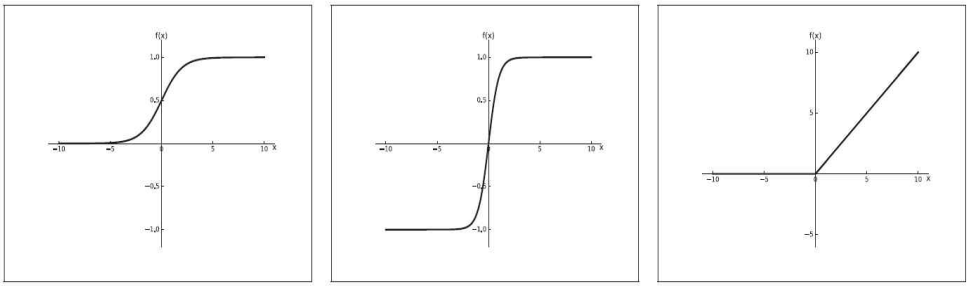
5) 이 외에도 오차가 출력층에서 입력층 방향으로 전달되면서 점점 사라지는 '사라지는 경사(Vanishing gradient)' 문제가 있다. 이는 활성화 함수를 설명하는 다음 소절에서 다룬다.

경사하강법은 학습데이터가 많은 대규모 문제에서는 계산량이 많아지기 때문에 모든 데이터를 적용하는 것이 상당히 부담되는 방법이다. 또한 초기 설정한 값에 따라 근사해가 지역최솟값(local minimum)에 빠지는 경우가 발생한다. 이러한 문제를 보완하기 위하여 확률적으로 선택한 하나의 샘플만을 사용하는 확률적 경사하강법(Stocastic gradiend descent)이나 전체 데이터세트를 몇 개의 미니 데이터세트로 나눈 후 각 세트에 경사하강법을 적용하는 미니배치 경사하강법(Mini-batch gradient descent) 또한 함께 사용되고 있다(김의중, 2015; Bouten et al., 1995).

하지만 최적화 문제가 해결된다고 하더라도, 과적합 문제가 해결되지 않으면 학습모델의 정확성을 보장하지 못한다. 과적합(overfitting) 문제란 학습모델에 훈련자료의 잡음(noise)이나 이상치(outlier)까지 모두 반영되어, 새로운 평가자료에서는 제대로 작동하지 않는 것을 의미한다. 이러한 문제를 해결하기 위해 손실함수에 구속조건(Penalty)을 적용하여 잡음이나 이상치의 영향을 줄이는 방법인 규제화(regularization)나 입력층과 은닉층에 있는 노드 중 일부를 임의로 제거한 후 축소된 신경망으로 학습하는 방법인 드롭아웃(drop-out) 방법을 적용한다(오창석, 2018; 최희열, 민윤홍, 2015).

나) 활성화 함수와 기울기 상실

활성화 함수(activation function)는 각 노드에서 앞 계층의 신호를 다음 계층으로 전파할 때, 값을 변환하여 준다. 이때 특정 임계값(threshold)을 기준으로 어떤 범주로 출력할 것인지를 결정하며, 다층퍼셉트론 모형에 비선형적인 특성을 부여하는 장치이다(오창석, 2018; Rasamoelina et al., 2020).



[그림 Ⅱ-6] 시그모이드(왼쪽), 쌍곡선탄젠트(가운데), ReLU(오른쪽) 함수 그래프

심층인공신경망에서 사용하는 대표적인 활성화 함수[6]로는 3가지가 있다. 첫 번째는 시그모이드(sigmoid) 함수로, 이 함수는 다음과 같이 표현할 수 있다.

$$f(x) = \frac{1}{1+e^{-x}} \qquad \text{식 (Ⅱ-9)}$$

함수 형태를 미루어 보면 로짓이 아주 작을 때 로지스틱 함수의 결과가 0에 매우 가깝고, 로짓이 아주 클 때는 로지스틱 함수의 결과가 1에 가깝다는 것을 의미하며, 이는 로지스틱 함수의 특수한 형태라 할 수 있다. 이 함수는 S자 형태를 띄며, [그림 Ⅱ-6] 왼쪽에 있는 그래프와 같다.

쌍곡탄젠트(Hyperbolic tangent) 함수는 비슷한 종류의 S자 비선형성을 사용하지만, 결과의 범위가 0에서 1사이가 아니라 -1에서 1사이이다.

$$f(x) = \tanh(x) = \frac{\sinh(x)}{\cosh(x)} = \frac{(e^{x}-e^{-x})/2}{(e^{x}+e^{-x})/2} \qquad \text{식 (Ⅱ-10)}$$

6) 단층 퍼셉트론에서는 0과 1을 출력하는 계단함수(step function)을 활성화함수로 사용하였으나 최적의 모수를 찾기 위해 자료를 바탕으로 기울기하강(gradient descent) 방법을 사용해서 학습하는 과정을 사용하게 되면서 미분이 가능한 시그모이드(sigmoid) 함수를 사용하게 되었다

출력 y와 로짓 x사이의 관계에 대한 결과는 [그림 Ⅱ-4]의 가운데 그래프에 나타나 있다. S자 비선형성을 사용할 때 tanh 함수는 시그모이드 함수와 달리 중심이 0이 되므로 출력 값의 분포가 안정적이라는 장점이 있으며, 계산에 효율성을 높인다.

그러나 신경망 모형의 층이 깊어짐에 따라 가중치 조정 정도가 약해지는 문제가 발생한다. 신경망 모형의 최적화 과정에서는 미분 값을 기준으로 최적 해의 방향으로 이동하는 기울기 하강(gradient descent) 방법을 활용하게 되는데, 이때 매 층을 지날 때마다 시그모이드 함수를 미분하게 되면서 그 값이 거의 0에 가까워지게 되는 것[7]이다 이를 기울기 상실 문제(vanishing gradient)라고 한다. 기울기가 0에 가까워진다는 의미는 지금의 가중치가 적절한 가중치에 가깝다는 의미이므로, 에러 정보가 낮은 층까지 잘 전달되지 않아 낮은 층의 가중치는 초기 값에서 크게 달라지지 않게 된다. 결국 학습 과정에서 상위 몇 개 층의 연결 강도만을 조정하게 되어, 적은 깊이의 모델을 사용한 것과 같은 결과가 되는 것이다(최희열, 민윤홍, 2015) 또한 기울기 상실의 문제는 신경망 구조 어디에서 발생하는지 찾기 어려우며, 시그모이드 함수보다 효율적이라 알려져 있는 쌍곡탄젠트 함수에서도 미분값이 사라지는 문제(missing slope problem)가 여전히 존재한다(Ciaburro & Venkateswaran, 2017).

따라서 이러한 문제를 해결하기 위한 활성화 함수가 ReUL(Restricted linear unit) 함수로서, x가 양수일 때는 $y=x$의 형태를 출력하고 음수일 때는 0을 출력하는 형태이다. ReLU 함수는 미분 가능하므로 여전히 경사하강법[8]을 사용할 수 있으면서도 계산이 매우 단순해짐에 따라 학습

---

7) 이는 시그모이드를 미분해보면 쉽게 이해할 수 있다. 시그모이드를 미분하면 $f'(z)=f(z)(1-f(z))$ 와 같은 도함수를 얻을 수 있다. 여기에 신술기하평균 공식을 적용하면 $\frac{f(z)+(1-f(z))}{2} \geq \sqrt{f(z)(1-f(z))}$ 이므로, 이 도함수의 최대값은 0.25가 된다. 이 과정을 매 층마다 반복하게 되면 매 단계마다 미분값을 곱하게 되는데, 이러한 최적화 과정에서 기울기 값은 급속하게 작하지게 된다.

8) 이와 관련된 내용은 역전파 알고리즘을 설명하는 다음 절에서 확인할 수 있다.

속도가 빨라진다는 장점이 있다. 또한 0이하의 값은 0으로 출력되면서 자료의 특징적인 부분이 부각된다. 이는 이후 최적화 문제에서 다루게 될 드롭아웃(drop out)과 비슷한 효과 상당히 유사한 부분이기도 하다.

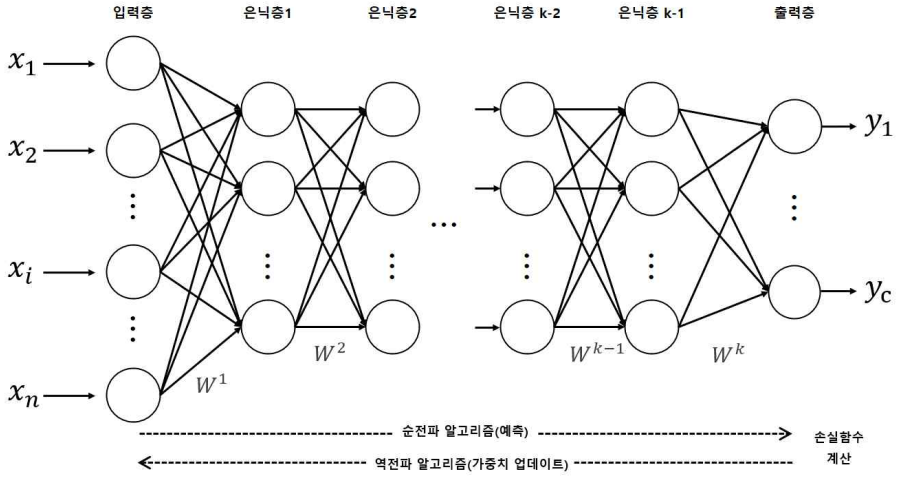
$$ReLU = x^{+} = \max(0, x) \qquad \text{식 (Ⅱ-11)}$$

따라서 ReLu 함수는 시그모이드 함수나 쌍곡탄젠트 함수보다 은닉층에서 쓰기에 적합한 것으로 알려져있어(Ciaburro & Venkateswaran, 2017), 최근까지도 가장 많이 사용되고 있는 활성화 함수이다.

다) 다층퍼셉트론 모형과 합성곱신경망

이상의 내용과 같이, 심층인공신경망 모형은 다층퍼셉트론 모형을 기본으로 하며, 손실함수를 최소로 하는 가중치 행렬을 찾아내는 모형 최적화가 이루어진다. 이때, 역전파 알고리즘을 적용하여 출력층에서 은닉층, 다시 은닉층에서 입력층 방향으로 단계마다 가중치 업데이트가 가능하게 되었고, 그 방향은 경사하강법을 적용하여 결정하였다. 이러한 다층퍼셉트론은 층의 깊이와 노드 수를 증가시킴으로써 주어진 자료에 나타난 복잡한 패턴을 잘 잡아내게 되었다. 또한 다층퍼셉트론의 기본 개념을 바탕으로 자료의 형태나 특성에 따라 다양한 심층인공신경망 모형들이 제안되기도 하였다. 대표적으로 이미지 데이터의 패턴 인식에 주로 쓰이고 있는 합성곱신경망과 데이터의 시계열적 속성을 반영하는 순환신경망이 있으며, 이 중에서도 합성곱신경망은 이미지데이터 뿐만 아니라 정형데이터에서도 활용 가능성을 넓히고 있다(Artzi et al., 2021; Du et al., 2021). 이 연구에서는 정형데이터인 패널자료를 분석에 활용하므로, 여러 심층인공신경망 모형 중 다층퍼셉트론 모형과 합성곱신경망 모형을 중점적으로 살펴보고자 한다.

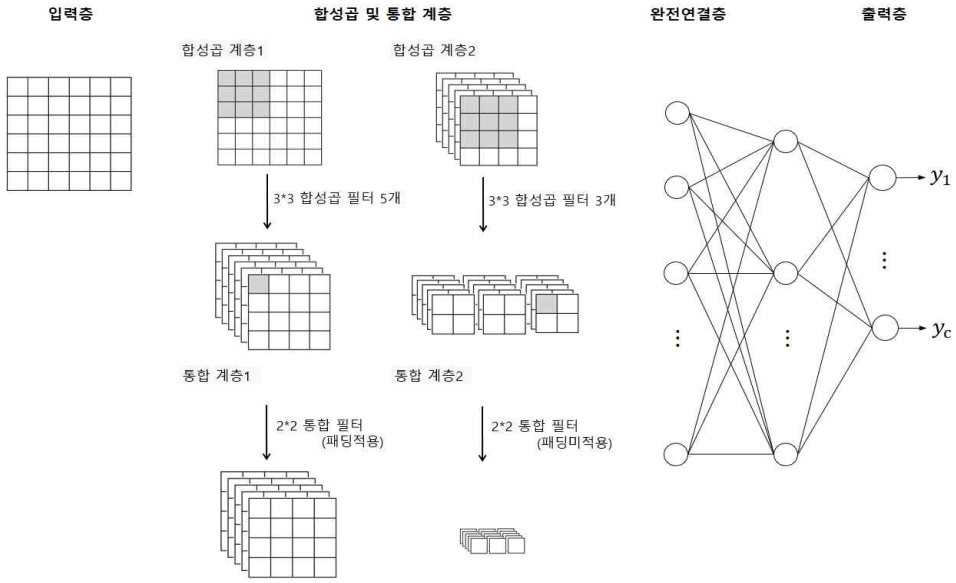
실제 자료를 예측 또는 분류하고자 하는 경우에는 은닉층의 깊이가 깊은 다층퍼셉트론을 활용하게 되며, 이후에서 언급하는 다층퍼셉트론은 은닉층의 수가 많은 경우를 가정한다. [그림 Ⅱ-7]에서 볼 수 있듯이 다층퍼셉트론는 그레디언트 계산과 가중치 갱신을 더 많은 단계에 걸쳐 수행하게 된다. 상대적으로 모형의 깊이가 얕은 다층퍼셉트론에서는 시그모이드 함수를 은닉층의 활성화 함수로 사용하며, 손실함수는 평균제곱오차(Mean squared error)를 활용하였으나, 일반적으로 다층퍼셉트론에서는 ReLU 함수를 주로 은닉층의 활성화 함수로 사용하며, 손실함수로는 교차 엔트로피(Cross-entropy) 또는 로그우도(Log likelihood)를 활용한다.



[그림 Ⅱ-7] 깊은 다층퍼셉트론의 구조

다음으로 합성곱신경망(Convolutional neural network)은 자료의 특징을 추출하는 과정을 추가하여 보다 특징적인 패턴을 학습하고, 이를 활용하여 예측 또는 분류를 하는 모형이다. 다층퍼셉트론에서는 연구자가 활용하는 자료의 특징을 바로 입력층 노드에 입력하였다. 반면 합성곱신경망에서는 [그림 Ⅱ-8]의 합성곱 및 통합 계층(Convolution and Pooling layer)과 같이 입력된 자료의 특징 다시 추출하는 부분이 추가된다. 여기에서는 저차원 특징부터 고차원 특징까지 계층적으로 추출하며, 추출된

특징을 완전연결층(Fully connected layer)에 입력하여 분류에 활용하는 방식이다.

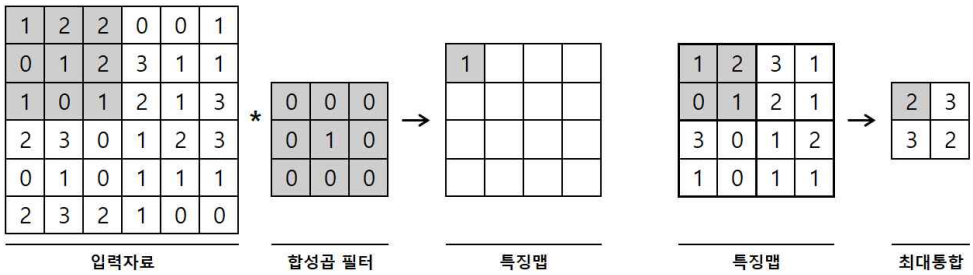


[그림 II-8] 합성곱신경망의 구조

계층적으로 특징을 추출하는 부분에서는 각 요소끼리 곱하고 그 결과를 더하는 선형 연산인 합성곱(Convolution)과 합성곱 결과의 대푯값을 결정하는 통합(Pooling)이 반복되는 형태이며, 이후에는 추출된 특징을 1 1차원의 상수값으로 변환하고, 특징의 대푯값을 기준으로 크기를 축소하는 통합 계층이 반복되는 형태이며, 특징 추출 단계에서 1×1로 저장된 데이터, 즉 1차원 벡터 형태로 분류신경망 모형의 입력값으로 저장되는 것이다. 이후 순전파 및 역전파 과정을 거치면서 학습이 진행된다. 이 단계는 기존의 인공신경망 모형과 동일하게 입력층, 은닉층, 출력층으로 구성되며, 특징 추출 부분과는 달리 층 간의 모든 뉴런들이 연결되어 있는 구조이다.

합성곱 계층(Convolution layer)에서는 필터(filter)를 사용하여 특징맵(Feature map)이라고 하는 새로운 정보를 출력하며, [그림 II-8]과 같이

나타낼 수 있다. 이때 여러 개의 필터를 병렬적으로 사용하여 다양한 특징들을 추출하는데, [그림 Ⅱ-8]을 보면 합성곱 계층1과 2의 특징맵이 각각 5개, 3개로 나타난 것을 볼 수 있다. 합성곱 계층에서 특징을 추출하는 방법은 [그림 Ⅱ-9]와 같다. 3×3 크기의 정 가운데 있는 값만 출력하는 필터[9]가 6×6 크기의 입력자료를 왼쪽에서 오른쪽, 위에서 아래로 전 영역을 훑고 지나가면서[10] 가중치를 적용하여 그 결과를 오른쪽 4×4 크기의 특징맵을 출력[11]한다.



[그림 Ⅱ-9] 합성곱 계층과 통합 계층의 적용 과정

통합 계층(Pooling layer)은 합성곱 계층에서 추출한 특징 맵의 크기를 줄여주는 동시에 덜 중요하거나 왜곡된 정보를 약화시키면서 특징적인 부분만 남겨주는 역할을 한다. [그림 Ⅱ-9]는 통합 계층에서 정보를 처리하는 과정[12]을 보여주는 예이며. 2×2 크기의 필터가 합성곱 계층에서 출력한 결과[13]에서 최대값을 선택하는 방법으로 통합하고 있다.

9) 만약 필터의 크기가 2×2이라면 출력되는 영역은 5×5 크기가 될 것이다. 필터의 크기와 개수는 주어진 자료의 특성을 고려하여 경험적으로 지정한다.
10) 필터를 적용하는 위치의 간격을 스트라이드(stride)라고 하며, 여기에서는 필터를 한 칸씩 이동하며 계산 값을 산출(stride=1)하는 상황을 가정하였다.
11) [그림 Ⅱ-9]의 입력자료 음영부분과 같은 위치 필터값의 곱들의 합을 특징맵을 추출하며, 다음의 식과 같이 계산된다.
$1\times0+2\times0+2\times0+0\times0+1\times1+2\times0+1\times0+0\times0+1\times0=1$
12) 통합하는 방법은 다양하나 대표적으로 평균통합(mean pooling)이나 최대통합(max pooling)을 사용한다. 평균통합은 특정 영역에 있는 입력값들의 평균을 구하고 평균값을 대푯값으로 활용하는 것이며, 최대통합은 특정 영역에 있는 입력값 중 최댓값으로 표현하는 것을 의미한다(오창석, 2018)

다. 불균형 자료와 심층인공신경망

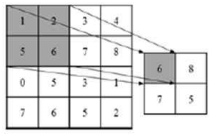
1) 불균형 자료의 개념 및 특징

학교 중퇴 또는 우울이나 공격성이 위험 수준 이상인 학생 예측, 온라인 수업에서 중도 이탈자 확인, 학습장애 판별 등 소수집단의 사례 수와 다수집단의 사례 수의 차이가 상당히 큰 자료를 불균형 자료(imbalanced data)라고 한다. 자료의 불균형 정도는 불균형 비율(Imbalanced Ratio, IR)[14]을 기준으로 살펴볼 수 있으며, 3 이하는 낮은 불균형, 4에서 9 이하는 중간 정도의 불균형, 10 이상은 높은 불균형을 띄는 자료임을 나타낸다(Fernandez et al., 2008).

분석자료가 불균형 자료인 경우 분류 및 예측 문제에서 여러 문제가 발생하게 된다. 그 이유는 분류함수가 다수집단에 편향된 추정을 함으로써 대부분의 자료를 다수집단으로 분류하는 결과로 이어지기 때문이다(방성완, 김재오, 2020; Owen, 2007; Oommen et al., 2011). 이는 소수집단 분류 및 예측 문제에서 가장 큰 문제점이라고 할 수 있으며, 그 이유는 일반적으로 소수집단을 잘못 분류하는 것이 다수집단을 잘못 분류하는 경우보다 더 큰 문제가 된다는 점에 기인한다. 예를 들어, 일반학생에서 학습장애 학생을 분류하는 경우, 일반학생을 학습장애라 분류하는 것은 추가적인 지원을 제공해주는 정도의 문제일 수 있다. 그러나 학습장애 학생을 일반학생이라 분류하게 된다면 학습장애 학생은 도움을 받지 못하여 학습 결손이 누적 및 심화되는 결과를 초래하게 된다. 따라서 불균형 자료 문제에서는 소수집단의 분류 정확도를 향상시키는 것이 핵

---

13) 여기서 출력된 영역은 출력된 영역 중 통합필터가 적용되는 영역을 의미한다. 즉, 합성곱 계층을 거쳐 출력된 영영의 크기가 4×4라면, 출력 영역 중 2×2크기만큼 해당하는 영역 내에서의 최댓값을 선택하고, 최종적으로 오른쪽 그림과 같이 2×2크기의 영역을 출력할 것이다.



14) IR = (다수집단 사례 수)/(소수집단 사례 수)

심이라 할 수 있다(방성완, 김재오, 2020). 이를 위하여 분석 시 소수집단의 정보가 다수집단의 정보만큼 반영될 수 있도록 하는 여러 방법과 불균형 자료에 적절한 분석 모형이 무엇인지에 관한 논의가 이어져 오고 있다.

그러나 집단 간 비율 차이로 인하여 집단 간 분포가 균등하며 독립적이라는 통계적 기본 가정을 충족시키기 어렵다는 점에서 전통적인 통계분석에서는 불균형 자료를 다루는 것이 쉽지 않았다(박근우, 정인경, 2019; Marquez-Vera et al., 2013). 따라서 최근에는 상대적으로 통계적 가정에 자유로우며 자료의 형태에 따른 모형 수정이 유연한 기계학습(Machine Learning) 분야에서 불균형 자료 분석 문제를 해결해보고자 하는 연구들이 진행되고 있다(Tang et al., 2008; Wang et al., 2017). 구체적으로 모형을 훈련할 때 소수집단의 정보가 누락되지 않도록 하기 위한 여러 기법에 따른 효과를 탐색하고 그 결과를 비교한 연구들이 수행되고 있으며, 이를 여러 기계학습 방법에 적용한 연구들이 그 예라고 할 수 있다(김한용, 이우주, 2017; 박근우, 정인경, 2019; 이종회, 2019; Barros et al., 2019).

2) 심층인공신경망 모형을 활용한 불균형 자료 분석

앞서 다루었듯이 심층인공신경망에서 전체 자료 중 일부를 사용하여 모형을 훈련시킨 후, 훈련에 사용하지 않은 자료를 이용하여 모형을 검증하는 과정을 거친다. 따라서 모형 훈련 시 소수집단의 사례를 이상치(outlier)로 분류하지 않고 그 정보를 모형 훈련에 반영되도록 하기 위한 기법을 적용하게 된다. 이를 위해 가장 많이 활용되는 방법 중에는 표집(Sampling)방법과 비용민감(Cost-sensitive)방법이 있다.

우선 표집방법은 훈련자료를 구성할 때 집단 간 분포를 맞추는 것을 말한다. 기본적으로는 소수집단의 사례를 원 사례보다 과대추출

(over-sampling)하거나, 다수집단의 사례를 과소추출(under-sampling) 방법을 활용하여 집단 간 비율의 균형을 맞춰주는 방식을 취한다. 이 중 과소추출은 자료를 없애는 방식을 취하기 때문에 정보가 손실되고, 삭제되는 자료에 따라 결과가 달라지는 문제가 있다(김한용, 이우주, 2017; Wang et al., 2017). 아울러 자료의 불균형 정도가 심한 경우에는 손실되는 정보가 더 많아진다.

이에 따라 상대적으로 과대추출을 기본 매커니즘으로 하는 연구들이 많이 이루어지고 있다. 과대추출 방법을 간단하게 적용하는 경우에는 소수집단의 사례를 임의로 선택하여 반복 추출하게 된다(Longadge & Dongre, 2013). 이러한 임의과대추출(Random over-sampling) 방법은 분석 모형에 기존 정보를 중첩적으로 추가하기 때문에 과적합(overfitting)의 문제를 야기할 수 있으며 분석 사례 수가 많은 경우에는 모형 구축에 시간이 많이 걸리기도 한다(방성완, 김재오, 2020; He, Garcia, 2009).

이 중 가장 문제가 되는 과적합 문제를 해결하기 위하여 SMOTE(Synthetic minority over-sampling technique) 방법을 비롯하여 이를 발전시킨 방법들이 적용되고 있으며(Chawla et al., 2002; He et al., 2008), 우수한 성능을 보이고 있다. SMOTE 방법은 소수집단 사례 하나를 뽑아 k-최근접 이웃(k-nearest neighbor) 알고리즘을 활용하여 k개의 가장 가까운 사례 중 임의로 한 사례를 선택한다. 이후에는 새로 생성된 사례와 원 사례의 직선거리의 가운데 지점을 중심으로 알고리즘을 반복하게 된다. 이는 기존의 데이터를 약간 조정한 사례를 추가하는 방식이므로 과적합 문제를 일부 개선해준다(Chawla et al, 2002). 이를 발전시킨 여러 방법 중 ADASYN(Adaptive synthetic sampling approach)는 경계 값에 더 큰 가중치를 두는 방법으로 집간 간 경계가 다소 모호한 자료 분류에 잘 작동하는 것으로 알려져 있다(He et al, 2008).

이러한 표집방법은 자료 수준에서 집단 간 균형을 조정하는 방법이라 할 수 있다. 이와 달리 비용민감방법은 알고리즘 수준에서 각 집단의 영향력을 조정하는 방법이다. 심층인공신경망 모형에서는 손실함수(Loss function)의 편향(bias)에 기초하여 반복적으로 모형을 훈련을 시키게 되

는데, 이때 소수집단의 편향을 다수집단과 유사한 정도로 가중하여 모형에 반영해 주는 것이다. 비용민감방법은 가중치 조절에 초점을 두기도 하고, 손실함수 자체를 수정할 수 있다(Zhou & Liu, 2005).

교육학 분야에서 불균형 자료를 분류 또는 예측하는 목적으로 심층인공신경망이 활용되고 있는 모습을 살펴보면, 주로 학업성취 및 중도탈락을 예측하는 연구들이 많았으며, 일부 연구에서는 학습장애 진단, 학생의 특정 성향을 예측을 다루는 연구도 있었다(김동일, 홍성두, 2009; 이혜주, 정의현, 2013).

우선, 학업성취 예측을 주제로 한 연구이다. Rachburee & Punlumjeak (2021)의 연구에서는 공과 대학 학생들을 대상으로 GPA 성적이 매우 낮은 학생들을 예측하고자 하였다. 이들은 전체 6,882명 중 114명(6%)으로 불균형 정도가 높았으며, 이를 위해 SMOTE, ADASYN, Borderline-SMOTE 방법을 적용하였다. 또한 다층퍼셉트론 모형뿐만 아니라 의사결정 나무(Decision tree), 서포트백터머신(Support vector machine), KNN(K-nearest neighborhood) 방법을 활용하여 그 결과를 비교하였다. 분석결과, Borderline-SMOTE를 적용한 다층퍼셉트론 모형의 성능이 가장 좋게 나타난 것을 확인할 수 있었다.

이와 유사한 연구로서 Nabil 외 (2021) 연구에서도 불균형 자료 문제를 과대추출 방법으로 해결해보고자 하였으며, 심층인공신경망 모형과 여러 기계학습 방법들의 예측 성능을 비교하였다. 분석에는 학생들의 인구통계학적 정보, 학업적 특징, 행동적 특징 등 총 16개의 변인을 활용하여 해당 학기 수업 중 Fail을 받은 학생(8.8%)을 예측해보고자 하였다. SMOTE를 적용한 경우와 그렇지 않은 경우를 비교하였으며, SMOTE를 적용한 다층퍼셉트론의 모형 성능이 가장 우수하였다.

국내 연구에서는 학생들의 학업성취 수준을 예측한 이창묵 외(2020) 연구가 있었다. 이 연구에서는 앞의 연구들이 주로 다층퍼셉트론 모형을 적용한 것과 달리 시계열적 특성이 있는 자료에 적합한 순환신경망 모형을 적용하였다는 점이 특징적이다. 분석자료는 국가수준 학업성취도평가 자료의 2011년 초6부터 2016년 고2까지 세 시점 연계자료를 활용하였으

며, 약 8~21% 정도 되는 기초학력 이하 학생을 예측하고자 하였다. 불균형 자료 조정하는 기법을 적용하진 않았으나, 같은 시점의 정보를 활용한 일부 모형에서는 좋은 모형 성능을 보여주기도 하였다. 전통적인 통계모형에서 다루기 힘든 종속변수의 비율이 극단적으로 차이나는 경우에 심층인공신경망이 대안이 될 수 있음을 논의하기도 하였다.

중도탈락을 주제로 한 연구는 단기 프로그램, 학교, 온라인 강좌 등 다양한 상황에서의 중도탈락 문제를 다루고 있었다. 우선, 하만석, 안현철(2019) 연구에서는 합성곱신경망을 활용하여 직업훈련 중도탈락을 예측하고자 하였다. 2015년 1월부터 2018년 9월까지 1440시간의 장기 교육과정 수강생 767명을 대상으로 하였으며, 이 중에 수료자는 605명(약 79%), 중도 탈락자는 162명(약 21%)인 불균형 자료이다. 이 연구에서는 분석 자료의 불균형 정도가 심각하지 않아, 과소추출 방법을 적용하여 훈련자료의 집단 간 비율을 1:1로 구성하였다. 심층인공신경망이 다양한 얼 자료를 활용할 수 있음을 고려하여 수집된 수강생 정보을 담은 정형자료와 함께 비정형 자료인 상담일지 자료를 분석에 활용하고자 하였다. 분석결과 예측률은 약 87% 정도로 나타났으며, 비정형자료를 함께 활용한 경우에는 예측하는 경우에는 95%의 예측 정확도를 보이기도 하였다.

Barros 외(2019) 연구에서는 다층퍼셉트론 모형, 의사결정 트리, 균형배깅(Balanced Bagging) 방법을 활용하여 학교 중도탈락을 예측하였다. 4년 동안 중등학교 7718명 학생을 대상으로 하였으며, 중도탈락자는 약 4.5%였다. 분석자료가 불균형 정도가 높음을 고려하여 훈련자료를 구성할 때 원자료의 비율을 그대로 유지한 경우와 과소추출, SMOTE, ADASYN 방법을 적용한 결과를 비교하여 제시하였다. 분석결과 집단 간 비율을 조정하여 훈련자료를 구성하는 경우에 모형 성능이 향상되었으며, 그 방법에 따라서는 결과가 크게 달라지지는 않았다.

이와 같이 교육 분야에서도 소수집단을 예측하고자 하는 연구가 이루어지고 있으며, 불균형 문제를 완화하기 위한 다양한 방법을 탐색하는 움직임도 확인할 수 있었다. 그러나 대부분 표집방법에 집중하고 있는 모습이 나타났으며, 일부 연구에서는 불균형 자료에 대한 고민을 하지

않은 경우도 확인할 수 있었다(김동일, 홍성두, 2009; 이혜주, 정의현, 2013). 교육 외의 자료를 활용한 여러 연구에 따르면 표집방법과 비용민감방법을 비교한 연구에서는 비용민감방법을 적용하는 경우, 자료의 분포나 자료의 정보를 변할 필요가 없기 때문에 표집방법에 비해 안정적인 성능이 나타나는 경향도 확인할 수 있었다(Mienye & Sun, 2021; Zhang et al., 2011). 국내에서는 홍유정, 박현정(2021)에서 표집방법과 비용민감방법을 적용한 심층인공신경망 예측모형을 구현하고 모형 성능을 비교한 결과, 비용민감기법을 적용한 모형의 성능이 가장 좋은 성능을 보이기도 하였다.

Ⅲ. 연구방법

1. 연구 절차

이 연구에서는 교육패널자료를 활용하여 기초학력 미달(이하 기초미달) 학생 예측을 위한 심층인공신경망 모형을 탐색하는 데 목적이 있다. 이를 위하여 한국교육종단연구2013(Korean Education Longitudinal Study 2013, KELS2013)의 중학교 2학년 시기의 정보를 활용하여 중학교 3학년 시점에서 개별 학생의 국어, 영어, 수학 기초미달 여부를 예측하는 심층인공신경망 모형을 구현하고자 하였으며, [그림 Ⅲ-1]과 같은 절차에 따라 연구를 수행하였다.

우선, 분석을 위한 준비 단계로 중2 학생과 관련된 변인들의 전처리를 실시하였다. 이후 분석의 첫 번째 단계로, 실질적인 심층인공신경망 예측모형을 구축하기에 앞서, 학업성취도와 관련이 있는 주요 설명변인을 확인하였다(연구문제 1). 분석에는 상관없는 변인을 0으로 추정하여 일부 변인을 선택하면서도 각 변인의 계수 추정이 가능한 LASSO(Least Absolute Shrinkage and Selection Operator) 회귀분석을 적용하였으며, 이를 다음 단계의 예측변인으로 활용하였다. 이때 실제 예측상황에서는 향후 기초미달 여부를 알 수 없다는 점을 고려하여 중3 시점이 아닌 중2 시점의 학업성취도와 관련 있는 변인을 확인하고자 하였다.

두 번째 단계에서는 기초미달 학생 예측을 위한 심층인공신경망 모형을 상세화하고 그 결과를 비교하였다(연구문제 2). 이때 예측변인의 개수, 소수집단 정보 반영방법, 심층인공신경망 종류를 달리하여 예측모형을 구축하였으며, 이에 따른 모형 성능을 확인하였다. 예측변인 상세화 과정에서 고려한 사항과 관련 연구는 <표 Ⅲ-1>에 정리하였으며, 자세한 내용은 다음과 같다.

단계	연구 내용	적용 모형
자료 전처리	KELS2013 4차(중2) 학생, 학부모, 학교 변인 확인 5차(중3) 학생 학업성취 수준 변인 확인	
	변인별 결측 확인, 중복 변인 확인 필요한 경우 유의미한 값으로 변환	
	결측처리	KNN imputation
[연구문제1] 주요 설명변인 선정 모형	중2 과목별 학업성취도를 설명하는 주요 설명변인 선정 - 종속변인: 중2 학업성취도 척도 점수 - 독립변인: 중2 학생, 학부모, 학교 변인 361개	LASSO
[연구문제2] 기초미달 학생 예측 모형	중3 과목별 기초미달 여부를 예측하는 예측모형 성능 비교 (1) 예측변인의 개수 (2) 소수집단 정보 반영방법 (3) 심층인공신경망 종류 - 종속변인 : 중3 기초미달 여부 - 예측변인 : 과목별 상위 주요 예측변인(상위 9, 16, 25, 36, 49, 64, 81개 변인)	심층인공신경망

[그림 III-1] 연구절차 및 주요 연구내용

<표 Ⅲ-1> 심층인공신경망 예측모형 상세화

구 분	상세구분	관련 연구
예측변인의 개수	• 과목별 상위 9개, 16개, 25개, 36개, 49개, 64개 81개 변인 *LASSO 분석결과 기준	- García et al. (2016) - Snieder et al.(2019)
소수집단 정보 반영방법	• 표집방법 : 과대임의표집, B-SMOTE, ADASYN • 비용민감방법 : 집단별 손실함수 가중치 조정	- Chawla et al. (2002) - Zhang et al. (2010) - Bang & Jhun (2014) - Mienye & Sun (2021)
심층인공신경망 종류	• MLP • CNN	- Johnson & Khoshgofraar (2019) - Yan et al. (2015)

*B-SMOTE=Borderline-SMOTE, MLP=다층퍼셉트론 모형, CNN=합성곱신경망 모형

첫째, 예측에 활용하는 예측변인의 개수를 달리하여 모형을 설정하였다. 앞서 실시한 LASSO 분석에서의 변인별 계수 절대값을 기준으로 관련성이 높은 변인을 확인한 후, 상위 변인을 기준으로 9개, 16개, 25개, 36개, 49개, 64개, 81개 변인을 예측변인으로 활용하고 그 결과를 비교하였다. 이와 같이 예측변인 개수를 지정한 이유는 합성곱신경망 모형에서는 표형식의 교육패널자료를 이미지화하는 과정에서 변인을 n^2개씩 활용하게 되는 점을 고려한 것이며, 모형 간 동일한 기준으로 비교하기 위하여 다층퍼셉트론모형에도 동일하게 예측변인 개수를 지정하였다. 또한 이 연구에서 활용하는 교육패널자료는 인구통계학적 정보나 자기보고식의 주관적 응답이 다수 포함되어 있어, 집단 간 구분이 다소 명확하지 않을 가능성이 높은 자료[15]이다. 또한 각 변인들이 독립적인 의미를 지니는 자료이기 때문에 모형을 설정할 때 모든 자료를 활용하지 않을 수 있는 특징[16]이 있다. 따라서 학업성취도와의 관련이 높은 변인을 중심으

15) 상대적으로 기초미달 학생 예측에 중간고사, 기말고사 결과나 각 과목별 평가 개별 문항 응답값을 예측변인으로 활용하는 경우라면 교육패널 자료에 비해 집단 간 구분 가능성이 높을 것이라 추측해볼 수 있다.

16) 상대적으로 문자자료나 이미지 등의 비정형 자료의 경우에는 각 음소의 순서나 픽셀 위치에 대한 정보도 중요하기 때문에 음소 또는 픽셀을 모두 활용하는 것이 필요할 수 있다.

로 예측변인의 개수를 다르게 하여 기초미달 학생을 예측하고 그 결과를 비교하고자 하였다.

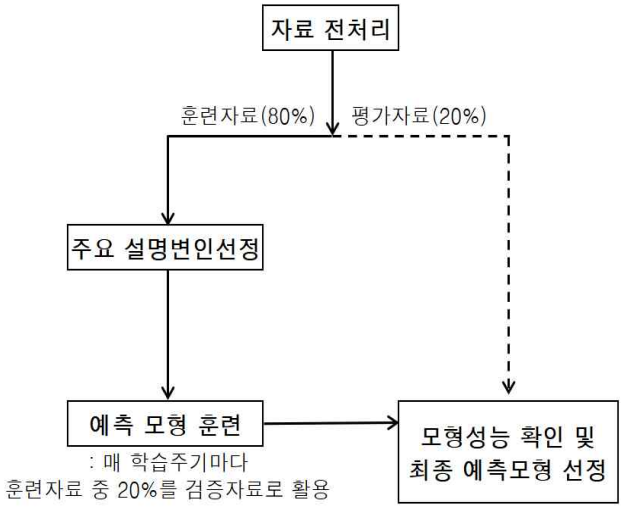
둘째, 기초미달 학생이 소수집단임을 고려하여, 모형 훈련 시 소수집단 정보를 다수집단만큼 반영해 줄 수 있는 방법으로 표집(Sampling)방법과 비용민감(Cost-Sensitive)방법을 적용하여 그에 따른 결과를 살펴보았다. 표집방법은 자료 수준에서 불균형 정도를 조정하는 방법으로, 소수사례를 다수사례만큼 단순 반복추출하는 임의과대표집 방법과 불균형 자료의 특성을 고려하여 새로운 소수사례를 생성하는 방법인 Borderline-SMOTE(Borderline Synthetic minority over-sampling technique)와 ADASYN(Adaptive synthetic sampling approach) 방법을 적용하였다(Chawla et al., 2002; He et al., 2008).

비용민감방법은 알고리즘 수준에서 불균형 정도를 조정하는 방법으로 손실함수의 가중치를 조정하거나 손실함수의 식 자체를 조정할 수 있다. 손실함수 식 자체에 대한 조정은 분석자료의 특징과 전체 알고리즘에 대한 이해를 바탕으로 함수 형태를 세밀하게 조정해야 한다(Castro & Braga, 2013). 상대적으로 집단별 손실함수 가중치를 조정하는 방법은 불균형 비율만 고려하고 적용할 수 있다는 장점이 있다. 교육연구에서의 불균형 자료를 활용한 심층인공신경망 예측모형 연구는 아직 초기 단계라 할 수 있기 때문에 이 연구에서는 적용이 용이한 가중치 조정의 비용민감 방법을 적용하고 표집방법을 적용한 모형과의 성능을 비교해보고자 하였다.

셋째, 심층인공신경망 종류로는 다층퍼셉트론 모형(Multi-layer perceptron, 이하 MLP)과 합성곱신경망 모형(Convolutional neural network, 이하 CNN)으로 예측모형을 구현하였다. MLP는 패널자료와 같은 표형식의 정형자료를 적용하기에 적절한 심층인공신경망 모형이며, CNN은 상대적으로 이미지 자료에 적합한 모형이지만 모형 내에서 주요 특징을 추출하는 과정이 있어 집단 간 정보가 뚜렷하게 구분되지 않는 경우에도 분류에 좋은 성능을 보인다고 알려져 있다(Ray et al., 2016). 따라서 MLP와 함께 CNN을 기초미달 학생 예측에 적용하고 그 성능을

비교해보고자 하였다.

한편 심층인공신경망 모형에서는 전체 자료 중 일부만을 활용하여 모형을 학습시킨 후, 훈련에 사용하지 않은 자료로 학습시킨 모형의 성능을 평가한다. 또한 훈련 과정에서는 손실함수의 값에 따라 가중치를 수정하는 과정이 반복되는데 이때, 매 학습주기(epoch)마다 일부 자료를 따로 떼어 놓고 학습을 시킨 뒤 떼어낸 자료로 학습된 모형을 평가하는 과정을 거친다. 이는 최종 모형이 평가자료에만 최적화되는 것을 막고 범용적으로 성능이 좋은 모형을 확인하기 위한 조치이다. 이에 따라 이 연구에서도 전체 자료를 약 8:2의 비율로 훈련자료(training data)와 평가자료(test data)로 나누어 연구를 진행하였다. 훈련자료를 활용하여 주요 설명변인을 살펴본 후, 심층인공신경망 모형을 상세화한 후 예측모형을 훈련시켰다. 이때 심층인공신경망 모형 훈련 시 매 학습주기마다 훈련자료 중 20%를 훈련에 활용하지 않고, 훈련자료의 학습과정을 평가하는 검증자료(validation data)로 활용하였다. 이후 학습된 모형에 평가자료를 적용하여 기초미달 여부를 예측해봄으로써 예측 성능을 확인하였다. 연구 과정별 분석자료를 활용은 [그림 Ⅲ-2]와 같이 나타낼 수 있다.



[그림 Ⅲ-2] 연구 과정별 분석자료의 활용

2. 분석 자료

「한국교육종단연구」는 학생들의 인지적·정의적 발달 상황과 가정과 학교에서의 교육경험을 종단적으로 조사하는 교육 조사 연구로(김양분 외, 201), 청소년들의 건강한 성장과 발달을 위해 각 교육 단계에 나타나고 있는 상황을 정확하게 진단하는 동시에 학생이 경험하는 교육 맥락들이 학생의 인지적·정의적 발달에 어떠한 영향을 미치는지 파악할 수 있는 정보를 제공한다. 또한 학생이 성장함에 따라 매년 추적 조사가 이루어졌을 뿐만 아니라 자체적인 학업성취평가를 통해 절대적 기준으로 구분된 성취수준 정보를 제공하고 있다는 특징이 있다. 이에 이 연구에서는 기초미달 학생 예측 및 이해를 위하여 「한국교육종단연구 2013」(Korean Education Longitudinal Study 2013, 이하 KELS2013)의 4차 5차 자료를 활용하였다.

2013년부터 초등학교 5학년 학생 7,324명을 대상으로 시작된 KELS2013은 교육 주체의 교육 행위 맥락을 고려하여 학생의 성장과 성취에 유의한 영향을 미치는 변인들을 조사하였다(김양분 외, 2014). 조사변인은 학교 내 교육활동과 학교 구성 및 학교의 인구 통계학적 특성 뿐만 아니라, 학교 외 교육경험, 방과후 활동 및 학생의 가정 배경까지 포함하고 있으며, 고등학교 및 대학 진학과 관련된 생각과 활동까지 다루고 있다.

교육활동 전반에 대한 정보와 함께, KELS2013에서는 국어, 수학, 영어 교과 성취도를 별도의 기초능력검사를 통해 측정한다(이규민 외, 2013). 무엇보다 KELS2013은 학업성취도의 수준의 종단적 비교를 주요 과제로 삼고 있어, 학업성취 수준의 수직 연계화가 이루어진 자료이다. 수직 척도화가 이루어진 검사점수는 동일 학생의 능력 수준의 변화 및 서로 다른 학생들 간 수준 차이 대한 정보를 확인할 수 있도록 한다. 따라서 KELS2013에서는 특정 학년에서 기초학력 미달 수준이었던 학생이 다음 학년에서는 작년에 비해 성취수준이 상승하였는지 확인할 수 있으

며, 학생 변화와 관련된 변인 들 간 관계를 탐색할 수 있는 자료라 할 수 있다(김양분 외, 2016).

KELS2013의 기초능력검사에서는 학생의 수준을 '우수', '보통', '기초', '기초미달'로 설정[17]하였고, 국어, 영어, 수학 각 교과별로 3개의 도달 기준 점수를 설정하여 각 수준을 구분하였다(이규민 외, 2015). 기초능력검사에서는 피험자의 능력 수준을 구분하기 위한 '분할점수(cut score)' 혹은 '기준(standard)'을 설정하고, 이에 근거하여 매년 학생의 수준을 확인한다. 따라서 해당 기준에 포함되는 점수를 획득한 피험자는 그 수준의 지식과 능력을 소유하고 있는 것으로 해석이 가능하다.

가. 연구대상

이 연구에서는 중학교 학생의 기초학력 미달(이하 기초미달) 여부를 예측하는 모형을 구현하고자 하였다. 초·중학교 시기에 기초학력이 제대로 갖춰지지 않은 경우에는 학습결손이 누적되어 학교급이 올라갈수록 학습이 더욱 어려워진다(김양분 외, 2016). 특히 중학교는 초등학교와 교육과정 운영 방식이 다르고 학업에 대한 압력이 커지기 때문에, 많은 학생들이 중학교 진학 후 학업에서의 어려움을 겪는 시기이기도 하다(김대현, 김현주, 2003). 실제로 KELS2013의 학업성취 수준 기초미달 비율[18]은 초등학교 시기에는 15% 내외였으나, 중학교 시기가 끝나는 시점에는 약 28% 내외로 급격하게 증가한 모습을 보이기도 하였다(박경호 외, 2020).

또한 KELS2013의 코호트 2015~2016년 시기의 중학생 학생들이며, 자

17) 2020년 이후 4단계 성취수준은 4수준(우수학력), 3수준(보통학력), 2수준(기초학력), 1수준(기초학력 미달)로 제시되고 있다(김희경 외, 2019).

18) 한국교육종단연구에서 자체 개발한 기초능력검사(초5-중3)의 우수, 보통, 기초, 기초미달 수준 중 학년별로 국어, 영어, 수학 가운데 한 과목이라도 기초미달에 해당한 경우의 비율을 의미한다(박경호 외, 2020).

유학기제가 적극 시행된 시기이기도 하다. 2015년에는 전국의 80%의 연구학교 및 희망학교를 대상으로 시범 운영되었으며, 2016년부터는 모든 중학교에서 시행되기 시작하였다(교육부, 2015: 8). 이에 따라 교육 전반에서는 자유학기제 운영에 따른 학업성취 수준 저하를 우려하는 목소리와 함께 기초학력에 대한 관심이 높아졌던 시기라 할 수 있다(남궁지영 외, 2017).

따라서 이 연구에서는 기초학력 미달 학생의 비율이 급격히 높아지며, 기초학력 유지에 대한 관심이 높았던 시기인 중학교 시기의 기초학력 미달 학생을 예측하는 모형을 구축해보고자 하였다. 구체적으로는 중학교 2학년 시기의 학생, 학부모, 학교 정보를 이용하여 개별 학생의 중학교 3학년 기초학력 미달 여부를 예측하고자 하였다. 이는 KELS2013의 4차, 5차에 해당하는 자료이다. 조사 성공률은 초등학교 5학년 시기의 원표본 대비 각 90% 내외로 보고되어 전반적으로 양호한 수준이었으며, 성별, 학교규모, 지역규모, 설립유형에 따른 무응답 비율을 확인한 결과 집단별 무응답 비율은 대체로 균등하게 나타나 표집편향(sampling bias)이 없는 자료라 판단할 수 있었다(남궁지영 외, 2018, pp. 108, 118)[19].

최종적으로 4차 연도 설문에 응답한 6,677명 중 학부모 조사지에 응답하지 않은 사례와 학생 및 학부모 조사지의 전체 문항 중 50% 이상 결측한 사례를 분석에서 제외하였다. 또한 학교 조사지에 응답하지 않거나 응답 결과에 신뢰도가 낮은 경우(학교 전체 학생 수를 0명, 교사 수를 1명으로 보고함)에 해당되는 3개 학교에 속한 학생들 역시 분석에서 제외하였다. 마지막으로 4차 연도의 학생 및 학교 정보는 있으나 5차 연도 기초학력능력검사 결과가 보고되지 않은 사례는 기초학력 미달 결과 예측에 활용할 수 없어 제외하였다. 따라서 최종적으로 분석에서 활용한 사례는 수학 교과 6,407명, 영어 교과 6,398명, 국어 교과 6,402명이다.

---

19) 종단적 추적 조사 연구에서는 추적 시기가 길어짐에 따라 패널 자체에서 이탈되는 사례가 증가하게 된다. 이때, 이탈한 사례들이 특정 집단의 속해 있다면 1차 표집에서 가정했던 모집단 분포의 편의(bias)가 발생할 수 있다.

나. 주요변인

1) 종속변인

앞서 언급하였듯이, KELS2013에서는 별도의 기초능력검사를 실시하여 학생의 학업성취를 확인하며, 기초능력검사 정답 개수, 수직 연계화가 이루어진 척도 점수(이하 척도점수)[20], 절대적 수행기준에 따른 학업성취 수준(우수, 보통, 기초, 기초미달)(이하 성취수준)에 대한 정보를 제공한다. 이 중 척도점수는 수직연계화가 이루어진 점수이며, 척도점수를 기준으로 성취수준을 나누게 된다. 따라서 이 연구에서는 척도점수와 성취수준을 종속변인으로 활용하였다.

'주요 설명변인 선정 모형' 에서는 국어, 영어, 수학 중2 학업성취도 척도점수[21]를 종속변인으로 활용하였다. 이 연구에서는 중3 시기 성취수준에 대한 정보를 확인할 수 있으나, 실제 예측 상황에서는 다음 학년의 성취수준을 알지 못하는 상태에서 학생의 기초미달 여부를 예측하게 된다. 따라서 학업성취도가 자기상관이 높은 변인이라는 점을 고려하여,

---

20) KELS2013의 학업성취도 척도점수는 추정된 문항모수를 기반으로 학생들의 능력모수를 추정한 후, 추정된 능력모수를 평균 200, 표준편차 40으로 변환하였다(이규민 외, 2013).

21) 이 연구의 최종 목적이 기초미달 학생 여부를 예측하는 것이긴 하나, 예측변인 선정 모형에서는 기초미달 여부가 아닌 학업성취도 점수를 종속변수로 활용하였다. 이는 예측변인 확인에서 활용한 LASSO(least absolute shrinkage and selection operator) 분석에서도 불균형 자료로 인한 문제가 발생하였기 때문이다. Lasso 알고리즘에서는 종속변인을 잘 예측하는 데 기여하는 변인에 가중치를 주게 되며, 종속변인이 이분변인(binary variable)인 경우에는 가장 예측이 높을 때를 기준으로 기준으로 λ값을 결정하게 된다. 그러나 기초미달 여부를 종속변인으로 하여 예측변인 선정 모형을 설정한 결과, 예측률이 상당히 낮았으며, 주요 예측변인으로 선정된 변인 및 그 회귀계수의 방향도 신뢰하기 어려웠다. 따라서 대안적 방법으로 연속변인(continuous variable)인 학업성취도 척도점수를 종속변인으로 활용하였다. 따라서 결과를 해석할 때, 학업성취도 점수에 영향을 미치는 변인과 기초미달 여부에 영향을 미치는 변인이 다르게 나타날 수 있음을 고려해야할 필요가 있다.

중2 시기의 학업성취도를 설명하는 변인을 확인하고자 하였다. 중2 학업성취도의 기술통계 결과는 <표 Ⅲ-2>와 같다.

<표 Ⅲ-2> 중2 과목별 기초능력검사 척도점수 기술통계

과목	사례수	최소값	최대값	평균	표준 편차	왜도	첨도
국어	6,402	140	329	230.09	46.35	-0.20	-0.82
영어	6,398	150	331	250.17	51.52	-0.14	-1.00
수학	6,407	147	324	231.67	50.96	0.09	-1.07

‘기초미달 학생 예측모형’에서는 국어, 영어, 수학 중3 기초학력 미달 여부를 종속변인으로 활용하였다. 우수학력, 보통학력, 기초학력, 기초학력 미달의 4수준으로 구분된 학업성취 수준 정보를 기준으로 기초학력 미달에 해당하는 학생을 1, 우수학력, 보통학력, 기초학력에 해당하는 학생을 0으로 변환하여 기초미달 여부를 나타내는 이분변인(binary variable)을 생성하였다. 중학교 3학년 시기의 과목별 기초미달 학생의 비율은 <표 Ⅲ-3>과 같으며, 국어 7.6%, 영어 14.9%, 수학 20.6%로 국어 과목의 불균형 정도가 상대적으로 높은 것을 확인할 수 있었다.

<표 Ⅲ-3> 중3 과목별 성취수준에 따른 학생 수 및 비율

과목		기초미달	기초학력	보통학력	우수학력	총 학생 수
국어	학생 수	**487**	1805	3259	851	6,402
	비율(%)	**7.6%**	28.2%	50.9%	13.3%	100%
영어	학생 수	**953**	1,324	1,843	2,278	6,398
	비율(%)	**14.9%**	20.7%	28.8%	35.6%	100%
수학	학생 수	**1,320**	2,608	1,294	1,185	6,407
	비율(%)	**20.6%**	40.7%	20.2%	18.5%	100%

2) 독립변인

이 연구에서는 주요 설명변인을 선정한 후, 이를 예측모형의 예측변인으로 활용한다. 활용하는 변인은 동일하나 주요 설명변인 선정에서는 동일한 시점의 학업성취도와의 관련성을 확인하고, 예측모형에서는 다음 해 기초미달 여부를 예측하는 데 활용한다는 점에서 명칭을 구분하여 사용하였다.

주요 설명변인을 선정하기 위한 '설명변인 선정 모형'에서는 KELS2013에서 제공하는 중학교 2학년 시기의 학생, 학부모, 학교 변인들을 문항 단위로 활용하였다. 이는 같은 구인에 속하는 문항 중에서도 문항 내용에 따라 학생 성취도 및 성취수준에 미치는 영향이 다를 수 있다는 점을 반영하고자 하였기 때문이다. 이후, 과목별로 나타난 주요 설명변인과 중2 학업성취도 척도점수를 '기초학생 미달 예측모형'의 예측변인으로 활용하였다. 설명변인 선정 모형에서 활용한 총 361개 예측변인은 <표 Ⅲ-4>에 제시되어 있다. KELS2013에서 제공하는 변인 중 분석에 적절하지 않은 변인들은 삭제하였으며, 일부 변인은 분석에 활용 가능하도록 값을 변환하였다. 구체적인 사항은 다음과 같다.

우선 KELS2013에서 제공하는 학생, 학부모, 학교 조사지의 문항 결측이 20% 이상인 변인[22]은 결측대체를 하더라도 그 값의 신뢰도가 높지 않다고 판단하여 분석에서 제외하였으며, 학생의 학업성취와 관련 없는 변인[23]들도 삭제하였다. 또한 학생, 학부모, 학교 조사지 중 중복되는 내

---

22) 결측이 20%가 넘은 변인은 다음과 같다. 우선 학생 조사지에서는 학급임원 경험 여부, 교육프로그램 경험여부 및 도움 정도(예체능), 전학 여부, 전학시기, 전학 이후 만족도를 제외하였다. 학부모 조사지에서는 방과후 교육활동(사이버 가정학습 프로그램 이용 정도), 방과후학교·학원·과외·학습지/통신/인터넷 참여여부·참여시간·참여비용, 진로 및 진학결정을 위한 활동 경험여부, 학교 연락 경험, 전학여부, 전학시기, 전학 이후 만족도를 제외하였다. 학교 조사지에서는 자유학기제 시행현황(2016년), 역량 기반 교육 관련 교사 연수 실시 여부 및 연수 시간, 사회·과학·역사평가방식 및 수행평가 비율이 포함된다.

23) 학생, 학부모의 설문 참여여부, 학부모의 응답자 변인은 분석에서 제외하였

용이 있는 경우[24]에는 문항 내용에 따라 하나의 응답만을 활용하였다. 또한 학교 조사지 중 학교의 교육적 맥락과는 다소 거리가 있는 물리적 정보의 변인들도 분석에서 제외하였다.

제외되지 않은 변인들 중 한 문항의 응답 값이 2개 이상의 범주를 가지며 비서열성을 보이는 경우에는 각 범주의 응답 여부가 하나의 변인으로 표현되도록 값을 변환하였다[25]. 왜도가 상당히 높은 변인들은 로그변환을 통해 극단적 변수의 분포를 조정[26]하였으며, 교육연한과 같이 의미있는 값으로 변환이 필요한 경우는 재코딩을 하였다. 또한 학교 변인 중 학교 규모 대비 해당 학생들의 비율을 활용하는 것이 의미 있다고 판단되는 경우에는 원래 값이 아닌 전체 학생 수 대비 해당 학생들의 비율로 변환한 후 활용하였다.

마지막으로 변인의 응답 중 '잘 모르겠다', '기타'의 선지가 있는 경우[27]에는 응답 비율이 매우 적고 다른 변인들에 비해 유의미한 정보를 주지 않는다고 판단하여 응답 해당 응답을 결측처리한 후 분석에 활용하였다.

다. 또한 학부모 조사지 중 가족 내의 모든 자녀에 대한 월평균 교육비 변인은 활용하지 않았으며, 설문지를 가져온 자녀에대한 교육비 변인만 분석에 활용하였다.

24) 학교정보 중 도시규모(특별시, 대도시, 중소도시, 읍면지역), 설립유형(국공립, 사립), 학교유형(남학교, 여학교, 남녀공학)은 학부모, 학교 조사지에서 중복으로 조사되었으며, 이 연구에서는 학교 조사지의 응답을 사용하였다.

25) 예를 들어 학교유형은 여학교, 남학교, 남녀공학으로 범주가 구분되어 있으므로, 여학교 여부, 남학교 여부, 남녀공학 여부 3개 변인으로 변환하였다. 이는 참조범주의 모든 값을 0으로 부여하는 참조변인과는 다르며, 원핫인코딩(one hot ecoding)이라고도 한다.

26) LASSO 분석방법이나 심층인공신경망 모형과 같은 기계학습 모형에서는 예측변인의 분포에 대한 가정을 하지 않으므로, 왜도가 심한 변인을 정규화하는 것이 필수적인 것은 아니다. 그러나 모형 훈련을 할 때 값의 크기나 범위가 큰 변인이 최적화 과정에 큰 영향을 미칠 수 있으며, 분석에 활용하는 변인의 범위 및 특성이 유사한 경우 모수가 안정적으로 추정된다고 알려져 있다(Ahsan et al., 2021). 따라서 이 연구에서도 왜도가 심한 변인들은 로그변환하여 분석에 활용하였다.

27) 학부모 조사지에서 자녀의 기대학력 변인 중 '잘 모르겠다' 선지의 응답 비율은 약 3.4%이며, 가족구조(동거가족), 방과후학교 참여이유, 사교육 참여이유 선지에 응답 비율은 0.3~1.1%였다.

<표 Ⅲ-4> 설명변인 선정 모형에서 활용한 독립변인

설문대상 [변인 개수]	영 역 [변인 개수]	세부 변인 [변인 개수]
학생 [237]	개인특성[1]	성별[1]
	인지적·비인지적 특성[83]	창의성[2], 자아개념(사회, 가족, 신체, 학업)[20], 공동체의식(참여의식, 타인배려, 협력행동)[12], 규칙준수[5], 자기관리[3], 다문화수용성(외국인에 대한 인식, 다문화 이웃/친구와의 관계)[13], 진로성숙도(자기이해, 진로계획성, 일에 대한 태도)[15], 자기결정성[4], 정신건강[9]
	수업참여[19]	수업이해도(국, 영, 수)[3], 수업집중도(국, 영, 수)[3], 수업태도[5], 수업분위기(부정, 긍정)[8]
	학교의 교육 및 환경[17]	수업방식(개별화, 상호작용)[8], 교사특성(성취압력, 교사열의)[9]
	학교생활과 적응 [47]	교사와의 관계[6], 교우관계[6], 학교폭력(학교 내 폭력 정도, 폭력피해경험)[12], 스트레스(시험, 학업)[8], 학생 의견반영정도[5], 진로 및 직업 관련 프로그램 참여 및 도움 정도[7], 교육프로그램 경험여부 및 도움정도(국, 영, 수)[3]
	여가 및 방과후 시간[39]	독서량[1], 독서향유정도[5], 운동시간[1], 학습활동 및 여가활동[7], 문화활동[7], 봉사활동 참여여부 및 인식[5], 선행학습 정도(영, 수)[2], 공부시간[1], 컴퓨터/스마트 미디어 사용 유형[5], 휴대전화 중독[5]
	부모-자녀 관계[20]	부모지원(학업, 정서)[9], 상호작용[5], 부모애착(소외)[6]
	진로 및 포부[11]	교육포부[1], 고교진학계획(학교유형 10개)[10]

설문대상 [변인 개수]	영 역 [변인 개수]	세부 변인 [변인 개수]
학부모 (78)	가족의 기본특징[5]	동거가족(유형 5개)[5]
	경제수준[1]	월평균 가구소득[1]
	교육기대 및 계획 [11]	고등학교진학계획(학교유형 10개)[10], 기대학력[1]
	방과후 교육활동 및 교육비[10]	월평균 교육비(설문대상 자녀)[1], 방과후 자기보호 주당 일수[1], 방과후 자기보호 일일 시간[1], 방과후 교육활동[2], 방과후 교육활동 참여 이유[5]
	가정생활 및 교육지원[31]	자녀와의 상호작용[4], 교육적 지원[10], 교육적 지원[10], 진학 지원[5], 자녀의 어학연수 및 해외학교 재학 경험[2]
	학교 교육[20]	학교 참여 및 정보제공[6], 학교만족도[14]
학교 (46)	학교통계[2]	학생 수[1], 교사 수[1]**
	학교배경[6]	학교 소재 도시규모[1]*, 설립유형(국공립)[1], 학교유형(남학교, 여학교, 남녀공학)*[1], 기초생활보장 수급 대상 학생 수[1]**, 원클릭 교육비 수급 대상 학생 수[1]**, 다문화 가정 학생 수[1]**
	학교 교육활동[28]	동아리 운영 개수[2], 동아리 인원 수[2]**, 국영수 평가방식 및 수행평가 비율(지필평가, 수행평가, 선다형, 서술형)[12], 역량 기반 교육을 위한 학교 교육활동[6], 학교구성원들의 학교 운영 활동[6]
	학교정책 및 운영[10]	자유학기제 시행여부(2015년)[1], 정책 사업 참여 현황(위클래스, 혁신학교, 진로상담교사 및 진로진학 상담실 운영, 교육복지우선지원사업)[4], 학부모 지원[5]

주1. 영역 분류는 김양분 외(2016)(pp. 41-61)의 기준을 참고하였다.
주2. 변인 개수는 설문지 문항 수를 기준으로 표기하였다. 예를 들어 '도시규모' 변수는 1개 문항으로 표기하였으나 실제 분석에서는 '특별시', '대도시', '중소도시' 3개 변수로 더미화하여 활용하였다. 또한 이와 같이 더미화한 변수에는 *표시를 하였다.
주3. 전체 학생 대비 해당 변인의 비율로 변환한 경우 **표시 하였다.

3. 연구모형 및 분석절차

가. 주요 설명변인 선정 모형

심층인공신경망 모형은 통계적 가정 및 변인 간 다중공선성(multicollinearity) 등의 문제로부터 자유로운 특징이 있어, 많은 변인을 예측변인으로 활용할 수 있다. 하지만 종속변인과 직접적인 관련이 없는 예측변인이 많아지면, 오히려 모형의 성능이 낮아지는 문제가 발생하기도 한다(García et al., 2016; Snieder et al., 2019). 이에 따라 심층인공신경망 예측모형 상세화에 앞서, '주요 설명변인 선정 모형' 에서는 중2 국어, 영어, 수학 학업성취도 점수에 영향을 미치는 중2 학생, 학부모, 학교 특성을 살펴보고자 하였다.

이를 위하여 많은 변인 중 종속변인과 관련이 높은 일부 변인을 선택할 수 있으며, 상관에 기반을 둔 방법이나 랜덤 포레스트 등 여러 변인 선택 방법 중 LASSO 방법은 종속변인 및 설명변인의 종류 및 특징에 따른 변화가 크지 않은 방법이라고 알려져 있는(Fu et al., 2019; Zhao & Yu, 2006) LASSO(Least Absolute Shrinkage and Selection Operator) 분석 방법을 적용하였다(Tibshirani, 1996).

LASSO는 손실함수(Loss function)인 OLS(Ordinary Least Square)에 추정 계수들의 합을 제한하는 제약함수(Penalty function)을 추가하여 종속변인에 영향을 미치는 정도가 작은 경우에는 계수 값을 0으로 수렴시키는 동시에, 각 변인의 계수도 추정할 수 있는 방법이다(Fan et al., 2014; Varian, 2014). 일반적인 회귀분석에서는 OLS 방법을 적용하여 분산(Variance)이 최소인 불편(unbiased) 추정치를 얻는다. 하지만 이는 분석에 활용되는 변인 간 독립성을 전제하기 때문에, 이 연구에서와 같이 많은 독립변인을 활용하는 연구에서는 활용하기 어렵다. 따라서 기존 회귀모형에서 약간의 편향(bias)을 허락하는 대신 분산을 줄이는 방식으로 평

균제곱오차(Mean squared error, 이하 MSE)를 조절하여 추정의 정확성과 해석상의 이점을 모색하는 방법을 적용해볼 수 있다(James et al., 2013).

따라서 주요 설명변인을 살펴보기 위하여 국어, 영어, 수학 학업성취도 척도점수를 종속변인으로, 중2 시점의 학생, 학부모 학교 변인 361개를 독립변인으로 하여 LASSO 분석을 실시하였으며, 이를 위한 LASSO 식은 다음과 같다(Tibshirani, 1996; Li et al., 2017).

$$\hat{\beta} = \arg\min \left\{ \frac{1}{2} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\} \qquad \text{식 (Ⅲ-1)}$$

이는 n명의 학생에 대한 정보를 활용하여, 학업성취도 척도점수 y에 영향을 미치는 p개의 독립변인 x에 대한 회귀 계수 β를 추정함을 의미한다. λ는 제약모수(penalty parameter)로서 λ 값에 따라 회귀 계수 β를 얼마나 축소 추정할지 결정하게 되며, 회귀계수가 0이 아닌 변인은 주요 설명변인이라 해석할 수 있다. 이 연구에서는 식 (Ⅲ-1)과 같이 주요 설명변인 선정 모형을 설정하고, 교차타당화(Cross-validation)을 적용하여 과목별 λ값을 설정하였다. 교차타당화에서는 분석에 활용하는 자료[28]를 훈련자료(training data) 중 일부를 검증자료(validation data)로 이용하여 예측오차의 평균를 계산하였으며, 최소 λ와 최대 λ의 비율 0.001 ($\lambda_{\min} = 0.001\lambda_{\max}$)에서의 100개 λ 값에 대하여 20-fold 교차타당화[29]를 적용하였다. λ값은 MSE가 가장 작을 때의 λ값을 선택하거나, 최소 MSE에서 교차타당화 오차의 1 표준오차를 넘어서지 않는 λ값을 선택하는 '1 표준오차 법칙(one standard error rule)' 을 적용하여 상대적으로 모

28) 주요 변인 선정모형에서 활용하는 자료는 전체자료 중 약 80%인 훈련자료를 의미한다.

29) n-fold 교차타당화는 예측오차(prediction error)를 추정하는 데 널리 사용되는 방법이다. 전체 자료를 n부분으로 동일하게 나눈 후 $n-1$개 부분은 훈련자료, 나머지 1개 부분은 평가자료로 설정한 후, 훈련자료로 모형을 만들고 평가자료로 모형의 타당성을 검정하게 되며 이를 n번 반복하면서 MSE를 계산한다(Friedman et al., 2010).

형의 간명성(parsimony)을 고려한 λ값을 선택하게 된다(Hastie et al., 2009, p. 244). 이 연구에서는 후자의 방법에 준하여 λ값을 선택하고, 이때의 예측변인의 계수를 확인하였다. LASSO 모형 분석에는 R 4.0.4.의 glmnet 패키지를 활용하였으며(Friedman et al, 2010), 분석에 활용한 R 명령어를 분석 과정에 따라 <표 Ⅲ-5>에 제시하였다.

<표 Ⅲ-5> LASSO 분석을 위한 R 명령어(R 4.0.4.)

분석 과정	명령어
활용 패키지 로드	library(glmnet)

분석데이터 준비
* 전체 자료 중 80%에 해당하는 훈련 자료 활용(traindata)
- 종속변인을 숫자형 변인으로 변환 후, 전체 data를 예측변인만으로 구성된 X와 종속변인만으로 구성된 Y로 구분
	traindata$종속변인명
	<- as.numeric(traindata$종속변인명)
	
	X = traindata[,-traindata$종속변인명]
	X = scale(X) # scale : mean = 0, std=1
	y = scale(traindata[,traindata$종속변인명])

적절한 람다 값을 찾기 위한 교차타당화 실시	cv20 <- cv.glmnet(x=X, y=y, family='gaussian', type.measure = 'mse', nfolds = 20, alpha=1, trace.it = 1) * family: 종속변인에 따른 분포 설정 * type.measure: 측정지표 설정 * nfolds: 자료를 몇 개의 fold로 나누어 검증할 것인지 설정 * alpha: LASSO 세부모형 지정(1=LASSO, 0=GroupLASSO에 해당) * trace.it: 분석과정 출력여부(1=출력, 0=미출력)
LASSO 분석	lasso <- glmnet(x=X, y=y, lambda=cv20$lambda.1se, family='gaussian', alpha=1) * lambda: CV에서 확인한 lambda 지정 * family: cv.glmnet() 옵션과 동일 * alpha: cv.glmnet() 옵션과 동일

나. 기초미달 학생 예측모형

‘기초미달 학생 예측모형’에서는 앞서 상위 변인으로 나타난 설명변인과 중2 학업성취도 점수를 심층인공신경망 모형의 예측변인으로 활용하여 중3 기초미달 학생을 예측하는 모형을 구축하고자 하였다. 심층인공신경망 모형은 예측변인의 개수, 소수집단 정보 반영방법, 심층인공신경망 모형의 종류에 따라 예측모형을 상세화하였으며, 이를 기초미달 비율이 다른 국어(7.6%), 영어(14.9%), 수학(20.6%) 자료에 적용하여 모형 성능을 비교해보고자 하였다.

심층인공신경망 모형 훈련 시 기초미달 학생 정보가 기초이상 학생 정보만큼 반영될 수 있도록 표집방법과 비용민감방법을 적용하였으며, 심층인공신경망 모형의 종류로는 MLP 모형과 CNN 모형을 고려하였다. 이때, CNN에서는 표형식의 교육패널자료를 이미지화[30]하는 과정에서 변인을 n^2개씩 활용하게 되므로 예측변인의 개수를 n^2개로 지정하였고, 앞서 주요 설명변인을 선정하는 과정에서 계수가 0이 아닌 변인의 개수가 국어 91개, 영어 83개, 수학 84개로 나타남에 따라 예측변인의 개수를 9개, 16개, 25개, 36개, 49개 64개, 81개로 설정하였다. 동일한 조건으로 모형을 설정하기 위하여 MLP에도 예측변인의 수를 동일하게 적용하였다.

소수집단 정보 반영방법과 심층인공신경망 종류에 따른 모형 설정 상세 내용과 분석 절차는 다음과 같다.

1) 불균형 문제 완화를 위한 소수집단 정보 반영방법

이 연구에서 예측하고자 하는 국어, 영어, 수학 기초미달 학생은 전체

30) 합성곱신경망 입력자료를 위한 패널자료 변환 과정은 이후 합성곱신경망 모형 상세화에서 자세하게 다루었다(p.81 참고).

학생 중 7.6%, 14.9%, 20.6%에 해당하는 학생으로, 불균형 자료에 속한다고 할 수 있다[31]. 따라서 훈련자료를 활용하여 예측모형을 훈련시킬 때 원자료를 그대로 활용한다면, 기초이상 학생을 잘 예측하는 방향으로 가중치가 갱신될 가능성이 높다. 따라서 이를 완화에 주기 위한 방법을 고안할 필요가 있으며, 이 연구에서는 표집방법과 비용민감방법을 적용해 보고자 하였다.

표집방법 중에는 임의과대표집(Random Over Sampling, 이하 ROS) 방법과 함께 선행연구에서 심층인공신경망 모형에 자주 활용되며 불균형 문제 완화에 좋은 성능을 보였던, Borderline-SMOTE(Borderline Synthetic minority over-sampling technique, 이하 B-SMOTE)와 ADASYN(Adaptive synthetic sampling approach)도 함께 비교하였다. 각 표집방법의 적용 알고리즘은 다음과 같다.

우선 ROS은 소수집단의 사례를 중복하여 추출하는 방법을 취하므로, 국어, 영어, 수학 기초미달 학생을 기초이상 학생과 1:1 비율이 될 때까지 재추출하였다.

다음으로 B-SMOTE는 SMOTE를 확장한 방법으로서 SMOTE 알고리즘을 적용하되, 경계값에 있는 사례만을 활용하여 자료를 추출하는 방법이다(Han et al., 2005). 구체적인 알고리즘은 다음과 같으며, 최근접 사례는 5개로 지정하였다.

---

31) 자료의 불균형 정도는 불균형 비율(Imbalanced Ratio, IR) (다수집단 사례 수/소수집단 사례 수) 을 기준으로 살펴볼 수 있으며, 4 미만은 낮은 불균형, 4에서 10 미만은 중간 정도의 불균형, 10 이상은 높은 불균형을 띄는 자료임을 나타낸다(Fernandez et al., 2008).

알고리즘 1. Borderline-SMOTE

소수집단을 T라 한다.

1. 집단 T의 모든 사례에 대하여 최근접 사례 k개를 확인한다.
2. k개 사례 중 절반 이상이 다수집단의 사례(경계면에 있는 사례)를 집단 t라 한다.
3. 집단t에서 임의의 한 사례를 선택한 후 최근접 사례 k개를 확인한다.
4. k개 사례 중 임의의 한 사례를 선택한다.
5. 선택된 표본 쌍의 직선 거리에 0과 1사이의 값을 곱한다.
6. 그 지점에 위치하는 새로운 사례를 생성하고 T에 추가한다.
7. 이를 T와 다수집단의 비율이 1:1이 될 때까지 반복한다.

ADASYN 역시 SMOTE를 확장한 방법 중 하나로, SMOTE는 소수집단의 사례에서 동일한 개수의 사례를 생성하였지만, ADASYN은 각 사례마다 생성하는 사례의 수를 다르게 한다는 특징이 있다. 각 사례의 최근접 사례에 다수집단이 있는 정도를 가중치로 하여, 다수집단과 구분이 어려운 사례의 정보를 늘려주고자 하는 방법이다. 구체적인 알고리즘은 다음과 같으며, 역시 최근접 사례는 5로 지정해주었다.

알고리즘 2. ADASYN

소수집단을 T라 한다.

1. 집단 T의 모든 사례에 대하여 최근접 사례 k개를 확인한다.
2. k개 사례 중 다수집단의 사례의 비율을 구하고 이를 r_i라 한다.
3. 집단 T의 모든 사례에 대하여 r_i을 구하고 이를 표준화 한다.
   $\hat{r_i} = r_i / \sum_i r_i$
4. 생성해야 하는 사례의 개수를 구한다. ($g_i =$ 다수집단사례수 $- T$)
5. k개 사례 중 임의의 한 사례를 선택하고 $g_i \times r_i$를 확인한다.
6. $g_i \times r_i$가 0인 아닐 때, 최근접 사례 k개를 찾고 그중 한 사례를 임의로 선택한다.

7. 선택된 표본 쌍의 직선 거리에 0과 1사이의 값을 곱하고, 그 지점에 위치하는 새로운 사례를 생성한다. .
8. 6과 7을 $g_i \times r_i$번 반복한다.
9. 5에서 8을 집단 T의 모든 사례에 대하여 반복한다.

3가지 과대표집 자료 생성에는 R 4.0.4의 ROSE(Lunardon et al, 2014), smotefamily(Siriseriwan, 2019) 패키지를 사용하였으며, 분석에 활용한 R 명령어를 분석 과정에 따라 <표 Ⅲ-6>에 제시하였다.

<표 Ⅲ-6> 표집방법 적용을 위한 R 명령어(R 4.0.4.)

분석 과정	명령어
활용 패키지 로드	library(ROSE) #ROS library(smotefamily) #B-SMOTE, ADASYN
분석데이터 준비 * 전체 자료 중 80%에 해당하는 훈련 자료 활용(traindata)	X = traindata[,-traindata$종속변인명] y = traindata[,traindata$종속변인명]
ROS 방법을 적용하여 기초미달 학생 수를 기초이상 학생 수만큼 조정	ROS <- ovun.sample(종속변인명~., data = traindata, method = "over", N = n)$traindata * method: 적용하고자 하는 표집방법 지정 * N: 표집 적용 후 최종 사례수(과목별 불균형 비율에 따라 균형 자료가 되도록 지정)
B-SMOTE 방법을 적용하여 기초미달 학생 수를 기초이상 학생 수만큼 조정	SMOTE <- BLSMOTE(X, y, K=5, method = "type1") * K: 최근접사례 개수 지정 * C: Han et al. (2005) 연구에 제시된 type1, type2 방법 중 활용하고자 하는 방법을 지정
ADASYN 방법을 적용하여 기초미달 학생 수를 기초이상 학생 수만큼 조정	ADA <- ADAS(X, y, K=5) * K: 최근접사례 개수 지정

표집방법이 자료 수준에서 집단 간 균형을 맞춰주는 방법이라면, 비용민감방법은 알고리즘 수준에서 불균형 정도를 조정하는 방법이다. 심층인공신경망에서는 모형을 반복적으로 훈련시키는 과정에서 매 훈련 마다 예측 값과 실제 값의 차이를 반영하여 가중치를 갱신하는데, 비용민감방법에서는 소수집단의 손실함수(loss function) 값에 높은 가중치를 부여한다. 이 연구에서는 소수집단과 다수집단이 모형 훈련에 미치는 영향이 동일하게 반영될 수 있도록 집단별 가중치(w)를 다음과 같이 조정하였다.

$$w_{\text{소수집단}} = \frac{\text{다수집단 학생 수}}{\text{전체 학생 수}}, \quad w_{\text{다수집단}} = \frac{\text{소수집단 학생 수}}{\text{전체 학생 수}} \qquad \text{식 (Ⅲ-2)}$$

비용민감방법은 심층인공신경망 모형 훈련 시 가중치를 직접 지정해 주었고, 이를 적용한 Python 명령어는 <표 Ⅲ-7>과 같다.

<표 III-7> 비용민감기법 적용을 위한 Python 명령어(Python 3.7.1.)

분석 과정	명령어
class 가중치 지정	class_weight = {1 : $w_{\text{소수집단}}$, 0 : $w_{\text{다수집단}}$}
심층인공신경망 모형 훈련 시 옵션으로 지정(볼드체 부분에 해당함)	model.fit(train_x, train_y, epochs=epochs, batch_size=batch_size, verbose=1, validation_split=0.2, **class_weight=class_weight**)

2) 다층퍼셉트론 모형

연구에서 활용한 '다층퍼셉트론 모형(Multi-layer perceptron, 이하 MLP)'은 [그림 Ⅲ-3]과 같다. 입력층 노드는 예측변인 개수와 동일하며, 출력층은 기초미달 여부를 1과 0으로 출력하는 1개 노드로 구성되어 있다. 은닉층과 노드 수 설정을 위하여 교육패널자료와 유사한 정형자료를 활용한 선행연구의 모형을 참고하여 은닉층 3~8개인 MLP 모형을 설정[32] 한 후, 과적합이 의심되지 않으며 모형 성능이 높게 나타난 모형을 선정하였다. 최종적으로 은닉층 총 6개이며 입력층에 이어지는 은닉층부터 64개 노드에서 1/2만큼 감소하는 형태로 설정하였다.

단, 예측변인의 개수를 64개, 81개 활용하는 모형에서는 예측변인의 개수가 적은 경우와 동일하게 모형을 설정하는 경우, 과적합으로 인한 성능 감소가 의심되는 부분이 있었다. 모든 모형에서 예측변인의 개수 64개를 기점으로 모형 성능이 크게 하락하는 모습이 나타났으며, 모형 훈련 과정에서의 검증자료(validation data)로 확인한 모형 성능과 모형 훈련이 끝난 후 평가자료(test data)로 확인한 모형 성능 차이가 예측변인 개수가 적은 모형에 비해 큰 경향이 있었다.

따라서 64개, 81개 예측변인을 활용하는 경우에는 입력층과 가까운 은닉층을 1개(노드 수 128) 추가[33]하여 모형을 훈련시키고 그 결과를 제

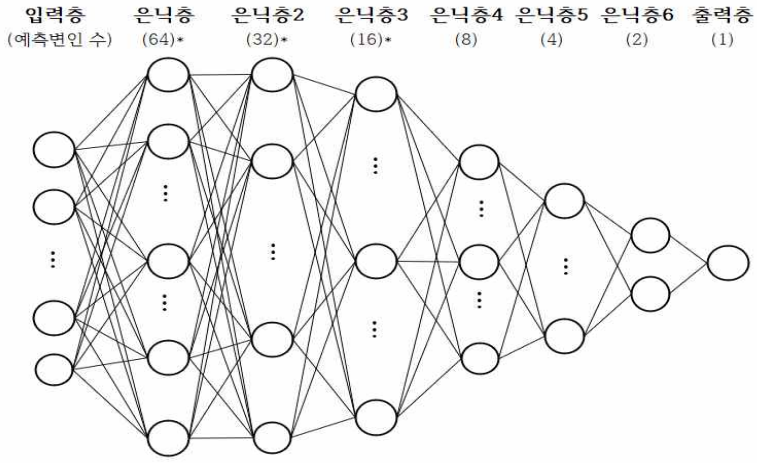
---

32) 심층인공신경망 모형의 초모수(Hyperparameter)를 지정하기 위한 분명한 기준은 찾아보기 어렵다. 하지만 일반적으로 모형의 성능이 높게 나타나는 모형을 기준으로 은닉층 및 노드 개수 등을 선정하게 된다(Hagan et al., 2014/2018, p. 783). 따라서 모형 훈련 시 패널자료와 유사한 정형데이터를 활용한 선행연구의 모형을 참고하여 은닉층의 수가 3개에서 8개인 다층퍼셉트론 모형을 설정한 후, 모형 성능이 높게 나타난 은닉층이 6개인 모형을 활용하였다. 노드 수는 출력층으로 갈수록 정보가 수렴될 수 있도록 노드 수가 적어지게 설정하였으며, 이는 역시 선행연구의 모형을 참고하였으며, 출력층 노드가 한 개인 이진분류 상황인 점을 고려하였다.

33) 적절한 모형을 찾기 위하여 은닉층을 1~3개를 추가하여 모형을 훈련시키고 그 결과를 살펴본 결과, 추가한 은닉층 개수에 따른 모형 성능의 차이는 크지 않았다. 하지만 기존에 설정한 총 은닉층 6개인 모형에 비해 모형 성능이 개선된 모습이 나타났고, 모형 훈련 시 검증자료(validation data)로 확인한 모

시하였다.

은닉층 활성화 함수는 ReLU(Rectified Linear Unit) 함수로 지정하였으며, 출력층 활성화 함수는 이진분류에 적합한 시그모이드(sigmoid) 함수로 지정하였으며 임계점(threshold) 0.5를 기준으로 하였다. 또한 출력 값이 이분 변인이므로 손실함수(loss function)는 이분 교차엔트로피(Binary cross entropy)로 설정하였다. 모형의 최적화를 위하여 Adam(Adaptive moment equation) 방법으로 가중치를 조정하도록 하였으며, 학습률(learning rate)은 10^{-5}로 지정하였다. 가중치 추정 과정에서 검증오차(validation error)의 변화가 안정적이게 될 때까지 가중치 갱신을 충분히 반복하였으며, 이를 위해 학습 횟수를 500주기(epoch)로 설정하였다. 또한 가중치 추정과정에서 과적합을 방지하기 위하여 노드의 수가 상대적으로 많은 입력층 이후의 3개 은닉층(예측변인 개수가 61개, 81개인 모형에서는 입력층 이후 4개 은닉층)의 노드 중 20% 노드에 드롭아웃(dropout)을 적용하였다.



[그림 III-3] 다층퍼셉트론 연구모형

주. 드롭아웃을 적용한 은닉층에 *표시함.

형 성능과 최종 모형을 평가자료(test data)로 확인한 모형 성능 간 차이가 줄어드는 모습을 확인할 수 있었다.

MLP 모형 분석에는 Python(3.7.1.) 기반의 텐서플로(Tensorflow) 2.0을 활용하였으며, Google Colab GPU 환경에서 분석을 진행하였다. 분석 과정에 따른 Python 명령어는 <표 Ⅳ-8>와 같다.

<표 Ⅲ-8> 다층퍼셉트론 모형 훈련을 위한 Python 명령어(Python 3.7.1.)

분석 과정	명령어
활용 패키지 로드 및 축약어 지정	import tensorflow as tf from tensorflow import keras import numpy as np from sklearn.metrics import confusion_matrix

분석자료 준비
- 훈련자료(80%)와 평가자료(20%)로 구분
	#주요 설명변인 선정 단계의 결과에 따라 과목별 분석 자료의 변인 순서를 재조정
	od_var = {1순위 변인명 : 0, 2순위 변인명 : 1,
	..., 81순위 변인명: 80}
	train_x = sorted(traindata.drop('종속변인명' , axis=1),
	key=lambda x : od_var)
	train_y = traindata[['종속변인명']]
	
	test_x = sorted(testdata.drop('종속변인명' , axis=1),
	key=lambda x : od_var)
	test_y = test[['종속변인명']]

초모수 지정	epochs = 500 batch_size = 10 class_weight = {1 : $w_{\text{소수집단}}$, 0 : $w_{\text{다수집단}}$} mlp_feature = 예측변인 수 * mlp_feature: 연구에서는 9, 16, 25, 36, 49, 64, 81개로 지정하였으며, 상위 n개 변인만 예측변인으로 활용하도록 함.
MLP 모형 설정	np.random.seed(5) model = tf.keras.Sequential([tf.keras.layers.Dense(64, input_shape=(n_feature,) , activation='relu'), tf.keras.layers.Dropout(.2), tf.keras.layers.Dense(32, activation='relu'), tf.keras.layers.Dropout(.2), tf.keras.layers.Dense(16, activation='relu'), tf.keras.layers.Dropout(.2),

분석 과정	명령어
	tf.keras.layers.Dense(8, activation='relu'), tf.keras.layers.Dense(4, activation='relu'), tf.keras.layers.Dense(2, activation='relu'), tf.keras.layers.Dense(1, activation='sigmoid')]) model.compile(optimizer=tf.keras.optimizers.Adam (learning_rate=1e-5), loss='binary_crossentropy', metrics=['accuracy', tf.keras.metrics.AUC(), tf.keras.metrics.Recall(), tf.keras.metrics.Precision()])
모형훈련	model.fit(train_x, train_y, epochs=epochs, batch_size=batch_size, verbose=1, validation_split=0.2, class_weight=class_weight)
모형검증 * 임계점 0.5 기준	model_sigmoid = model.predict(test_x) model_pridict = tf.cast(model_sigmoid > 0.5 , dtype = tf.float32) output = confusion_matrix(test_y, model_pridict)

3) 합성곱신경망 모형

연구에서 활용한 ‘합성곱신경망 모형(Convolutional neural network, 이하 CNN)’은 [그림 Ⅲ-5]와 같다. 앞서 2절에서 언급하였듯이 다층퍼셉트론에서는 자료의 특징(예측변인)을 지정해주고 이에 따라 결과를 추정하지만, CNN에서는 자료의 특징을 모형에서 추출하고 이를 활용하여 결과 값을 출력하게 된다. 따라서 합성곱 계층(convolutional layer) 및 통합 계층(pooling layer)과 완전연결층(fully-connected layer) 분석과 관련된

초모수(hyperparameter)를 각각 지정해 주어야 한다.

우선 합성곱신경망 모형의 입력층은 상위 주요변인 p개를 활용하는 경우 $\sqrt{p} \times \sqrt{p}$ 형태로 입력하였다. 이를 위해 행(사례)×열(변인) 형태의 표형식으로 저장되는 정형데이터인 패널자료를 이미지 형태로 변환해 주는 과정이 필요하며, CNN 입력을 위한 자료 변환 방법은 [그림 Ⅲ-4]과 같다. [그림 Ⅲ-4]의 (a)에서는 교육패널자료의 일반적인 저장 형태의 예를 볼 수 있으며, MLP에서는 각 변인의 하나의 입력 노드로 지정되었다. CNN에서는 입력자료가 (b)와 같이 이미지 형태가 되며, 예측변인을 9개 사용하는 경우 (a)의 9개 변인이 (b)의 각 셀에 위치하게 된다. 따라서 CNN 분석에서는 (a) 형태의 교육패널자료를 (c)에 제시된 학생1, 학생2 자료와 같이 변환하여 활용하였다.

(a) 교육패널자료의 일반적인 저장 형태

	변인1	변인2	변인3	변인4	변인5	변인6	변인7	변인8	변인9	…
학생1	8	2	3	3	2	2	3	3	2	
학생2	9	1	2	3	1	1	2	3	1	
학생3	10	1	3	1	5	1	3	1	5	
학생4	2	2	5	5	2	2	5	5	2	
학생5	1	2	5	5	4	2	5	5	4	
⋮										

(b) CNN 입력의 예

변인1	변인2	변인3
변인4	변인5	변인6
변인7	변인8	변인9

(c) 학생1, 학생2 변환 자료

8	2	3
3	2	2
3	3	2

9	1	2
3	1	1
2	3	1

[그림 Ⅲ-4] CNN 입력을 위한 자료 변환의 예

출력층은 MLP 모형과 동일하게 기초미달 여부를 1과 0으로 출력하는 1개 노드이다. 이후 '합성곱 및 통합 계층'에서는 필터를 3×3 합성곱 필터를 스트라이드(stride) 1로 하여 32개를 적용한 후, 2×2 통합 필터를 스트라이드 2로 최대통합(max pooling)하였다[34]. 이는 최소로 활용하는

예측변인이 9개임을 고려하여 이때 최대로 활용할 수 있는 필터 값을 지정한 것이라 할 수 있다.

그러나 예측변인 49개 이하를 사용한 모형과 달리 예측변인 64개, 81개를 활용하는 모형에서는 동일하게 모형을 설정하였을 때 과적합이 의심되는 부분이 있었고, 모형 성능이 크게 하락하는 모습이 나타나기도 하였다. 이는 예측변인이 많아져 고차원 입력자료가 되는 상황이지만 저차원 입력자료와 동일하게 모형을 설명했기 때문일 수 있다. 따라서 예측변인 64개, 81개를 활용하는 모형에서는 동일한 필터를 사용하는 합성곱 계층과 통합 계층을 1개 추가[35]하여 모형을 설정하였다.

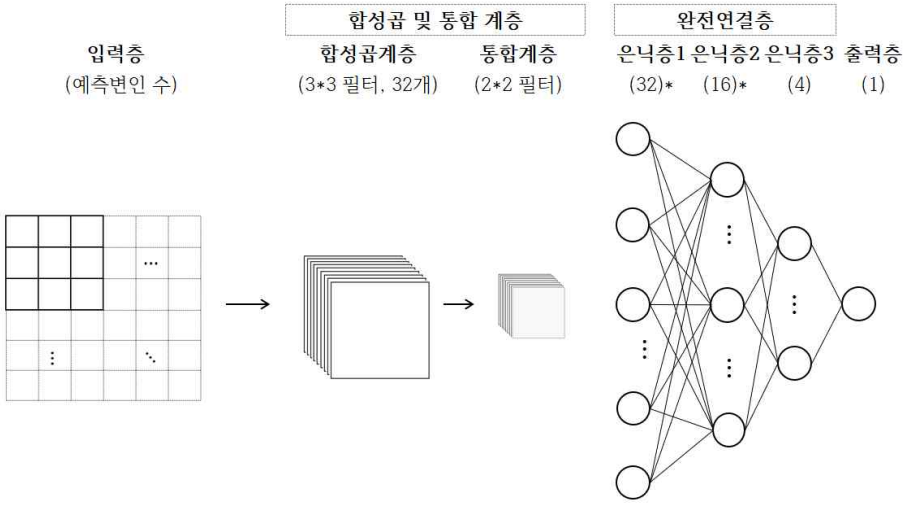
이후 '완전연결층'에서는 합성곱 및 통합 계층 결과를 1×1 벡터(vector)로 변환하여 입력 자료로 활용하였으며, 모든 모형에 동일하게 은닉층 3개[36], 각 은닉층의 노드는 32개, 16개 4개로 지정하였다. 은닉층 및 출력층의 활성화 함수는 다층퍼셉트론 모형과 동일하게 지정하였으며, 손실함수 및 가중치 조정을 위한 최적화 방법, 학습률 또한 동일하게 지정하였다. 또한 합성곱 신경망에서도 과적합을 방지하기 위하여 노드 수가 상대적으로 많은 앞 2개 은닉층의 노드 중 20% 노드에 드롭아웃을 적용하였다.

---

34) 합성곱신경망 모형의 최소 예측변인은 9개인 점을 고려하여 9개 예측변인을 활용할 때 최대로 설정할 수 있는 합성곱 필터를 기준으로 모형을 상세화하였다.

35) 예측변인의 개수가 많은 모형에 적합한 모형을 탐색하기 위하여 합성곱 계층 및 통합 계측의 필터는 예측변인 개수가 적을 때와 동일하게 유지한 채 완전연결층의 은닉층의 개수를 늘려 그 결과를 살펴보았으나, 완전연결층 은닉층 개수 증감에 따른 모형 성능 변화는 크지 않았다.

36) 합성곱신경망 모형에서는 합성곱 및 통합 계층에서 주요 특징을 추리는 과정이 있기 때문에 완전연결층의 깊이가 깊지 않은 경향이 있다. 따라서 '완전연결층'의 은닉층의 개수를 결정하는 과정에서 은닉층 개수를 2~6개로 지정하여 모형 성능을 살펴본 결과 완전연결층 은닉층 개수에 따른 결과 차이가 크지 않았다. 따라서 전반적으로 모형 성능이 높게 나타난 은닉층 3개 모형을 최종 모형으로 설정하였다.



[그림 III-5] 합성곱신경망 연구모형

주. 드롭아웃을 적용한 은닉층에 *표시함.

CNN 모형 분석에는 Python(3.7.1.) 기반의 텐서플로(Tensorflow) 2.0을 활용하였으며, Google Colab GPU 환경에서 분석을 진행하였다. 분석 과정에 따른 Python 명령어는 <표 Ⅳ-9>와 같다.

<표 III-9> 합성곱신경망 모형 훈련을 위한 Python 명령어(Python 3.7.1.)

분석 과정	명령어
활용 패키지 로드 및 축약어 지정	import tensorflow as tf from tensorflow import keras import numpy as np from sklearn.metrics import confusion_matrix
분석자료 준비 - 훈련자료(80%)와 평가자료(20%)로 구분	train_x, train_y, test_x, test_y * MLP 모형에서 활용한 훈련자료와 평가자료 활용 train_x = train_x.to_numpy() train_x = train_x.reshape(train_x.shape[0], cnn_feature, cnn_feature, 1) * cnn_feature = 예측변인 수 양의 제곱근 값 #train_x 변환을 train_y, test_x, test_y에도 동일하게 적용

분석 과정	명령어
초모수 지정	epochs = 500 batch_size = 10 class_weight = {1 : $w_{소수집단}$, 0 : $w_{다수집단}$} cnn_feature = 예측변인 수의 양의 제곱근 값 * cnn_feature: 연구에서는 3, 4, 5, 6, 7, 8, 9개로 지정 input_shape = (n_feature, n_rowcol, 1)
합성곱신경망 모형 설정	model = Sequential() model.add(Conv2D(32, kernel_size=(3, 3), strides=(1, 1), padding='same', activation='relu', input_shape=input_shape)) model.add(MaxPooling2D(pool_size=(2, 2), strides=(2, 2))) model.add(Flatten()) model.add(Dense(32, activation='relu')) model.add(Dropout(.2)) model.add(Dense(16, activation='relu')) model.add(Dropout(.2)) model.add(Dense(4, activation='relu')) model.add(Dense(1, activation='sigmoid')) model.summary() model.compile(optimizer=tf.keras.optimizers.Adam (learning_rate=1e-5), loss='binary_crossentropy', metrics=['accuracy', tf.keras.metrics.AUC(), tf.keras.metrics.Recall(), tf.keras.metrics.Precision()])
모형훈련	model.fit(train_x, train_y, epochs=epochs, batch_size=batch_size, verbose=1, validation_split=0.2, class_weight=class_weight)
모형검증 * 임계점 0.5 기준	model_sigmoid = model.predict(x_test_sampling) model_pridict = tf.cast(model_sigmoid > 0.5 , dtype = tf.float32)

4) 소수집단 예측모형에서의 모형평가

심층인공신경망 모형의 평가는 모형 훈련에 활용하지 않은 평가자료(test data)를 훈련된 모형에 적용하고, 그 결과를 살펴보는 과정을 취한다. 따라서 이 연구에서도 평가자료를 예측모형에 적용하였을 때 나타난 결과에 비추어 각 예측모형을 평가하고, 여러 예측모형 간 예측 성능 차이를 비교하였다.

먼저, 예측모형의 성능을 평가하기 위하여, 예측 결과를 <표 Ⅲ-10>과 같은 혼동행렬(Confusion matrix)를 활용할 수 있다. 여기서 True Negative(TN), True Positive(TP)는 실제 값과 예측한 결과가 일치한 사례, False Negative(FN), False Positive(FP)는 실제 값과 예측한 결과가 일치하지 않은 사례를 의미하며, 예측모형은 FN과 FP를 최소화하는 것을 목표로 한다.

<표 III-10> 기초미달 학생 실제 값과 예측 값에 대한 혼동행렬

		예측	
		기초학력 이상(0)	기초미달(1)
실제	기초학력 이상(0)	True Negative(TN)	False Positive(FP)
	기초미달(1)	False Negative(FN)	True Positive(TP)

혼동행렬 각 셀의 값을 활용하여 다음과 같은 모형 지표를 확인할 수 있다. 정분류율(Accuracy)는 전체 사례 중 제대로 분류한 사례의 비율을 말한다. 민감도(Sensitivity 또는 Recall)는 실제 기초미달 학생 중에서 기로미달 학생이라고 제대로 예측한 사례의 비율을, 특이도(Specificity)는 기초이상 학생 중 제대로 예측한 사례의 비율을 의미한다. 마지막으로 정밀도(Presicision)는 예측모형에서 기초미달 학생이라고 예측한 사례 중에서 실제 기초미달 학생인 학생의 비율을 말한다.

$$\text{정분류율}(Accuracy) = \frac{TN + TP}{TN + FP + FN + TP}$$

$$\text{민감도}(Sensitivity, Recall) = \frac{TP}{FN + TP}$$

$$\text{특이도}(Specificity) = \frac{TN}{TN + FP}$$

$$\text{정밀도}(\text{Pr}ecision) = \frac{TP}{FP + TP})$$

일반적으로는 전체 사례 중 제대로 분류한 비율인 '정분류율'을 기준으로 모형을 평가하지만, 불균형 자료에서는 정분류율로 모형 성능을 평가하긴 어렵다. 만약 기초미달 학생이 전체의 1%라고 한다면, 단순하게 모든 학생을 기초학력 이상이라고 예측해도 정분류율이 99%로 나타나기 때문이다. 일반적으로 불균형 자료의 예측모형은 소수집단을 제대로 예측하는 데 초점이 있으며, 다수집단을 잘못 예측(FP)하는 것에 비해 소수집단을 잘못 예측(FN)하는 것이 더 큰 문제를 초래할 수 있다고 본다(방성완, 김재오, 2020; Weng & Poon, 2008). 예를 들어 실제 기초학력 이상 학생을 기초미달로 잘못 분류(FP)를 한 경우에는 추가적으로 교육적 지원을 받게 되지만, 실제 기초미달 학생을 기초학력 이상으로 잘못 분류(FN)를 한 경우에는 지원 대상에서 제외됨으로써 학습 결손이 심화될 수 있기 때문이다.

따라서 불균형 자료를 활용한 예측모형의 모형 성능을 확인하기 위해서는 정분류율과 함께 이를 보완할 수 있는 모형 지표들을 함께 확인해야 한다. 가장 기본적인 지표로는 실제 기초미달 학생을 제대로 분류한 비율을 의미하는 '민감도'가 사용되며, 이와 함께 민감도 향상에 따라 기초학력 이상 학생 예측이 하락하는 정도를 동시에 고려한 지표인 AUC, F-score, G-mean 지표를 함께 살펴보는 것이 권장된다(Barros et

al., 2019; Weng & Poon, 2008).

AUC(Area Under the Curve; ROC Area)는 임계값(threshold) 조정에 따른 민감도와 1-특이도 간 균형을 확인할 수 있는 ROC(Receiver Operating Characteristic) 곡선 아래 영역을 수치화한 지표이다(Weng & Poon, 2008). 임계값에 따른 변화를 확인함으로써 기초학력 이상 학생을 잘못 분류하는 비율을 최소화하면서 기초미달 학생을 제대로 분류하는 성능을 평가하는 데 유용하게 활용된다.

F-score는 민감도와 정밀도의 조화평균 값이며, 베타 값을 지정하여 민감도와 정밀도의 중요도를 다르게 설정할 수도 있다. 두 값을 동등하게 반영한 경우는 베타가 1인 F1-score이며, 불균형 자료에서는 FP보다 FN 사례에 더 가중치를 두기 때문에 베타를 2로 지정한 F2-score을 사용한다(Jeni et al., 2013).

$$F-score = (1+\beta^2)\times\frac{\text{정밀도}\times\text{민감도}}{(\beta^2\times\text{정밀도}+\text{민감도})}$$

G-mean은 민감도와 특이도의 기하평균을 말한다. 전체 사례를 모두 0 또는 1로 분류하는 경우 민감도 또는 특이도는 1이 되지만 그 외 지표는 0이 때문에 두 개의 값을 모두 고려할 필요가 있다(Wu & Chang, 2003).

$$G-mean = \sqrt{\text{민감도}\times\text{특이도}}$$

AUC, F-score, G-mean 모두 값이 1에 가까울수록 모형의 예측 성능이 높다는 것을 의미한다[37]. 세 지표 모두 불균형 자료 예측 성능을 측

37) AUC는 0.5에서 1의 값을, F-score, G-mean은 0에서 1 사이의 값을 가진다. AUC의 경우, Joseph(2005)가 13단계로 제시한 판단기준에 따르면, 0.5는 분류능력이 전혀 없는 모형임을 의미하며, 0.7-0.8 사이의 값을 가지면 괜찮은 모형, 0.8 이상이면 우수한 모형이라 판단할 수 있다(홍종선, 장동환, 2020). 하지만 불균형 자료의 모형평가에서는 불균형 정도가 클수록 모형 지표는 더

정하는 데 적절한 지표이나, G-mean에 비해 F-score과 AUC가 불균형 자료 예측 성능 판단에 더 적절하다고 알려져 있으며(Davis & Goadrich, 2006; Jeni et al., 2013; Weng & Poon, 2008), 불균형 정도가 심해질수록 F-score가 안정적일 수 있다는 연구결과도 있다(Jeni et al., 2013). 따라서 기초미달 학생 예측모형의 성능 평가에는 민감도, AUC, F-score, G-mean을 종합적으로 살펴보고자 하였다.

또한, 예측모형 간 성능 비교를 위하여 McNemar 검정을 활용하였다(Naqa et al., 2015, pp. 52-53). McNemar 검정은 동일한 자료에 적용한 두 개의 모형 간 분류 및 예측 성능 차이가 있는지 확인할 수 있는 방법이다. 예측모형의 혼동행렬에 비추어 다음과 같은 검정 지표를 계산하고 알 이를 χ^2분포에서 검증한다.

$$\chi^2 = (|C_{01} - C_{10}| - 1)^2 / (C_{01} + C_{10})$$

비교하고자 하는 두 개의 예측모형이 있을 때, C_{01}은 두 번째 모형에서는 제대로 예측되었지만, 첫 번째 모형에서는 잘못 예측된 사례, C_{10}는 첫 번째 모형에서는 제대로 예측되었지만, 두 번째 모형에서는 잘못 예측된 사례를 의미한다. 이를 토대로 두 예측모형에 따른 예측 결과 차이가 통계적으로 유의미한지 확인할 수 있다.

낮게 나타날 수 있다(Jeni et al., 2013).

Ⅳ. 연구결과

1. 주요 설명변인 선정 결과

가. 과목별 설명변인 선정 결과

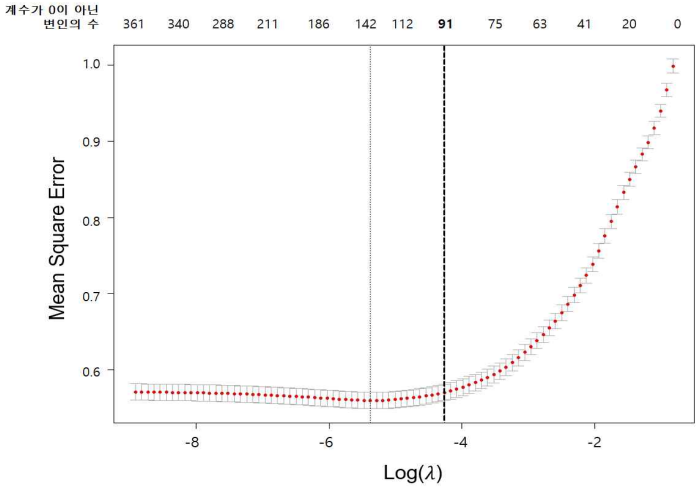
기초미달 학생 예측모형의 예측변인을 선정하기 위하여, 학생의 학업성취도와 관련이 높은 설명변인을 살펴보고자 하였다. 이를 위하여 LASSO(Least Absolute Shrinkage and Selection Operator) 분석을 실시하였으며, 국어, 영어, 수학 과목별 중2 학업성취도 척도점수와 관련이 높은 설명변인을 확인하였다. 총 361개 설명변인 중 계수가 0이 아닌 변인은 국어 91개, 영어 84개, 수학 83개로 나타났다.

1) 국어

중2 시기의 국어 학업성취도를 설명하는 데 기여하는 변인을 확인하기 위하여 먼저, 20-fold 교차타당화(Cross-validation)를 통해 제약모수인 λ 값에 따른 평균제곱오차(Mean Square Error, 이하 MSE)와 표준오차를 살펴보았으며, 이는 [그림 Ⅳ-1]과 같다. MSE가 가장 적은 값을 기준으로 1 표준오차 법칙(one standard error rule)[38]을 적용한 결과, 제약모수 λ 값은 0.014였으며, 이를 기준으로 LASSO 모형을 적합하였을 때 계수가 0이 아닌 변인은 91개로 나타났다.

---

38) 1 표준오차 법칙(one standard error rule)이란 교차타당화에서 모형의 간명성(parsimony)을 고려한 λ값을 선택하고자 할 때 적용하는 방법으로, 최소 MSE에서 교차타당화 오차의 1 표준오차를 넘어서지 않는 최대의 λ값을 선택한다(Hastie et al., 2009, p. 244).



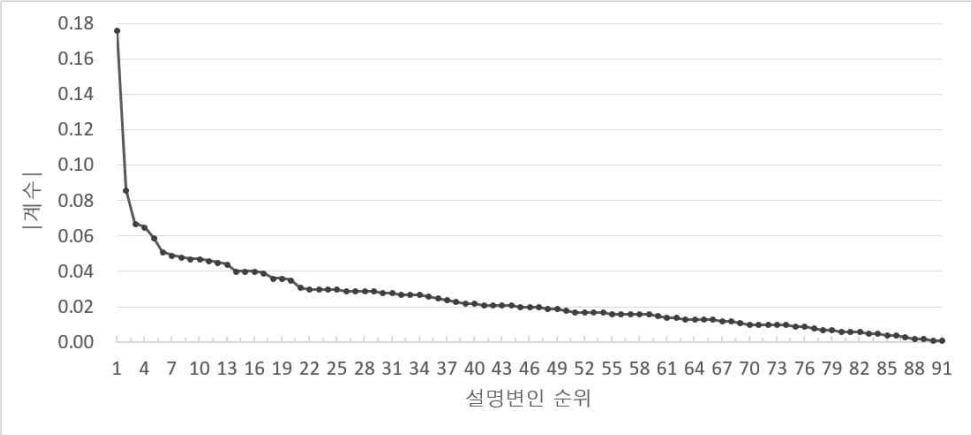
[그림 Ⅳ-1] 국어 중2 학업성취도 제약모수의 20-fold 교차타당화 결과

국어 학업성취도를 설명하는 91개 변인과 계수는 <표 Ⅳ-1>과 같으며, 분석에 활용하는 모든 변인은 값을 표준화하였기 때문에 LASSO 회귀 크기에 비추어 상대적인 기여도를 확인할 수 있다. 따라서 계수 절대값을 기준으로 기여도가 높은 순으로 설명변인을 정리하였으며, 각 변인의 세부 문항은 <표 Ⅲ-2>와 부록에 제시하였다.

설명변인 순위에 따른 LASSO 계수 절대값 크기는 [그림 Ⅳ-2]와 같다. 그래프를 살펴보면 1위, 2위 변인의 기여도가 나머지 변인에 비해 상당히 높은 모습이며, 상위 5번째 변인 이후 점진적으로 감소하고 있는 모습을 나타났고, 상위 70개 변인 이후로는 계수 절대값이 0.01 이하로 거의 0에 가까운 계수가 나타나기도 하였다. 계수 절대값이 0.01이라는 것은 해당 변인이 1 표준편차만큼 변화할 때, 국어 학업성취도가 0.01 표준편차 변화한다는 것을 의미하므로 학업성취도에 따른 학생 간 특징을 구분하는 데 중요한 역할을 하지 못할 수 있음을 짐작해볼 수 있다.

<표 Ⅳ-1> 중2 국어 학업성취도 주요 설명변인 선정 결과

순위	변인명	계수	순위	변인명	계수	순위	변인명	계수
1	수업이해도(국)	0.176	31	SCH수행비율(국)	0.028	61	정신건강(마음편함)	0.014
2	성별(남)	-0.086	32	규칙준수(차례)	0.027	62	진학계획(자사고)	0.014
3	수업이해도(수)	0.067	33	학교크기(학생수)	0.027	63	다문화수용성(이웃)	0.013
4	수업이해도(영)	0.065	34	학교만족도(기초능력향상)	0.027	64	수업집중도(국)	0.013
5	부정적 다문화수용성	-0.059	35	자아개념(어울림)	-0.026	65	PA계획(특성화고)	-0.013
6	수업태도(휴대전화)	-0.051	36	선행학습정도(수)	0.025	66	상호작용(책읽고대화)	0.013
7	SCH선다형비율(국)	0.049	37	수업방식(질문수용)	0.024	67	휴대전화중독(불편)	0.012
8	독서활동(책끝까지)	0.048	38	PA계획(국/공립대)	0.023	68	수업방식(맞춤형과제)	0.012
9	문화활동(콘서트)	-0.047	39	SCH혁신학교	0.022	69	자기관리(정리)	0.011
10	직업관(인생에 중요)	0.047	40	학업지원(일정관리)	-0.022	70	교사관계(이해)	0.010
11	정신건강(죽고싶다)	-0.046	41	여가활동(방과후학교)	0.021	71	여가활동(TV)	-0.010
12	직업관(수입만크면됨)	-0.045	42	직업관(좋아하는일선택)	0.021	72	여가활동(숙제)	0.010
13	수업분위기(딴짓)	-0.044	43	학폭경험(SNS)	-0.021	73	직업관(수단)	-0.010
14	독서활동(취미)	0.040	44	수업태도(잠)	-0.021	74	진로상담실 운영	0.010
15	독서활동(서점,도서관)	0.040	45	학교의사결정(규칙공유)	0.020	75	진학계획(예고/체고)	-0.009
16	자아개념(단체생활)	-0.040	46	규칙준수(학급규칙)	0.020	76	정서적지원(격려)	0.009
17	SCH사립	0.039	47	진로프로도움(고대방문)	-0.020	77	선행학습정도(영)	0.008
18	PA기대학력	0.036	48	공동체의식(조별학습)	-0.019	78	진로성숙도(학교공부의미)	0.007
19	진로성숙도(구체화)	0.036	49	부모상호작용(학교활동참여)	0.019	79	다문화수용성(친구)	0.007
20	진로프로도움(현장체험)	-0.035	50	SCH수행평가비율(수)	0.018	80	SCH구성원학교운영(논의)	0.006
21	직업관(의미X)	-0.031	51	소외(스트레스)	-0.017	81	공동체의식(친구와상의)	0.006
22	PA계획(외고/국제고)	0.030	52	공동체의식(빌려줌)	0.017	82	진로프로도움(모의직업)	-0.006
23	PA계획(과고/영재고)	0.030	53	수업분위기(집중)	0.017	83	자아개념(신체)	-0.005
24	자아개념(운동신경)	-0.030	54	SCH특별시	0.017	84	공동체의식(친구도움)	0.005
25	공동체의식(투표참여)	0.030	55	문화활동(박물관)	0.016	85	자기관리(스스로챙김)	0.004
26	자기관리(미루지않음)	-0.029	56	공동체의식(소외된친구)	0.016	86	미디어사용(검색)	0.004
27	교우관계(고민나눔)	-0.029	57	자기관리(숙제<노는것)	-0.016	87	상호작용(학교생활관심)	0.003
28	미디어사용(연락)	-0.029	58	PA계획(자사고)	0.016	88	학교내폭력정도(따돌림)	-0.002
29	여가활동(집안일)	-0.029	59	학교내폭력(욕설)	0.016	89	학폭경험(맞음)	-0.002
30	공부시간	0.028	60	월평균 가정소득	0.015	90	SCH여학교	0.001
						91	학업스트레스	0.001



[그림 Ⅳ-2] 설명변인 순위에 따른 LASSO 계수 (국어)

국어 학업성취도를 설명하는 91개 변인 중 상위 30개 변인을 중심으로 각 변인별 학업성취도와의 관계를 살펴보면 다음과 같다. <표 Ⅳ-2>에 제시되어 있는 상위 30개 설명변인의 문항 세부사항과 척도 및 계수의 방향을 살펴보았을 때, 전반적으로 수업참여 관련 변인의 영향이 크게 나타났으며, 비인지적 특성 변인 중 진로와 관련된 변인과 부정적 다문화수용성, 정신건강과 관련된 변인의 설명력이 큰 것을 확인할 수 있었다. 또한 여가 및 방과후 시간 영역에서는 독서활동과 관련된 변인들이 특징적이었다. 영어, 수학 과목과는 달리 학부모의 교육기대 및 계획과 관련된 변인들은 상대적으로 하위 변인으로 나타났으며, 학교와 관련된 변인 중에는 교육활동과 관련된 변인의 영향이 크게 나타남을 확인할 수 있었다.

영역별 변인의 문항 내용과 국어 학업성취도와의 관계를 살펴보면 다음과 같다. 우선, 학생 응답 중 '수업참여' 영역에서는 국어, 수학, 영어 과목의 수업이해도가 높을수록 국어 학업성취도가 높게 나타났고, 응답 학생이 수업시간에 휴대전화를 보거나, 수업에서 딴짓하는 학생이 많다고 느낄수록 국어 학업성취도가 낮았다.

'비인지적 특성' 영역에서는 부정적인 다문화수용성 정도가 높을수록, 죽고 싶다는 생각을 자주 할수록, 단체생활을 잘하며, 운동신경이 좋

다고 인식할수록 학업성취도가 낮은 경향이 있었다. 반면 어른이 되어 선거나 투표에 참여하겠다고 생각하거나, 오늘 할 일을 미루지 않는 경향이 있는 학생들은 그렇지 않은 학생에 비해 국어 학업성취도가 높게 나타났다. 비인지적 특성 영역에서는 진로와 관련된 변인이 영어 수학 과목에 비해 상대적으로 많이 나타났다. 구체적으로는 미래 인생에 있어 직업 선택이 중요하다고 생각하고, 장래 희망을 이루기 위해 해야 할 일을 구체적으로 생각해볼수록 학업성취도가 높은 반면, 직업 선택에서 돈을 많이 버는 것이 중요하다고 생각하고, 직업 자체가 인생에 큰 도움이 되지 않는다고 생각할수록 학업성취도는 낮게 나타났다.

'여가 및 방과후 시간' 영역에서는 독서와 관련된 활동이 주요 설명 변인으로 나타났다. 학생이 한번 읽기 시작한 책을 끝까지 읽고, 독서를 가장 좋아하는 취미활동으로 생각하며, 서점이나 도서관에 가는 것을 좋아하는 경우에 국어 학업성취도가 높은 경향을 보였다. 또한 방과후에 공부시간이 많을수록 성취도 높게 나타났다. 하지만 문화활동으로 대중가수 콘서트에 자주 가거나, 방과 후에 집안일을 하는 시간이 많을수록 학업성취도가 낮았다.

이 외에도 학생 응답에서는 '개인특성' 에서 남학생이 여학생이 비해 학업성취도가 낮은 것을 확인할 수 있었으며, '학교생활과 적응' 영역에서 현장체험형 진로프로그램 도움 정도와 친구에게 내 고민을 털어놓는 정도는 국어 학업성취도와 부적인 관계를 보였다.

다음으로 학부모 응답에서는 '교육기대 및 계획' 영역에 속하는 변인들이 국어 학업성취도를 예측하는 변인으로 나타났다. 자녀에게 기대하는 학력이 높을수록, 외국어고/국제고, 과학고/영재고 진학계획이 있는 경우 학생의 학업성취도가 높은 경향을 보였다.

학교 응답에서는 학생이 속한 학교의 '교육활동' 중 국어 선다형 비율이 높을수록 학생의 국어 학업성취도가 높았고, '학교배경' 과 관련한 변인으로는 국공립 학교에 비해 사립학교 학생들의 학업성취도가 높은 모습을 확인할 수 있었다.

<표 Ⅳ-2> 중2 국어 학업성취도 주요 설명변인(상위 30개)

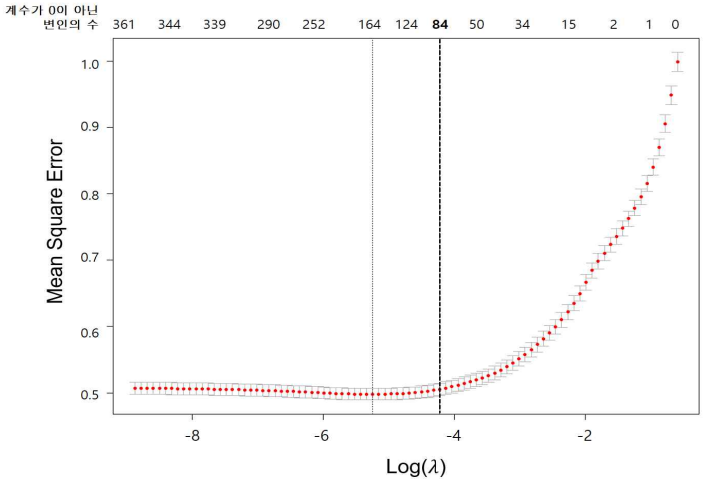
순위	설문 대상	영 역	변인명(문항내용)	척도	계수
1	학생	수업참여	수업이해도(국어)	5점[1]	0.176
2	학생	개인특성	성별(남자)	2점[2]	-0.086
3	학생	수업참여	수업이해도(수학)	5점[1]	0.067
4	학생	수업참여	수업이해도(영어)	5점[1]	0.065
5	학생	비인지적 특성	부정적 다문화수용성(우리나라에 살기 위해서 온 외국인들은 자신들의 전통이나 생활습관을 버려야 한다)	5점[2]	-0.059
6	학생	수업참여	수업태도(수업시간에 휴대전화를 본다)	5점[4]	-0.051
7	학교	교육활동	SCH선다형비율(국어)	%	0.049
8	학생	여가 및 방과후 시간	독서활동(한번 읽기 시작한 책은 끝까지 읽는다)	5점[2]	0.048
9	학생	여가 및 방과후 시간	문화활동(대중 가수 콘서트에 간다)	5점[5]	-0.047
10	학생	비인지적 특성	직업관(어떤 직업을 가지는가가 미래의 나의 인생에 중요한 영향을 미친다)	5점[2]	0.047
11	학생	비인지적 특성	정신건강(죽고 싶은 생각이 든다)	5점[2]	-0.046
12	학생	비인지적특성	직업관(돈을 많이 벌 수 있다면 어떤 직업이든 상관없다)	5점[2]	-0.045
13	학생	수업참여	수업분위기(우리반 아이들은 수업시간에 딴짓을 한다)	5점[4]	-0.044
14	학생	여가 및 방과후 시간	독서활동(독서는 내가 가장 좋아하는 취미 중 하나이다)	5점[2]	0.040
15	학생	여가 및 방과후 시간	독서활동(서점이나 도서관에 가는 것을 좋아한다)	5점[2]	0.040
16	학생	비인지적 특성	자아개념(단체생활을 잘한다)	5점[2]	-0.040
17	학교	학교배경	SCH사립	2점[3]	0.039
18	학부모	교육기대 및 계획	PA기대학력(자녀가 어느 수준까지 공부하기를 원하십니까)	5점[6]	0.036

순위	설문 대상	영 역	변인명(문항내용)	척도	계수
19	학생	비인지적 특성	진로성숙도(장래 희망을 이루기 위해 지금 무엇을 해야 할지 구체적으로 생각해 본다)	5점[2]	0.036
20	학생	학교생활과 적응	진로프로그램 도움정도(현장체험)	6점[1]	-0.035
21	학생	비인지적 특성	직업관(직업 자체가 내 인생에서 의미 있는 일은 아니다)	5점[2]	-0.031
22	학부모	교육기대 및 계획	PA진학계획(외고/국제고)	2점[1]	0.030
23	학부모	교육기대 및 계획	PA진학계획(과고/영재고)	2점[1]	0.030
24	학생	비인지적 특성	자아개념(운동신경이 발달되어 있다)	5점[2]	-0.030
25	학생	비인지적 특성	공동체의식(어른이 되면 선거나 투표에 참여할 것이다)	5점[2]	0.030
26	학생	비인지적특성	자기관리(오늘 해야 할 일을 미루지 않는다)	5점[2]	-0.029
27	학생	학교생활과 적응	교우관계(내 친구들에게 내 고민을 털어 놓는다)	5점[2]	-0.029
28	학생	여가 및 방과후 시간	미디어사용(문자, 채팅, 메신저, 이메일, 통화)	6점[2]	-0.029
29	학생	여가 및 방과후 시간	여가활동(집안일을 돕는다)	5점[4]	-0.029
30	학생	여가 및 방과후 시간	공부시간	7점[1]	0.028

주. 각 척도의 세부 내용은 다음과 같다.
2점[1] : 0: 없음, 1: 있음
2점[2] : 0: 여자, 1: 남자
2점[3] : 0: 아니오, 1: 예
5점[1] : 1: 20% 이하, 2: 21-40%, 3: 41-60%, 4: 61-80%, 5: 81% 이상
5점[2] : 1: 전혀 그렇지 않다, 2: 그렇지 않다, 3: 보통이다, 4: 그렇다, 5: 매우 그렇다
5점[4] : 1: 전혀 하지 않는다, 2: 별로 하지 않는다, 3: 보통이다, 4: 가끔 하는 편이다, 5: 자주 하는 편이다
5점[5] : 1; 연1회, 2: 학기에 1-2회, 3: 분기에 1-2회, 4: 월 1-2회, 5: 주 1-2회
5점[6] : 9: 고등학교, 11: 전문대학, 13: 4년제 대학, 15: 석사, 17: 박사
6점[1] : 0: 받은 적이 없다, 1: 전혀 도움이 되지 않는다, 2: 도움이 되지 않는 편이다, 3: 보통이다, 4: 도움이 되는 편이다, 5: 매우 도움이 된다
6점[2] : 0: 전혀 하지 않음, 1: 30분 미만, 2: 30분-1시간 미만, 3: 1-2시간 미만, 4: 2-3시간 미만, 5: 3시간 이상
7점[1] : 0: 전혀 없음, 1: 1시간, 2: 2시간, 3: 3시간, 4: 4시간, 5: 5시간, 6: 6시간 이상

2) 영어

중2 시기의 영어 학업성취도 주요 설명변인을 선정하기 위하여, 적절한 제약모수 λ 값을 찾기 위한 20-fold 교차타당화를 실시하였다. 교차타당화를 통해 MSE와 표준오차를 살펴본 결과는 [그림 Ⅳ-3]과 같다. 1 표준오차 법칙에 따라(Hastie et al., 2009, p. 244), MSE가 가장 적을 때의 값에서 1 표준오차 범위 내에 있는 최대의 λ 값을 확인하였으며, 0.015로 나타났다. 이를 기준으로 LASSO 모형을 적합하였고, 계수가 0이 아닌 변인은 84개였다.



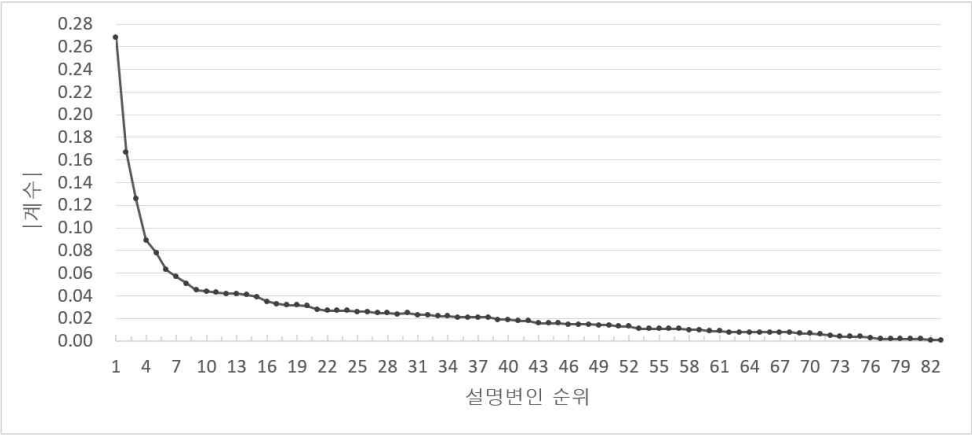
[그림 Ⅳ-3] 영어 중2 학업성취도 제약모수의 20-fold 교차타당화 결과

영어 학업성취도를 설명하는 84개 변인을 계수 절대값을 기준으로 기여도가 높은 순으로 설명변인을 정리하여 나타낸 결과는 <표 Ⅳ-3>과 같으며, 각 변인의 세부 문항은 <표 Ⅲ-4>와 부록에 제시하였다.

<표 Ⅳ-3> 중2 영어 학업성취도 주요 설명변인 선정 결과

순위	변인명	계수	순위	변인명	계수	순위	변인명	계수
1	수업이해도(영)	0.218	31	학교만족도(기초능력향상)	0.025	61	PA사교육참여이유	0.010
2	선행학습정도(영)	0.160	32	미디어사용(연락)	-0.025	62	PA계획(국/공립대)	-0.010
3	성별(남)	-0.124	33	SCH자율동아리(명)	0.025	63	교사관계(이해)	-0.009
4	수업이해도(수)	0.113	34	여가활동(집안일)	-0.024	64	수업분위기(무질서)	-0.009
5	수업태도(휴대전화)	-0.097	35	자아개념(단체생활)	-0.023	65	학교내폭력(욕설)	-0.008
6	월평균 가정소득	0.080	36	문화활동(박물관)	0.021	66	진로상담실 운영	0.008
7	SCH사립	0.067	37	상호작용(책읽고대화)	0.021	67	교육적지원(학부모모임)	-0.007
8	SCH특별시	0.059	38	PA계획(국/공립대)	0.020	68	교육적지원(진학관심)	0.007
9	부정적 다문화수용성	-0.051	39	공동체의식(빌려줌)	0.020	69	정신건강(마음편함)	0.006
10	PA기대학력	0.049	40	소외(스트레스)	0.018	70	진학계획(외고/국제고)	0.006
11	여가활동(학원,과외)	0.048	41	SCH학부모지원(보조)	0.018	71	진학계획(자사고)	0.006
12	문화활동(콘서트)	-0.043	42	SCH여학교	0.018	72	수업분위기(집중)	0.005
13	수업분위기(딴짓)	-0.043	43	진로프로도움(현장체험)	-0.017	73	독서활동(이야기)	0.005
14	진로프로도움(고·대방문)	-0.041	44	수업방식(질문수용)	0.017	74	SCH구성원학교운영(공유)	0.004
15	학폭경험(맞음)	-0.039	45	직업관(수단)	-0.017	75	교우관계(고민나눔)	-0.004
16	공동체의식(투표참여)	0.035	46	자아개념(외모)	-0.017	76	공부시간	0.003
17	PA계획(외고/국제고)	0.034	47	여가활동(TV)	-0.016	77	수업분위기(과제)	0.002
18	SCH수행비율(영)	0.034	48	PA계획(자사고)	0.016	78	공동체의식(조별학습)	-0.002
19	학교크기(학생 수)	0.032	49	소외(관심X)	0.016	79	부모상호작용(학교활동)	0.002
20	PA계획(특성화고)	-0.032	50	PA계획(과고/영재고)	0.015	80	규칙준수(차례)	0.002
21	선행학습정도(수)	0.031	51	수업방식(맞춤형과제)	0.014	81	자아개념(운동신경)	-0.001
22	진학계획(예고/체고)	-0.029	52	직업관(의미X)	-0.013	82	수업분위기(지시불이행)	-0.001
23	수업이해도(국)	0.029	53	자기관리(미루지않음)	0.013	83	PA진학계획(PA진학)	-0.001
24	수업태도(잠)	-0.028	54	다문화수용성(같은대우)	0.012	84	문화활동(문화활)	0.001
25	자기관리(숙제<노는것)	-0.028	55	SCH교복우사업	0.012			
26	독서활동(책끝까지)	0.028	56	SCH중소도시	-0.012			
27	직업관(인생에 중요)	0.027	57	교사특성(과제강조)	0.012			
28	진로성숙도(구체적생각)	-0.027	58	SCH혁신학교	0.011			
29	직업관(수입만크면됨)	-0.026	59	정신건강(죽고싶다)	-0.010			
30	독서활동(서점,도서관)	0.026	60	학폭경험(SNS)	-0.010			

영어 학업성취도 주요 설명변인 순위에 따른 LASSO 계수 절대값은 [그림 Ⅳ-4]와 같다. 상위 1~3번째 변인은 다른 변인에 비해 학업성취도에 기여하는 정도가 매우 높았고, 상위 11번째 변인을 전후로 하여 점진적으로 감소하는 모습이 나타났다. 상위 59번째 변인 이후로는 계수 절대값이 0.01 이하로 나타났으며, 이는 각 변인이 1 표준편차만큼 변화할 때, 영어 학업성취도가 0.01 표준편차 변화한다는 것을 의미한다. 이에 따라 계수가 거의 0에 가까운 후순위 설명변인들은 기초미달 학생을 예측하는 데 큰 영향을 미치지 못할 것이라 예상할 수 있다.



[그림 Ⅳ-4] 설명변인 순위에 따른 LASSO 계수 (영어)

다음으로 영어 학업성취도를 설명하는 84개 변인 중 상위 30개 변인을 중심으로 학업성취도와의 관계를 상세하게 살펴보면 <표 Ⅲ-4>와 같다. 국어 과목과 유사하게 성별과 수업참여 영역의 변인들이 주요 설명변인으로 나타난 반면, 국어 과목과는 달리 여가 및 방과후 시간 영역 중 학습 활동과 관련된 변인이 주요 설명변인으로 나타났다. 또한 가정의 경제활동 및 학교 배경 변인의 개수가 상대적으로 많았다. 비인지적 특성 변인에서는 진로와 관련된 변인이 상위 30개 변인 중 상대적으로 하위 순위의 변인으로 나타났다.

각 변인의 문항 내용과 영어 학업성취도와의 관계를 살펴보면 다음과

같다. 우선, 학생 응답 중 ‘수업참여’ 영역에서는 영어, 수학, 국어 수업이해도가 높을수록 영어 학업성취도가 높았으며, 특히 영어 수업이해도가 학업성취도에 미치는 정도(0.218)가 수학(0.113)과 국어(0.029)에 비해 상당히 크게 나타났다. 수업태도 및 분위기 관련 변인에서는 수업 중 휴대전화를 보거나, 수업시간에 졸거나 잘수록, 수업에서 딴짓하는 학생이 많다고 인식할수록 영어 학업성취도는 낮은 경향을 보였다.

‘비인지적 특성’ 영역에서는 국어와 유사한 변인들이 나타났다. 부정적 다문화 수용성이 높을수록, 숙제보다 TV를 보거나 노는 것을 먼저 하는 경우가 많을수록 영어 학업성취도 낮았던 반면, 어른이 되면 선거나 투표에 참여하겠다고 생각하는 것은 학업성취도와 정적인 관계를 보였다. 진로 관련 변인들은 영어 학업성취도와도 유사한 관계를 보이고 있었으나, 국어에 비해 상대적으로 하위 순위에서 확인되었다. 미래 인생에 있어 직업 선택이 중요하다고 생각하고, 장래 희망을 이루기 위해 해야 할 일을 구체적으로 생각해볼수록 영어 학업성취도가 높았고, 돈을 많이 벌 수 있다면 어떤 직업이든 상관없다고 생각할수록 학업성취도는 낮은 경향이 있었다.

‘여가 및 방과후 시간’ 영역에서는 영어, 수학 선행학습을 많이 하는 경우, 방과 후에 학원 강의나 과외 수업을 오랜 시간 받는 경우에 영어 학업성취도가 높게 나타났으며, 대중가수 콘서트에 자주 가는 것은 영어 학업성취도와 부적인 관계를 보였다. 국어에 비해 하위 순위이기는 하나, 독서활동과 관련된 변인도 주요 예측 변인으로 나타났으며, 한번 읽기 시작한 책을 끝까지 읽을수록, 서점이나 도서관에 가는 것을 좋아할수록 영어 학업성취도가 높았다.

이 외에도 남학생이 여학생이 비해 영어 학업성취도도 낮은 모습이 나타났으며, 예술고 또는 체육고에 진학하고자 하는 계획이 있는 경우에도 그렇지 않은 학생들에 비해 학업성취도가 낮았다.

다음으로 학부모 응답에서는 가정의 ‘경제수준’과 ‘교육기대 및 계획’ 영역과 관련된 변인들이 영어 학업성취도의 주요 설명변인으로 나타났다. 가정의 월평균 소득이 높을수록 영어 학업성취도가 높았으며,

학부모가 자녀에게 기대하는 학력이 높을수록, 외국어고/국제고 진학계획이 있고, 특성화고 진학계획이 없는 경우에 자녀의 학업성취도가 높은 모습이 나타났다.

학교 응답에서는 학생이 속한 학교의 영어 수행평가 비율이 높을수록 학생의 영어 학업성취도가 높은 경향을 보였으며, 국공립 학교에 비해 사립학교인 경우에, 학교가 특별시에 소재하는 경우에 학생들의 학업성취도가 높게 나타났다. 또한 학교의 학생 수가 많은 학교에 다닐수록 영어 학업성취도가 높은 경향을 보였다.

<표 Ⅳ-4> 중2 영어 학업성취도 주요 설명변인(상위 30개)

순위	설문 대상	영 역	변인명(문항내용)	척도	계수
1	학생	수업참여	수업이해도(영어)	5점[1]	0.218
2	학생	여가 및 방과후 시간	선행학습정도(어느 학년까지의 내용을 선행학습한 상태입니까?)(영어)	4점[1]	0.160
3	학생	개인특성	성별(남자)	2점[2]	-0.124
4	학생	수업참여	수업이해도(수학)	5점[1]	0.113
5	학생	수업참여	수업태도(수업시간에 휴대전화를 본다)	5점[4]	-0.097
6	학부모	경제수준	월평균 가정소득	만원	0.080
7	학교	학교배경	SCH사립	2점[3]	0.067
8	학교	학교배경	SCH특별시	2점[3]	0.059
9	학생	비인지적 특성	부정적 다문화수용성(우리나라에 살기 위해서 온 외국인들은 자신들의 전통이나 생활습관을 버려야 한다)	5점[2]	-0.051
10	학부모	교육기대 및 계획	PA기대학력(자녀가 어느 수준까지 공부하기를 원하십니까)	5점[6]	0.049
11	학생	여가 및 방과후 시간	여가활동(학원 강의나 과외 수업을 듣는다)	5점[3]	0.048

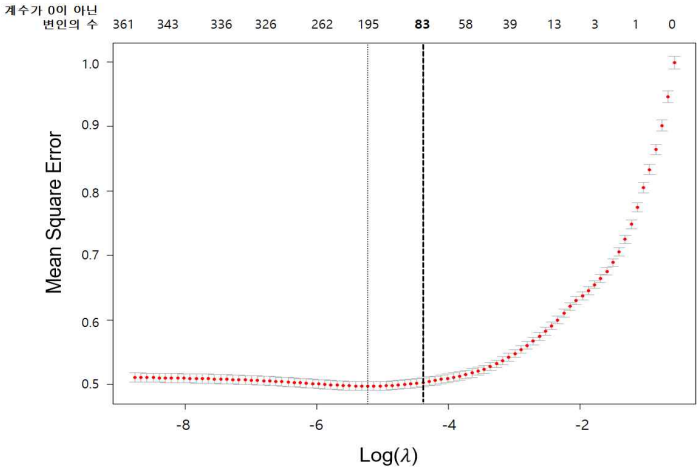
순위	설문 대상	영 역	변인명(문항내용)	척도	계수
12	학생	여가 및 방과후 시간	문화활동(대중 가수 콘서트에 간다)	5점[5]	-0.043
13	학생	수업참여	수업분위기(우리반 아이들은 수업시간에 딴짓을 한다)	5점[4]	-0.043
14	학생	학교생활과 적응	진로프로그램 도움정도(고등학교 또는 대학교 방문 및 학과체험)	6점[1]	-0.041
15	학생	학교생활과 적응	학폭경험(친구나 선후배에게 심하게 맞았다)	5점[7]	-0.039
16	학생	비인지적 특성	공동체의식(어른이 되면 선거나 투표에 참여할 것이다)	5점[2]	0.035
17	학부모	교육기대 및 계획	PA진학계획여부(외고/국제고)	2점[1]	0.034
18	학교	교육활동	SCH수행평가비율(영어)	%	0.034
19	학교	학교통계	학교크기(학생 수)	명	0.032
20	학부모	교육기대 및 계획	PA진학계획여부(특성화고)	2점[1]	-0.032
21	학생	여가 및 방과후 시간	선행학습정도(어느 학년까지의 내용을 선행학습한 상태입니까?)(수학)	4점[1]	0.031
22	학생	진로 및 포부	진학계획(예고/체고)	2점[1]	-0.029
23	학생	수업참여	수업이해도(국어)	5점[1]	0.029
24	학생	수업참여	수업태도(수업시간에 졸거나 잔다)	5점[4]	-0.028
25	학생	비인지적 특성	자기관리(숙제보다 TV를 보거나 노는 것을 먼저 하는 경우가 있다)	5점[2]	-0.028
26	학생	여가 및 방과후 시간	독서활동(한번 읽기 시작한 책은 끝까지 읽는다)	5점[2]	0.028
27	학생	비인지적 특성	직업관(어떤 직업을 가지는가가 미래의 나의 인생에 중요한 영향을 미친다)	5점[2]	0.027
28	학생	비인지적 특성	진로성숙도(장래 희망을 이루기 위해 지금 무엇을 해야 할지 구체적으로 생각해 본다)	5점[2]	-0.027
29	학생	비인지적 특성	직업관(돈을 많이 벌 수 있다면 어떤 직업이든 상관없다)	5점[2]	-0.026

순위	설문 대상	영 역	변인명(문항내용)	척도	계수
30	학생	여가 및 방과후 시간	독서활동(서점이나 도서관에 가는 것을 좋아한다)	5점[2]	0.026

주. 각 척도의 세부 내용은 다음과 같다.
2점[1] : 0: 없음, 1: 있음
2점[2] : 0: 여자, 1: 남자
2점[3] : 0: 아니오, 1: 예
4점[1] : 1: 중학교 3학년, 2: 고등학교 1학년, 3: 고등학교 2학년, 4:고등학교 3학년
5점[1] : 1: 20% 이하, 2: 21-40%, 3: 41-60%, 4: 61-80%, 5: 81% 이상
5점[2] : 1: 전혀 그렇지 않다, 2: 그렇지 않다, 3: 보통이다, 4: 그렇다, 5: 매우 그렇다
5점[4] : 1: 전혀 하지 않는다, 2: 별로 하지 않는다, 3: 보통이다, 4: 가끔 하는 편이다, 5: 자주 하는 편이다
5점[6] : 9: 고등학교, 11: 전문대학, 13: 4년제 대학, 15: 석사, 17: 박사
5점[7] : 1: 전혀 없다, 2: 거의 없다, 3: 가끔 있다, 4: 자주 있다, 5: 매우 많다
6점[1] : 0: 받은 적이 없다, 1: 전혀 도움이 되지 않는다, 2: 도움이 되지 않는 편이다, 3: 보통이다, 4: 도움이 되는 편이다, 5: 매우 도움이 된다

3) 수학

중2 수학 학업성취도 주요 설명변인을 선정하기 위하며, 먼저 제약모수 λ 값을 확인하기 위한 20-fold 교차타당화를 실시한 결과는 [그림 Ⅲ-5]와 같다. 1 표준오차 법칙에 따라(Hastie et al., 2009, p. 244), MSE가 가장 적을 때의 값에서 1 표준오차 범위 내에 있는 최대의 λ 값을 살펴보았다. 이때의 λ 값 0.012를 제약 모수로 하여 LASSO 모형을 적합하였을 때, 계수가 0이 아닌 변인은 83개임을 확인하였다.



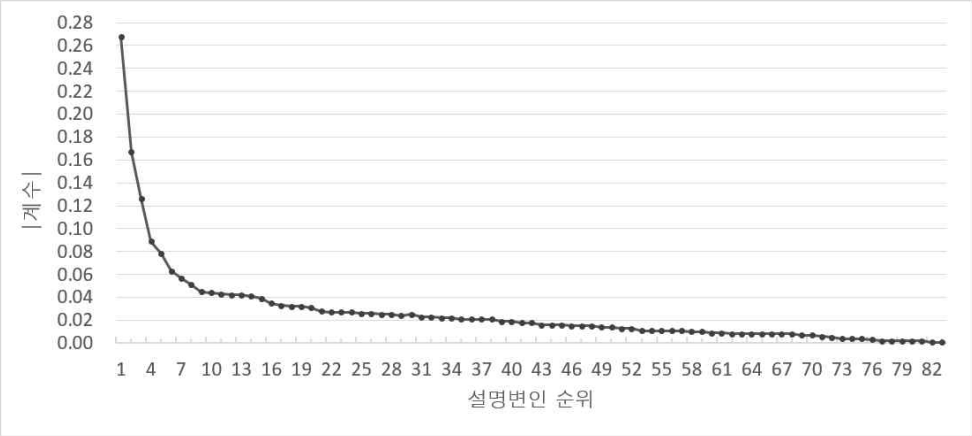
[그림 Ⅳ-5] 수학 중2 학업성취도 제약모수의 20-fold 교차타당화 결과

분석에 활용하는 모든 변인은 값을 표준화하였기 때문에 LASSO 회귀 크기에 비추어 상대적인 기여도를 확인할 수 있다. 이에 따라 수학 학업성취도를 설명하는 83개 변인을 계수 절대값 크기에 따라 정리하여 <표 Ⅳ-5>에 제시하였으며, 각 변인의 세부 문항은 <표 Ⅲ-6>과 부록에 제시하였다.

<표 Ⅳ-5> 중2 수학 학업성취도 주요 설명변인 선정 결과

순위	변인명	계수	순위	변인명	계수	순위	변인명	계수
1	수업이해도(수)	0.268	31	규칙준수(학급규칙)	0.023	61	PA계획(자공고)	0.009
2	선행학습정도(수)	0.167	32	학교만족도(기초능력향상)	0.023	62	부정적다문화수용성(범죄)	-0.008
3	수업이해도(영)	0.126	33	공동체의식(소외된친구)	0.022	63	학교내폭력(욕설)	-0.008
4	PA계획(자사고)	0.089	34	여가활동(TV)	-0.022	64	자기관리(숙제<노는것)	-0.008
5	여가활동(학원,과외)	0.078	35	SCH수행비율(수)	0.021	65	다문화수용성(친구)	0.008
6	자아개념(단체생활)	-0.063	36	직업관(의미X)	-0.021	66	진로상담실 운영	0.008
7	여가활동(집안일)	-0.057	37	진로프로도움(현장체험)	-0.021	67	PA계획(예고/체고)	-0.008
8	PA기대학력	0.051	38	상호작용(책읽고대화)	0.021	68	수업방식(질문수용)	0.008
9	SCH사립	0.045	39	PA계획(국/공립대)	0.019	69	PA교육적지원(숙제)	-0.007
10	진로성숙도(구체적생각)	0.044	40	학폭경험(맞음)	-0.019	70	학교크기(학생수)	0.007
11	PA계획(과고/영재고)	0.043	41	PA교육적지원(관심)	0.018	71	교사특성(과제강조)	0.006
12	직업관(수입만크면됨)	-0.042	42	정신건강(죽고싶다)	-0.018	72	학폭경험(SNS)	-0.005
13	진학계획(예고/체고)	-0.042	43	직업관(인생에 중요)	0.016	73	진로프로도움(고·대방문)	-0.004
14	문화활동(콘서트)	-0.041	44	SCH선다형비율(국)	0.016	74	학폭경험(놀림)	-0.004
15	월평균 가정소득	0.039	45	PA학교만족도(진로)	-0.016	75	문화활동(박물관)	0.004
16	PA사교육참여이유	0.035	46	SCH자율동아리(명)	0.015	76	공동체의식(조별학습)	-0.003
17	PA계획(특성화고)	-0.033	47	공부시간	0.015	77	교육프로도움(수)	-0.002
18	부정적 다문화수용성	-0.032	48	창의성(새로운생각)	-0.015	78	소외(스트레스)	0.002
19	진학계획(자사고)	0.032	49	미디어사용(오락)	-0.014	79	PA역량기반교육	-0.002
20	수업태도(휴대전화)	-0.031	50	학교크기(학생 수)	0.014	80	SCH구성원학교운영(참관)	0.002
21	SCH수행비율(수)	0.028	51	교우관계(고민나눔)	-0.013	81	자아개념(수업참여)	0.002
22	자기관리(미루지않음)	0.027	52	다문화수용성(다양성)	0.013	82	학교의사결정(의견반영)	0.001
23	수업분위기(딴짓)	-0.027	53	PA학교만족도(수준)	0.011	83	교사특성(상위권강조)	-0.001
24	수업태도(잠)	-0.027	54	교사특성(열심히공부)	0.011			
25	공동체의식(소외된친구)	0.026	55	SCH중소도시	-0.011			
26	공동체의식(투표참여)	0.026	56	진학계획(해외유학)	-0.011			
27	SCH혁신학교	0.025	57	독서활동(서점,도서관)	0.011			
28	미디어사용(연락)	-0.025	58	진학계획(외고/국제고)	0.010			
29	독서활동(책끝까지)	0.024	59	PA계획(외고/국제고)	0.010			
30	SCH특별시	0.025	60	다문화수용성(같은대우)	0.009			

수학 학업성취도 주요 설명변인 순위에 따른 LASSO 계수 절대값은 [그림 Ⅳ-6]와 같다. 수학에서도 상위 1~3번째 변인의 계수 절대값이 다른 변인에 비해 상당히 높은 모습이 나타났으며, 상대적으로 상위 9번째 변인 이후 점진적으로 감소하는 모습이 나타났다. 83개 설명변인 중에는 계수가 거의 0에 가까운 변인도 다수 확인할 수 있었다.



[그림 Ⅳ-6] 설명변인 순위에 따른 LASSO 계수 (수학)

수학 학업성취도를 설명하는 83개 변인 중 상위 30개 변인을 상세하게 살펴보면 <표 Ⅳ-6>와 같다. 수학에서는 부모의 교육기대 및 계획 관련된 변인 및 학생의 진로 및 포부 등 진학 관련 변인이 주요 상위 설명변인으로 나타났다. 또한 학원 강의 참여 및 선행학습과 관련된 변인들이 상대적으로 많이 나타난 것을 확인할 수 있었으며, 국어와 영어 과목과는 달리 학교생활과 적응 영역의 변인들은 상위 30개 주요 설명변인에서 확인하기 어려웠다.

상위 30개 설명변인과 수학 학업성취도와의 관계를 살펴보면 다음과 같다. 학생 응답 중 ‘수업참여’ 영역에서는 수학, 영어 수업이해도가 높을수록 수학 학업성취도가 높았으며, 수업시간에 휴대전화를 보거나, 졸거나 잘 때가 많으며, 수업에서 딴짓하는 학생이 많다고 인식할수록 수학 학업성취도가 낮은 모습이 나타나, 영어와 유사한 경향성을 확인할

수 있었다. 하지만 수업이해도 중 국어 수업에서의 수업이해도는 상위 30개 변인에서 나타나지 않았으며, 수업태도 및 수업분위기에 해당하는 변인들은 영어에 비해 상대적으로 하위 순위에서 확인할 수 있었다.

‘비인지적 특성’ 영역에서는 자아개념, 진로성숙도, 직업관, 공동체 의식 관련 변인이 국어, 영어 주요 설명변인과 유사하게 나타났으며, 학업성취도와의 관계 역시 비슷한 경향을 보였다. 장래 희망을 이루기 위해 해야 할 일을 구체적으로 생각하고, 오늘 할 일을 내일로 미루지 않으며, 소외된 친구가 있으면 먼저 따뜻하게 대해주고, 어른이 되면 선거나 투표에 참여할 의향이 높을수록 수학 학업성취도가 높게 나타났다. 반면 단체생활을 잘한다고 인식하며, 돈을 많이 벌 수 있다면 어떤 직업이든 상관없다고 생각하고, 부정적 다문화수용성이 높을수록 수학 학업성취도는 낮게 나타났다.

‘여가 및 방과후 시간’ 영역에서는 수학 선행학습을 많이 하고, 방과 후에 학원 강의가 과외 수업에 참여하는 시간이 많을수록 수학 학업성취도가 높았던 반면, 방과 후에 집안일을 돕는 시간이 많고, 대중 가수 콘서트에 가는 횟수가 많은 경우에는 학업성취도가 낮은 경향을 보였다. 또한 한번 읽기 시작한 책은 끝까지 읽는 성향이 수학 학업성취도와 정적인 관계가 나타났다.

수학 과목의 주요 예측 변인에는 학부모의 ‘교육기대 및 계획’ 영역의 변인의 수가 국어, 영어에 비해 많았으며, 학생 응답의 ‘진로 및 포부’ 영역에서 진학계획 변인들 또한 상위 30개 주요 설명변인에서 확인할 수 있었다. 구체적으로, 학부모가 자녀에게 기대하는 학력이 높을수록, 자사고, 과학고/영재고 진학계획이 있고, 특성화고 진학계획이 없을수록 자녀의 수학 학업성취도는 높았으며, 학생이 예술고/ 체육고에 계획이 없는 경우, 자사고 진학계획이 있는 경우에 수학 학업성취도가 높게 나타났다.

또한 ‘방과후 교육활동’에서는 상급학교 준비를 위에 사교육의 참여하는 경우, 가정의 월평균 가정소득이 수학 학업성취도 주요 설명변인으로 나타났으며, 가정소득과 수학 학업성취도는 정적인 관계가 나타났다.

학교 응답에서는 학생이 속한 학교가 사립학교일 때, 학교가 특별시에 소재하는 경우에 수학 학업성취도가 높게 나타났으며, 학교의 수학 수행평가 비율이 높을수록 수학 학업성취도가 높게 나타났다. 이는 영어 과목의 학교 응답과 상당히 유사하게 나타난 부분이다. 이 외에도 수학 과목에서는 혁신학교 여부가 학생의 수학 학업성취도에 정적으로 영향을 미치는 모습도 확인되었다.

<표 Ⅳ-6> 중2 수학 학업성취도 주요 설명변인(상위 30개)

순위	설문 대상	영 역	변인명(문항내용)	척도	계수
1	학생	수업참여	수업이해도(수학)	5점[1]	0.268
2	학생	여가 및 방과후 시간	선행학습정도(어느 학년까지의 내용을 선행학습한 상태입니까?)(수학)	4점[1]	0.167
3	학생	수업참여	수업이해도(영어)	5점[1]	0.126
4	학부모	교육기대 및 계획	PA진학계획(자사고)	2점[1]	0.089
5	학생	여가 및 방과후 시간	여가활동(학원 강의나 과외 수업을 듣는다)	5점[3]	0.078
6	학생	비인지적 특성	자아개념(단체생활을 잘한다)	5점[2]	-0.063
7	학생	여가 및 방과후 시간	여가활동(집안일을 돕는다)	5점[4]	-0.057
8	학부모	교육기대 및 계획	PA기대학력(자녀가 어느 수준까지 공부하기를 원하십니까)	5점[6]	0.051
9	학교	학교배경	SCH사립	2점[3]	0.045
10	학생	비인지적 특성	진로성숙도(장래 희망을 이루기 위해 지금 무엇을 해야 할지 구체적으로 생각해 본다)	5점[2]	0.044
11	학부모	교육기대 및 계획	PA진학계획(과고/영재고)	2점[1]	0.043
12	학생	비인지적 특성	직업관(돈을 많이 벌 수 있다면 어떤 직업이든 상관없다)	5점[2]	-0.042
13	학생	진로 및 포부	진학계획(예고/체고)	2점[1]	-0.042

순위	설문 대상	영 역	변인명(문항내용)	척도	계수
14	학생	여가 및 방과후 시간	문화활동(대중 가수 콘서트에 간다)	5점[5]	-0.041
15	학부모	경제수준	월평균 가정소득	만원	0.039
16	학부모	방과후 교육활동	사교육참여 이유(상급 학교 준비를 위해)	2점[1]	0.035
17	학부모	교육기대 및 계획	PA진학계획여부(특성화고)	2점[1]	-0.033
18	학생	비인지적 특성	부정적 다문화수용성(우리나라에 살기 위해서 온 외국인들은 자신들의 전통이나 생활습관을 버려야 한다)	5점[2]	-0.032
19	학생	진로 및 포부	진학계획(자사고)	2점[1]	0.032
20	학생	수업참여	수업태도(수업시간에 휴대전화를 본다)	5점[4]	-0.031
21	학교	교육활동	SCH수행평가비율(수학)	%	0.028
22	학생	비인지적 특성	자기관리(오늘 해야 할 일을 미루지 않는다)	5점[2]	0.027
23	학생	수업참여	수업분위기(우리반 아이들은 수업시간에 딴짓을 한다)	5점[4]	-0.027
24	학생	수업참여	수업태도(수업시간에 졸거나 잔다)	5점[4]	-0.027
25	학생	비인지적 특성	공동체의식(소외된 친구가 있으면 내가 먼저 따뜻하게 대해 준다)	5점[2]	0.026
26	학생	비인지적 특성	공동체의식(어른이 되면 선거나 투표에 참여할 것이다)	5점[2]	0.026
27	학교	학교정책 및 운영	SCH혁신학교	2점[3]	0.025
28	학생	여가 및 방과후 시간	미디어사용(문자, 채팅, 메신저, 이메일, 통화)	6점[2]	-0.025
29	학생	여가 및 방과후 시간	독서활동(한번 읽기 시작한 책은 끝까지 읽는다)	5점[2]	0.024
30	학교	학교배경	SCH특별시	2점[3]	0.025

순위	설문 대상	영 역	변인명(문항내용)	척도	계수

주. 각 척도의 세부 내용은 다음과 같다.
2점[1] : 0: 없음, 1: 있음
2점[3] : 0: 아니오, 1: 예
4점[1] : 1: 중학교 3학년, 2: 고등학교 1학년, 3: 고등학교 2학년, 4:고등학교 3학년
5점[1] : 1: 20% 이하, 2: 21-40%, 3: 41-60%, 4: 61-80%, 5: 81% 이상
5점[2] : 1: 전혀 그렇지 않다, 2: 그렇지 않다, 3: 보통이다, 4: 그렇다, 5: 매우 그렇다
5점[3] : 0: 전혀 하지 않음, 1: 1시간 미만, 2: 1시간 이상 - 2시간 미만, 3: 2시간 이상-3시간 미만, 4: 3시간 이상
5점[4] : 1: 전혀 하지 않는다, 2: 별로 하지 않는다, 3: 보통이다, 4: 가끔 하는 편이다, 5: 자주 하는 편이다
5점[5] : 1; 연1회, 2: 학기에 1-2회, 3: 분기에 1-2회, 4: 월 1-2회, 5: 주 1-2회
6점[2] : 0: 전혀 하지 않음, 1: 30분 미만, 2: 30분-1시간 미만, 3: 1-2시간 미만, 4: 2-3시간 미만, 5: 3시간 이상

나. 주요 설명변인 선정에 따른 예측모형 설정

과목별 중2 학업성취도를 설명하는 주요 설명변인을 살펴본 결과, 학생, 학부모, 학교와 관련된 361개 변인 중 국어 91개, 영어 84개, 수학 83개 설명변인이 각 과목 학업성취도를 설명하고 있는 것으로 나타났다. LASSO 분석을 실시하기 전 모든 변인을 표준화하였으므로 LASSO 계수 절대값 크기에 비추어 각 변인이 변화함에 따라 학업성취도가 변화하는 정도를 확인할 수 있다. 이에 따라 계수 절대값 크기가 큰 변인부터 작은 변인을 차례로 나열하고 그 크기를 살펴보았다.

LASSO 분석에 따른 주요 설명변인과 변인별 계수 절대값을 살펴보았을 때, 모든 과목에서 상위 1~3개 주요 설명변인의 계수 절대값이 다른 변인에 비해 매우 크게 나타났으며, 그 이후에는 계수가 점진적으로 작아지는 경향을 보였다. 또한 전체 변인 중 상대적으로 계수가 0에 가까운 변인들도 다수 포함되어 있는 모습을 확인할 수 있었다. 이는 모든 변인이 학업성취도를 설명하는 데 기여하고 있기는 하지만, 후순위 설명변인으로 갈수록 해당 변인의 값 변화에 따른 학업성취도 변화가 상대적

으로 적다는 것을 의미한다고 볼 수 있는 부분이다. 이에 따라 학업성취도를 설명하는 변인 중에서 학업성취도에 따른 학생 간 특징을 구분하기 어려울 수 있는 변인들이 있을 수 있음을 짐작할 수 있었다.

따라서 심층인공신경망 모형이 예측변인 개수 활용에 제약이 없는 모형이긴 하나, 구축할 때 설명력이 낮거나 관련이 없는 변인이 오히려 패턴 파악을 어렵게 할 수 있다는 논의에 따라(García et al., 2016; Snieder et al., 2019), 이후 심층인공신경망 예측모형을 상세화하는 과정에서 주요 설명변인을 예측변인으로 활용할 때, 각 변인의 계수 절대값 크기에 따라 예측변인으로 활용하는 것을 고려하였다.

구체적으로 상위 9, 16, 25, 36, 49, 64, 81개로 예측변인의 수를 늘려가며 예측모형을 설정하고자 하였다. 예측변인의 개수를 이와 같이 설정한 이유는 이후 합성곱신경망 모형을 적용할 때, 행(사례)과 열(변인)로 표현되는 표형식의 교육패널자료를 이미지화하는 과정에서 입력층 변인의 수를 n^2개씩 지정하게 되는 점을 고려하였기 때문이다. 또한 주요 설명변인 선정 결과 국어 91개, 영어 84개, 수학 83개까지의 설명변인 순위를 확인할 수 있고 그 이후 설명변인은 모든 계수가 0이기 때문에 순위를 확인할 수 없었다. 따라서 최대 81개 변인을 활용하는 것으로 예측변인의 개수를 지정하고자 하였다.

2. 기초미달 학생 예측모형 구현

앞서 주요 설명변인을 살펴본 결과를 바탕으로, 과목별 LASSO 계수가 0이 아닌 설명변인을 예측변인으로 하여 기초미달 학생 예측모형을 구현하고 모형 성능을 확인하였다. 국어, 영어, 수학 과목별 기초미달 학생 예측모형을 상세화하고, 연구문제 2에 따라 심층인공신경망 예측모형은 예측변인의 개수, 소수집단 정보 반영방법, 심층인공신경망 종류에 따른 결과를 살펴보았다.

가. 국어

국어 기초미달 학생의 비율은 7.6%로 다른 과목에 비해 불균형 정도가 크며, 소수집단 대비 다수집단의 비율(Imbalanced Ratio, 이하 IR)이 10 이상인 경우 높은 불균형을 보이는 자료라 볼 수 있을 때(Fernandez et al., 2008), IR 12.15으로 높은 불균형을 보이고 있었다.

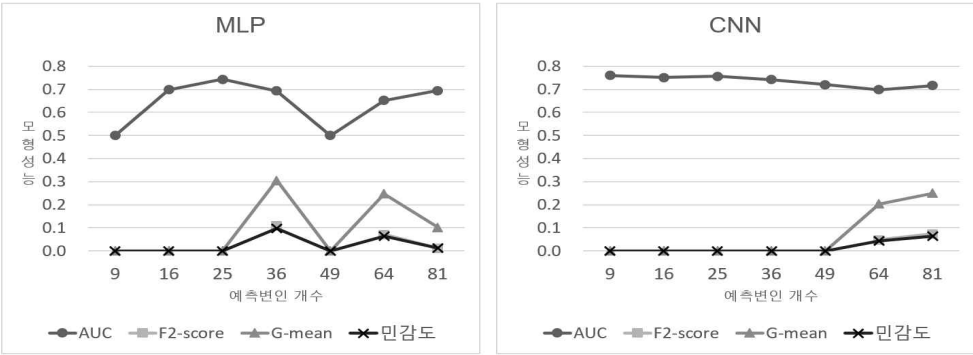
모형 훈련 시 불균형 정도를 조정해주지 않은 Original 모형의 경우, <표 Ⅳ-7>과 같이 MLP, CNN 모형 모두 정분류율은 약 0.9 이상이나 민감도 약 0.1 미만으로, 대부분의 학생을 기초이상으로 예측하는 모습이 나타났다. 이는 불균형 자료 예측모형을 구축할 때 불균형 정도를 조정해 주는 것이 필요함을 보여준다.

<표 Ⅳ-7> 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-Original)

모형	평가지표	예측변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.500	0.699	0.744	0.693	0.500	0.652	0.695
	F2-score	-	-	-	0.108	-	0.071	0.012
	G-mean	0.000	0.000	0.000	0.305	0.000	0.248	0.102
	정분류율	0.927	0.927	0.927	0.910	0.927	0.899	0.902
	민감도	0.000	0.000	0.000	0.096	0.000	0.064	0.011
	특이도	1.000	1.000	1.000	0.975	1.000	0.965	0.973
	McNemar-test	-	-	-	0.7	-	0.6	0.8
CNN	AUC	0.760	0.752	0.756	0.742	0.721	0.699	0.717
	F2-score	-	-	-	-	-	0.048	0.073
	G-mean	0.000	0.000	0.000	0.000	0.000	0.203	0.250
	정분류율	0.927	0.927	0.927	0.927	0.927	0.900	0.909
	민감도	0.000	0.000	0.000	0.000	0.000	0.043	0.064
	특이도	1.000	1.000	1.000	1.000	1.000	0.968	0.976
	McNemar-test	-	-	-	-	-	1.5	1.2

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-7] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (국어-Original)

표집방법을 적용하였을 때, 예측변인의 개수와 심층인공신경망 종류에 따른 모형 성능은 <표 Ⅳ-8>, [그림 Ⅳ-8]와 같다. 훈련자료에 기초미달 학생 정보가 기초이상 학생 정보만큼 반영되도록 조정한 경우에는 Original 모형에 비해 모형 성능이 향상되는 모습이 나타났다. <표 Ⅳ-8>에 나타난 것과 같이, ROS, B-SMOTE, ADASYN 모두 각 모형에서 모형 성능이 가장 높았던 모형(볼드체)의 G-mean 값이 0.657~0.685 이상이 되어 기초이상 학생과 기초미달 학생 예측의 불균형이 완화되었고, F2-score과 AUC 역시 Original 모형에 비해 값이 증가하였다.

예측변인의 개수에 따른 결과를 살펴보면([그림 Ⅳ-8]), 예측변인의 개수가 약 9~25개 정도일 때에는 모형 성능이 비슷한 수준이거나 오히려 상승하는 모습도 나타났다. 하지만 예측변인의 개수가 36~81개일 때에는 ROS를 적용한 CNN 모형을 제외한 모든 모형에서 예측에 활용하는 변인이 많아질수록 모형 성능이 감소하는 경향을 보였다.

심층인공신경망 모형 종류에 따른 모형 성능 결과를 살펴보면, MLP, CNN 모형 적용에 따른 예측 성능의 차이는 크지 않았으나, 예측변인 개수의 따른 모형 성능 변화에서는 차이를 보였다. 우선, <표 Ⅳ-8>에서 각 모형에서 성능이 높게 나타난 모형(볼드체)을 기준으로 비교하였을 때, ROS를 적용한 경우에는 AUC, F2-score는 CNN 모형이 더 높았고, G-mean, 민감도는 MLP 모형이 더 높아 MLP, CNN 모형의 우위를 정하기 어려웠으며, B-SMOTE, ADASYN을 적용한 경우에는 모형 간 분류성능 차이를 살펴볼 수 있는 McNemar-test를 실시하였을 때 두 모형의 분류성능 차이가 나타나지 않았다.

하지만 MLP 모형은 예측에 활용하는 변인의 개수가 늘어남에 따라 모형 성능이 크게 감소하는 모습이 나타난 반면, CNN 모형은 예측변인 수에 따른 모형 성능 감소폭이 적은 경향을 보였다([그림 Ⅳ-8]). MLP 모형의 경우, 81개 예측변인을 활용하여 국어 기초미달 학생을 예측한 모형 성능은 민감도가 ROS 0.128, B-SMOTE 0.106, ADASYN 0.128인 반면, CNN 모형에서는 민감도가 ROS 0.474, B-SMOTE 0.287, ADASYN 0.274로 나타났으며 AUC, F2-score, G-mean 값 또한 MLP에 비해 CNN이 높은

모습을 확인할 수 있었다.

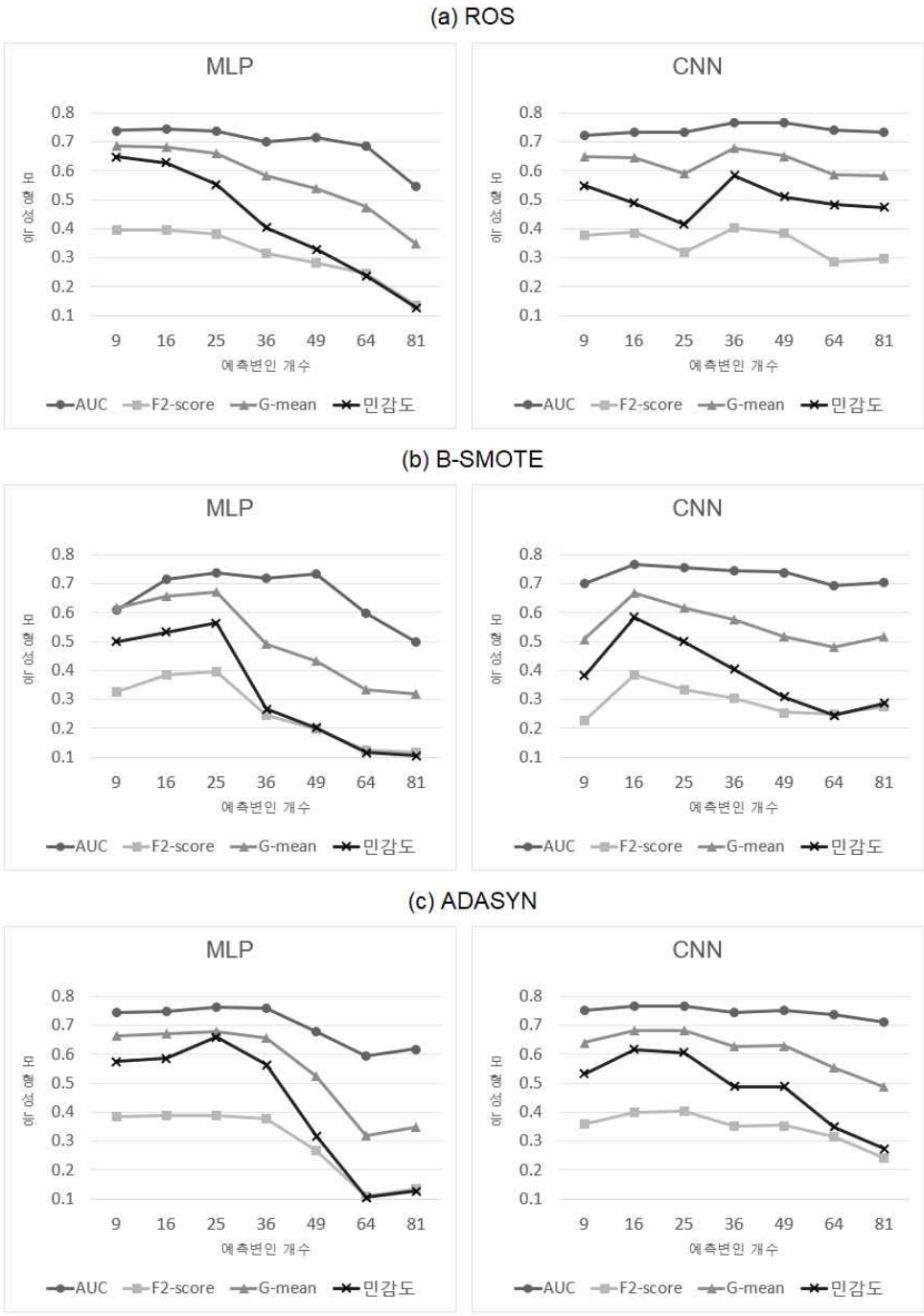
<표 Ⅳ-8> 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-표집방법)

모형		평가지표	예측변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ROS	MLP	AUC	**0.739**	0.746	0.736	0.702	0.714	0.687	0.546
		F2-score	**0.398**	0.396	0.381	0.315	0.283	0.246	0.135
		G-mean	**0.685**	0.681	0.659	0.583	0.539	0.475	0.349
		정분류율	**0.717**	0.730	0.768	0.808	0.841	0.898	0.891
		민감도	**0.649**	0.628	0.553	0.404	0.330	0.237	0.128
		특이도	**0.723**	0.739	0.785	0.840	0.882	0.953	0.952
		McNemar-test		2.4	0.8	38.3*	16.2*	21.2*	19.4*
	CNN	AUC	0.721	0.733	0.734	**0.766**	0.765	0.740	0.732
		F2-score	0.377	0.387	0.320	**0.404**	0.384	0.288	0.296
		G-mean	0.649	0.647	0.589	**0.680**	0.651	0.588	0.583
		정분류율	0.750	0.827	0.805	**0.774**	0.807	0.701	0.699
		민감도	0.550	0.489	0.415	**0.585**	0.511	0.483	0.474
		특이도	0.767	0.854	0.836	**0.789**	0.831	0.717	0.717
		McNemar-test		9.3*	1.4	35.3*	18.9*	9.2*	5.3*
B-SMOTE	MLP	AUC	0.610	**0.716**	0.739	0.718	0.735	0.598	0.611
		F2-score	0.326	**0.385**	0.398	0.246	0.197	0.126	0.119
		G-mean	0.617	**0.657**	0.672	0.492	0.433	0.335	0.321
		정분류율	0.742	**0.791**	0.783	0.862	0.873	0.898	0.907
		민감도	0.500	**0.532**	0.564	0.266	0.202	0.117	0.106
		특이도	0.761	**0.812**	0.800	0.909	0.926	0.960	0.970
		McNemar-test		5.2*	1.6	122.2*	4.7*	12.1*	4.1*
	CNN	AUC	0.701	**0.768**	0.754	0.744	0.739	0.694	0.704
		F2-score	0.226	**0.387**	0.333	0.305	0.255	0.250	0.276
		G-mean	0.508	**0.669**	0.617	0.577	0.516	0.482	0.516
		정분류율	0.653	**0.751**	0.743	0.793	0.822	0.897	0.880
		민감도	0.383	**0.585**	0.500	0.404	0.309	0.245	0.287
		특이도	0.675	**0.764**	0.762	0.824	0.863	0.949	0.927
		McNemar-test		5.4*	3.9*	40.5*	21.3*	14.7*	7.0*

모형		평가지표	예측변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ADASYN	MLP	AUC	0.746	0.749	**0.762**	0.758	0.680	0.596	0.618
		F2-score	0.384	0.388	**0.390**	0.377	0.267	0.112	0.135
		G-mean	0.665	0.670	**0.679**	0.658	0.526	0.318	0.349
		정분류율	0.755	0.753	**0.697**	0.754	0.828	0.887	0.893
		민감도	0.574	0.585	**0.660**	0.564	0.319	0.106	0.128
		특이도	0.770	0.766	**0.700**	0.769	0.868	0.949	0.954
		McNemar-test		0.09	18.4*	32.2*	84.1*	25.1*	11.1*
	CNN	AUC	0.753	**0.765**	0.765	0.745	0.753	0.736	0.712
		F2-score	0.358	**0.400**	0.404	0.351	0.354	0.315	0.244
		G-mean	0.640	**0.682**	0.684	0.627	0.629	0.553	0.487
		정분류율	0.753	**0.745**	0.759	0.780	0.785	0.826	0.813
		민감도	0.532	**0.617**	0.606	0.489	0.489	0.350	0.274
		특이도	0.771	**0.755**	0.771	0.803	0.809	0.875	0.867
		McNemar-test		3.1	2.1	13.6*	0.2	7.5*	14.8*

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-8] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교
(국어-표집방법)

3가지 표집방법 간 모형 성능 차이를 살펴보면, B-SMOTE와 ADASYN은 ROS에 비해 불균형 자료의 특성을 고려하여 자료를 생성하는 방법이지만, 소수집단 사례를 다수집단만큼 단순반복 추출하는 ROS를 적용한 모형과 비슷한 수준의 예측 성능을 보였다. 이는 <표 Ⅳ-9>는 각 표집방법에서 모형 성능이 높게 나타난 모형을 정리한 결과에서 살펴볼 수 있다.

<표 Ⅳ-9> 표집방법에 따른 MLP, CNN 모형 성능 비교 (국어)

신경망 모형	MLP			CNN		
소수집단 정보 반영방법	ROS	B-SMOTE	ADASYN	ROS	B-SMOTE	ADASYN
예측변인 수	9	16	25	36	16	16
AUC	0.739	0.716	0.762	0.766	0.768	0.765
F2-score	0.398	0.385	0.390	0.404	0.387	0.400
G-mean	0.685	0.657	0.679	0.680	0.669	0.682
정분류율	0.717	0.791	0.697	0.774	0.751	0.745
민감도	0.649	0.532	0.660	0.585	0.585	0.617
특이도	0.723	0.812	0.700	0.789	0.764	0.755

주. 동일한 신경망 모형 중 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

다음으로, 비용민감방법을 적용하였을 때의 예측변인의 개수와 심층 인공신경망 모형 종류에 따른 모형 성능은 <표 Ⅳ-10>, [그림 Ⅳ-9]와 같다. 손실함수 가중치를 조정하여 기초미달 학생 정보가 기초이상 학생 정보만큼 반영되도록 조정한 경우에도 Original 모형에 비해 모형 성능이 향상되는 모습이 나타났다. 민감도 0.455~0.755으로 기초미달 학생 예측 비율이 높아졌고, 실제 기초미달 학생을 제대로 예측한 비율(민감도)과 실제 기초이상 학생을 제대로 예측한 비율(특이도) 간 균형을 보여주는 G-mean 값은 0.562~0.703으로 집단별 예측 균형이 조정된 모습이 나타났다.

예측변인 개수에 따른 모형 성능을 살펴보면([그림 Ⅳ-9]), 표집방법과 유사하게 예측변인의 개수가 일정 수준 이상 많아지면 모형 성능이 감소하는 경향이 나타났다. 하지만 표집방법을 적용하였을 때 81개 예측변인을 활용한 경우 민감도가 MLP 0.106~0.128, CNN 0.274~0.474로 낮아진 반면, 비용민감방법을 적용하였을 때에는 MLP 0.455, CNN 0.511 정도로 감소폭이 적은 경향을 보였다.

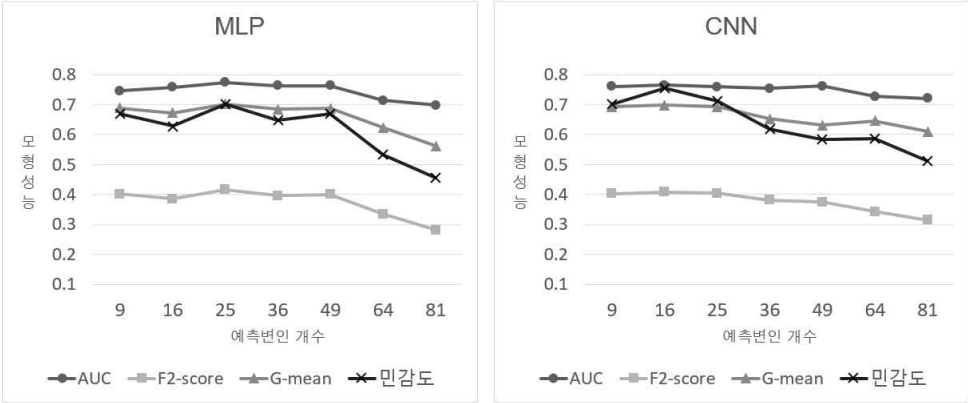
심층인공신경망 모형 종류에 따른 성능은 표집방법과 유사하게 MLP, CNN 모형 적용에 따른 예측 성능의 차이는 크지 않았으나, 예측변인 개수의 따른 모형 성능 변화에서는 차이를 보였다. <표 Ⅳ-10>의 모형별 모형 성능이 가장 좋은 모형(볼드체)을 기준으로 보았을 때, MLP 모형이 CNN 모형에 비해 모형 성능이 다소 높게 나타나긴 하였으나, 두드러지는 성능 차이가 나타나진 않았다. 하지만 예측변인 개수에 따른 모형 성능 변화는 종류에 따라 차이를 보였는데, CNN 모형이 MLP 모형에 비해 예측변인 개수가 증가하여도 모형 성능이 감소하는 정도가 MLP 보다 적은 모습이었다.

<표 Ⅳ-10> 예측모형 상세화 방법에 따른 모형 성능 비교 (국어-비용민감방법)

모형	평가지표	예측변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.747	0.758	**0.774**	0.764	0.764	0.714	0.698
	F2-score	0.402	0.386	**0.416**	0.397	0.400	0.335	0.281
	G-mean	0.689	0.673	**0.703**	0.684	0.688	0.623	0.562
	정분류율	0.706	0.716	**0.704**	0.715	0.704	0.715	0.676
	민감도	0.670	0.628	**0.702**	0.649	0.670	0.532	0.455
	특이도	0.709	0.723	**0.704**	0.720	0.707	0.729	0.694
	McNemar-test		3.4	16.2*	4.6*	32.0*		
CNN	AUC	0.761	**0.765**	0.760	0.755	0.762	0.728	0.721
	F2-score	0.403	**0.409**	0.404	0.381	0.375	0.343	0.315
	G-mean	0.692	**0.698**	0.693	0.652	0.631	0.646	0.610
	정분류율	0.684	**0.654**	0.677	0.681	0.673	0.703	0.712
	민감도	0.702	**0.755**	0.713	0.619	0.585	0.586	0.511
	특이도	0.683	**0.646**	0.674	0.642	0.681	0.712	0.727
	McNemar-test		21.9*	12.8*	10.0*	17.4*		

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-9] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (국어-비용민감방법)

모형 훈련 시 소수집단 정보를 반영하는 4가지 방법별 모형 성능이 좋게 나타난 대표모형의 결과는 <표 Ⅳ-11>과 같다. ROS 방법을 적용한 경우에는 예측변인 36개를 활용한 CNN 모형이, B-SMOTE와 ADASYN 방법을 적용한 경우에는 예측변인 16개를 활용한 CNN 모형의 모형 성능이 높게 나타났다. 또한 비용민감방법을 적용한 경우에는 25개의 예측변인을 적용한 MLP 모형의 평가 지표가 가장 높았다. 또한 각 대표 모형의 분류결과가 통계적으로 유의미한 차이가 있는지 확인하기 위하여 McNemar 검증을 실시하였으며, 4개 모형이 모두 유의수준 0.01에서 차이가 있었다.

ROS 대표모형은 다른 3개 모형에 비해 모형 성능이 낮았으며, B-SMOTE와 ADASYN은 비슷한 수준이었고, 비용민감방법 대표모형의 성능이 가장 높았다. CS 모형의 정분류율, 민감도, 특이도는 모두 약 0.7로 나타났으며, 기초미달 학생과 기초이상 학생을 모두 약 70% 정확하게 예측하는 모습을 확인할 수 있었다.

<표 Ⅳ-11> 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (국어)

소수집단 정보 반영방법	ROS	B-SMOTE	ADASYN	CS
신경망 모형	CNN	CNN	CNN	MLP
예측변인 수	36	16	16	25
AUC	0.766	0.768	0.765	0.774
F2-score	0.404	0.387	0.400	0.416
G-mean	0.680	0.669	0.682	0.703
정분류율	0.774	0.751	0.745	0.704
민감도	0.585	0.585	0.617	0.702
특이도	0.789	0.764	0.755	0.704

주. 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

이상의 국어 기초미달 예측모형 비교 결과를 정리하면, 불균형 정도를 조정하지 않은 Original 모형에서는 기초미달 학생을 거의 예측하지 못한 반면, 표집방법이나 비용민감방법을 적용하여 불균형 정도를 조정해 준 후에는 기초미달 학생 중 약 56~70%를 정확하게 예측하였다. 모형 설정에 따른 모형 성능을 살펴보면, 예측변인 개수가 일정 수준 이상 늘어나면 모형 성능이 감소하는 경향이 나타났다. 또한 비용민감방법은 표집방법에 비해 전반적인 모형 성능이 높게 나타났고, 예측변인 개수가 늘어남에 따른 모형 성능 감소도 적었다. 마지막으로 심층인공신경망 종류에 따른 성능 차이는 크지 않았으나, 표집방법과 비용민감방법 적용에 관계없이 CNN 모형은 MLP 모형에 비해 모형 성능 감소폭이 적은 모습을 확인할 수 있었다.

나. 영어

영어 기초미달 학생의 비율은 14.9%이며, 소수집단 대비 다수집단의 비율(Imbalanced Ratio, IR)이 4 이상 10 미만인 경우 중간 정도의 불균형을 보이는 자료라 볼 수 있을 때(Fernandez et al., 2008), IR 5.7로 중간 정도 불균형을 보이고 있었다.

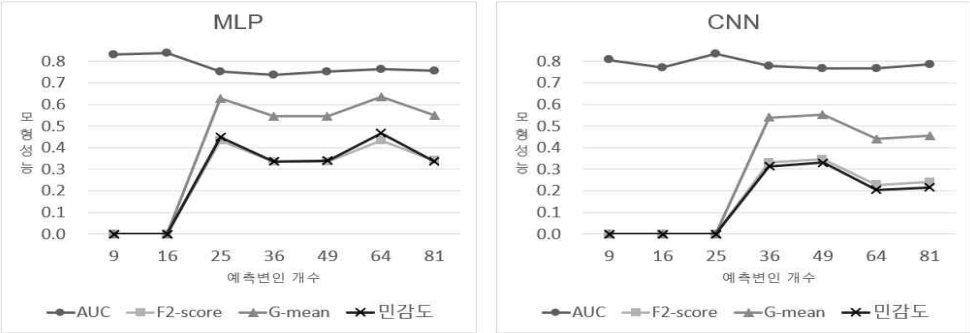
모형 훈련 시 원자료를 그대로 활용한 Original 모형의 예측 성능 결과는 <표 Ⅳ-12>, [그림 Ⅳ-10]과 같다. MLP 모형은 예측변인 25개, CNN 모형은 예측변인 36개 이상을 활용할 때부터 기초미달 학생을 예측하는 모습이 나타났으며, 국어에서는 민감도가 0.1 미만이었던 것에 비해 민감도가 높은 모습을 보였다. 하지만 MLP, CNN 모형에서 모형 성능이 높은 모형(볼드체)에서도 민감도가 0.449, 0.330으로 나타나, 예측모형을 구축할 때 불균형 조정이 필요한 모습이 나타났다.

<표 Ⅳ-12> 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-Original)

모형	평가지표	예측변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.833	0.839	**0.754**	0.736	0.753	0.765	0.755
	F2-score	-	-	**0.433**	0.335	0.335	0.432	0.341
	G-mean	0.000	0.000	**0.627**	0.545	0.546	0.614	0.550
	정분류율	0.855	0.855	**0.815**	0.808	0.798	0.816	0.821
	민감도	0.000	0.000	**0.449**	0.335	0.341	0.437	0.335
	특이도	1.000	1.000	**0.877**	0.888	0.875	0.866	0.903
	McNemar-test			68.9*	8.2*	9.1*	12.5*	24.0*
CNN	AUC	0.807	0.771	0.834	0.779	**0.767**	0.769	0.786
	F2-score	-	-	-	0.331	**0.345**	0.228	0.242
	G-mean	0.000	0.000	0.000	0.540	**0.552**	0.442	0.455
	정분류율	0.855	0.855	0.855	0.840	**0.839**	0.842	0.852
	민감도	0.000	0.000	0.000	0.314	**0.330**	0.205	0.216
	특이도	1.000	1.000	1.000	0.929	**0.925**	0.950	0.959
	McNemar-test				34.0*	29.1*	22.3*	2.2

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-10] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (영어-Original)

표집방법을 적용하여 모형 훈련 시 소수집단 정보를 다수집단만큼 반영하였을 때, 예측변인의 개수와 심층인공신경망 종류에 따른 모형 성능은 <표 Ⅳ-13>, [그림 Ⅳ-11]과 같다. 각 모형에서 모형 성능이 가장 높았던 모형(볼드체)을 살펴보면, 표집방법에 관계없이 기초이상 학생과 기초미달 학생 예측 불균형이 완화되었고(G-mean 0.736~0.765), 기초미달 학생 예측률이 71~79%(민감도 0.714~0.789)로 증가하였다. F2-score과 AUC 역시 Original 모형에 비해 값이 증가하였다.

예측변인의 개수에 따른 결과를 살펴보면, 예측변인의 개수가 9~25개 정도일 때 모형 성능이 높게 나타나는 경향을 보였으며(<표 Ⅳ-13>의 볼드체 참고), 예측변인이 49개 이상으로 늘어나는 경우에는 성능이 감소하는 모습이 나타났다.

심층인공신경망 종류에 따른 모형 성능 결과에서는 국어와 유사하게 MLP, CNN 모형 적용에 따른 예측 성능의 차이는 크지 않았고, 예측변인 개수의 따른 모형 성능에서는 차이를 보였다. <표 Ⅳ-13>에서 각 모형에서 성능이 높게 나타난 모형(볼드체)을 기준으로, ROS에서는 MLP, CNN 모형의 AUC(0.835>0.822), F2-score(0.617<0.618), G-mean(0.765>0.763), 민감도(0.714<0.746)를 비교하였을 때, 상대적으로 성능이 우수한 모형을 확인하기 어려웠으며, B-SMOTE, ADASYN을 적용한 경우에는 모형 간 분류성능 차이를 살펴볼 수 있는 McNemar-test를 실시하였을 때 모형 간 분류성능 차이가 나타나지 않았다.

하지만 예측에 활용하는 변인 개수에 따른 MLP, CNN 모형 성능은 차이가 있었다. [그림 Ⅳ-11]과 같이, MLP 모형은 예측에 활용하는 변인의 개수 36개 이상으로 늘어남에 따라 모형 성능이 감소하였으며, 81개 예측변인을 활용한 경우에는 민감도가 0.330~0.416 정도로 나타났다. 상대적으로 CNN 모형에서는 예측변인 수가 많아져도 따른 모형 성능 감소하는 정도가 적었으며, 특히 ROS를 적용한 CNN 모형에서는 예측변인 개수에 관계없이 비슷한 수준의 모형 성능이 나타났다.

<표 Ⅳ-13> 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-표집방법)

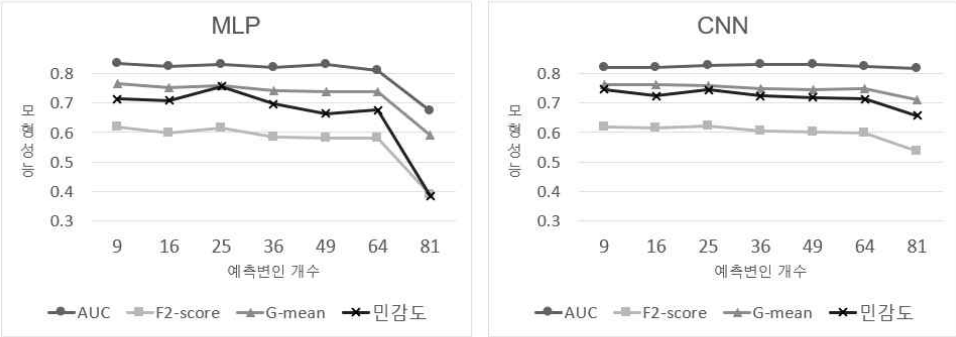
모형		평가지표	예측변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ROS	MLP	AUC	**0.835**	0.824	0.830	0.820	0.831	0.810	0.674
		F2-score	**0.617**	0.598	0.615	0.586	0.580	0.582	0.389
		G-mean	**0.765**	0.751	0.760	0.741	0.738	0.740	0.590
		정분류율	**0.805**	0.783	0.763	0.775	0.798	0.791	0.832
		민감도	**0.714**	0.708	0.757	0.697	0.665	0.676	0.384
		특이도	**0.820**	0.795	0.763	0.788	0.820	0.810	0.908
		McNemar-test		5.6*	16.3*	12.0*	2.0		
	CNN	AUC	**0.822**	0.821	0.828	0.831	0.830	0.826	0.818
		F2-score	**0.618**	0.615	0.622	0.604	0.602	0.598	0.537
		G-mean	**0.763**	0.763	0.765	0.749	0.747	0.750	0.711
		정분류율	**0.780**	0.792	0.780	0.768	0.767	0.778	0.754
		민감도	**0.746**	0.724	0.745	0.724	0.720	0.714	0.657
		특이도	**0.785**	0.804	0.787	0.776	0.776	0.789	0.769
		McNemar-test		3.9*	18.8*	8.6*	1.3		
B-SMOTE	MLP	AUC	0.812	0.819	**0.829**	0.815	0.806	0.807	0.771
		F2-score	0.579	0.578	**0.618**	0.484	0.520	0.481	0.404
		G-mean	0.736	0.735	**0.763**	0.667	0.695	0.665	0.603
		정분류율	0.760	0.764	**0.768**	0.795	0.791	0.777	0.809
		민감도	0.703	0.697	**0.757**	0.530	0.584	0.541	0.416
		특이도	0.770	0.775	**0.770**	0.840	0.826	0.817	0.875
		McNemar-test		1.1	18.2*	22.3*	15.2*		
	CNN	AUC	0.825	**0.827**	0.763	0.818	0.813	0.805	0.802
		F2-score	0.583	**0.609**	0.483	0.558	0.536	0.521	0.467
		G-mean	0.739	**0.756**	0.666	0.722	0.706	0.701	0.655
		정분류율	0.771	**0.755**	0.800	0.791	0.787	0.780	0.783
		민감도	0.697	**0.757**	0.524	0.638	0.611	0.610	0.519
		특이도	0.784	**0.754**	0.847	0.817	0.816	0.806	0.827
		McNemar-test		21.9*	16.3*	6.9*	1.9		

모형		평가지표	예측변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ADASYN	MLP	AUC	0.805	0.793	**0.825**	0.796	0.808	0.763	0.744
		F2-score	0.621	0.556	**0.583**	0.524	0.546	0.393	0.326
		G-mean	0.730	0.719	**0.736**	0.696	0.712	0.593	0.538
		정분류율	0.758	0.761	**0.740**	0.756	0.757	0.830	0.798
		민감도	0.685	0.665	**0.730**	0.622	0.654	0.389	0.330
		특이도	0.777	0.777	**0.742**	0.779	0.774	0.904	0.878
		McNemar-test		1.8	15.8*	11.2*	4.5*		
	CNN	AUC	**0.824**	0.812	0.806	0.819	0.808	0.790	0.746
		F2-score	**0.610**	0.531	0.571	0.571	0.544	0.474	0.423
		G-mean	**0.752**	0.702	0.730	0.731	0.712	0.656	0.620
		정분류율	**0.727**	0.784	0.763	0.784	0.794	0.823	0.790
		민감도	**0.789**	0.605	0.686	0.665	0.616	0.487	0.454
		특이도	**0.716**	0.815	0.776	0.804	0.824	0.883	0.847
		McNemar-test		17.0*	16.5*	3.2	9.6*		

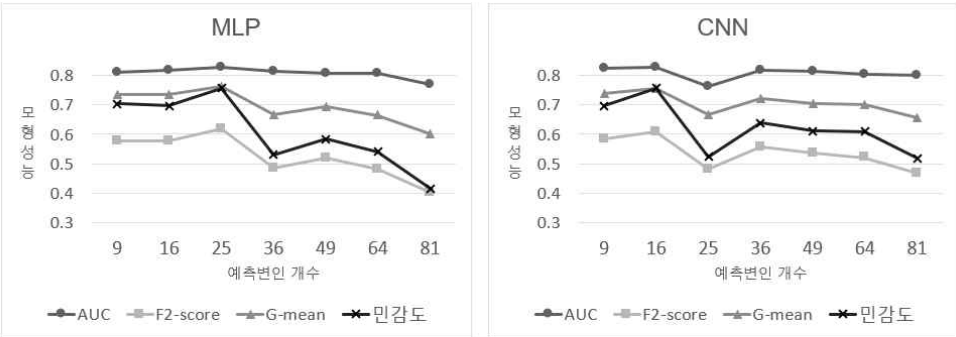
주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.

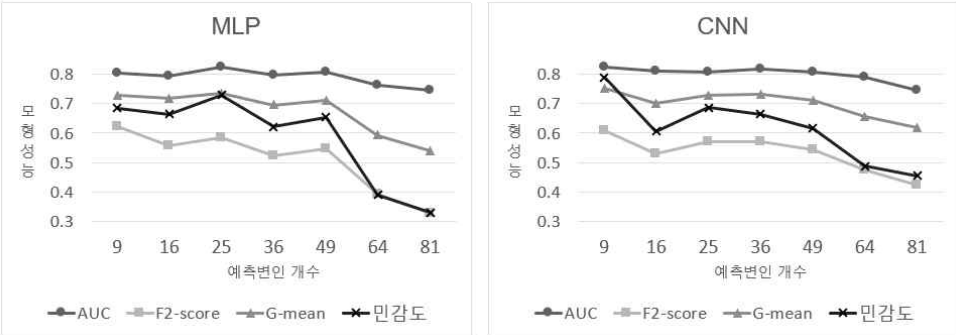
(a) ROS



(b) B-SMOTE



(c) ADASYN



[그림 Ⅳ-11] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교
(영어-표집방법)

3가지 표집방법 간 모형 성능 차이에서도 국어 기초미달 예측모형 결과와 유사한 모습이 나타났다. 표집방법별 모형 성능이 높은 모형을 정리한 <표 Ⅳ-14>를 살펴보면, B-SMOTE, ADASYN는 표집과정에서 불균형 자료의 특성을 반영해주는 방법이지만, ROS을 적용한 모형과 성능 차이를 확인하기 어려웠다. 오히려 ROS를 적용한 경우에 예측변인의 개수, 심층인공신경망 종류에 따른 모형 성능 변화가 상대적으로 안정적인 모습이 나타나기도 하였다([그림 Ⅳ-11]).

<표 Ⅳ-14> 표집방법에 따른 MLP, CNN 모형 성능 비교 (영어)

신경망 모형	MLP			CNN		
소수집단 정보 반영방법	ROS	B-SMOTE	ADASYN	ROS	B-SMOTE	ADASYN
예측변인 수	9	25	25	9	16	9
AUC	0.835	0.829	0.825	0.822	0.827	0.824
F2-score	0.617	0.618	0.583	0.618	0.609	0.610
G-mean	0.765	0.763	0.736	0.763	0.756	0.752
정분류율	0.805	0.768	0.740	0.780	0.755	0.727
민감도	0.714	0.757	0.730	0.746	0.757	0.789
특이도	0.820	0.770	0.742	0.785	0.754	0.716

주. 동일한 신경망 모형 중 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

다음으로, 비용민감방법을 적용하였을 때의 예측변인의 개수와 심층인공신경망 모형 종류에 따른 모형 성능은 <표 Ⅳ-15>, [그림 Ⅳ-12]와 같다. 표집방법을 적용한 모형과 같이 비용민감방법으로 불균형 정도를 조정해주었을 때에도 Original 모형에 비해 모형 성능이 향상되는 모습이 나타났다. 민감도 0.762~0.832로 기초미달 학생 예측 비율이 높아졌고, 민감도와 특이도 간 균형을 보여주는 G-mean 값은 0.755~0.766으로 집단별 예측 균형이 조정된 모습이 나타났다.

예측변인 개수에 따른 모형 성능을 살펴보면([그림 Ⅳ-12]), 예측변인

의 개수가 64개 이상 늘어나면 모형 성능이 감소하는 경향이 나타났다. 하지만 표집방법을 적용하였을 때에 비해 모형 성능이 감소하는 폭이 적은 모습을 볼 수 있었다.

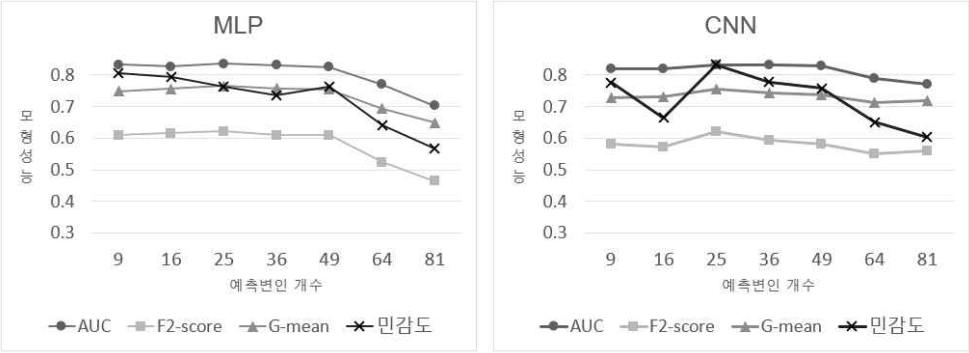
심층인공신경망 모형 종류에 따른 성능 차이는 크지 않았으나, CNN 모형에서는 MLP 모형에 비해 많은 예측변인을 활용하였을 때 모형 성능이 감소하는 정도가 적은 경향을 보였다. <표 Ⅳ-15>의 모형별 모형 성능이 가장 좋은 모형(볼드체)을 기준으로 보았을 때, MLP 모형의 AUC, F2-score, G-mean이 CNN보다 높게 나타나 상대적으로 좋은 성능을 보였으나, 그 차이가 크지 않았고 민감도는 CNN이 높은 모습이 나타났다. 또한 예측변인의 개수가 64, 81개일 때 CNN 모형의 감소폭은 적은 모습을 확인할 수 있었다.

<표 Ⅳ-15> 예측모형 상세화 방법에 따른 모형 성능 비교 (영어-비용민감방법)

모형	평가지표	예측변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.833	0.826	**0.836**	0.832	0.825	0.770	0.703
	F2-score	0.610	0.616	**0.622**	0.609	0.609	0.524	0.465
	G-mean	0.749	0.756	**0.766**	0.757	0.755	0.693	0.649
	정분류율	0.713	0.730	**0.768**	0.773	0.749	0.735	0.716
	민감도	0.805	0.795	**0.762**	0.735	0.762	0.640	0.568
	특이도	0.697	0.720	**0.769**	0.780	0.747	0.751	0.742
	McNemar-test		3.5	21.3*	13.3*	9.8*		
CNN	AUC	0.821	0.819	**0.833**	0.832	0.830	0.789	0.771
	F2-score	0.580	0.571	**0.621**	0.593	0.581	0.551	0.561
	G-mean	0.727	0.731	**0.755**	0.743	0.736	0.712	0.718
	정분류율	0.695	0.784	**0.705**	0.719	0.720	0.759	0.814
	민감도	0.776	0.665	**0.832**	0.778	0.758	0.651	0.604
	특이도	0.681	0.804	**0.684**	0.709	0.714	0.778	0.853
	McNemar-test		12.8*	12.8*	8.8*	5.6*		

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-12] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교
(영어-비용민감방법)

소수집단 정보 반영방법에 따라 모형 성능이 가장 우수하게 나타난 모형을 비교한 결과는 <표 Ⅳ-16>과 같다. ROS 방법을 적용한 경우에는 예측변인 9개를 활용한 MLP 모형, ADASYN 방법을 적용한 경우에는 예측변인 9개를 활용한 CNN 모형, B-SMOTE와 CS 방법을 적용한 경우에는 예측변인 25개를 활용한 MLP 모형의 모형 성능이 높게 나타났다. 또한 McNemar 검증을 실시하여 대표모형의 분류결과가 통계적으로 유의미한 차이가 있는지 살펴보았으며, 그 결과 4개 모형이 모두 유의수준 0.05에서 차이가 있는 것을 확인하였다. 3가지 표집방법에 비해 CS방법의 대표모형의 모형 성능이 가장 좋게 나타났으며, 기초이상 학생과 기초미달 학생 모두 약 76%를 제대로 예측하는 결과를 확인할 수 있었다.

<표 Ⅳ-16> 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (영어)

소수집단 정보 반영방법	ROS	SMOTE	ADASYN	CS
신경망 모형	MLP	MLP	CNN	MLP
예측변인 수	9	25	9	25
AUC	0.835	0.829	0.824	0.836
F2-score	0.617	0.618	0.610	0.622
G-mean	0.765	0.763	0.752	0.766
정분류율	0.805	0.768	0.727	0.768
민감도	0.714	0.757	0.789	0.762
특이도	0.820	0.770	0.716	0.769

주. 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

이상의 영어 기초미달 예측모형 비교 결과를 살펴보면, 예측변인 개수가 일정 수준 이상 늘어나면 모형 성능이 감소하는 경향이 나타났다. 그러나 모형 성능이 감소하는 정도는 소수집단 정보 방영방법, 심층인공신경망 종류에 따라 차이를 보였다. 표집방법에 비해 비용민감방법을 적용한 경우, MLP 모형에 비해 CNN 모형을 활용하는 경우에 상대적으로 모형 성능 감소폭이 적은 모습을 확인할 수 있었다. 또한 전반적으로 비용민감방법의 모형 성능이 표집방법에 비해 높은 모습이 나타났고, 3가지 표집방법에 따른 모형 성능 차이나, 심층인공신경망 종류에 따른 성능 차이는 크지 않았다.

다. 수학

수학 기초미달 학생의 비율은 전체의 20.6%였고, 다수집단 대비 소수집단의 비율(Imbalanced Ratio, IR)은 3.85로 나타났다. IR 4를 기준으로 4 미만은 낮은 불균형, 4 이상은 중간 정도의 불균형 자료라 할 때(Fernandez et al., 2008), 불균형 정도는 낮은 편에 가까운 모습을 확인할 수 있었다.

수학 기초미달 학생 예측을 위한 모형 훈련 시 불균형 정도를 조정해 주지 않은 Original 모형의 예측 성능은 <표 Ⅳ-17>과 같다. MLP, CNN 모형에서 모형 성능이 높은 모형(볼드체)을 기준으로 살펴보았을 때, 각 민감도가 0.456, 0.537로 나타나 기초미달 학생의 약 절반 정도를 정확하게 예측하는 모습이 나타났다.

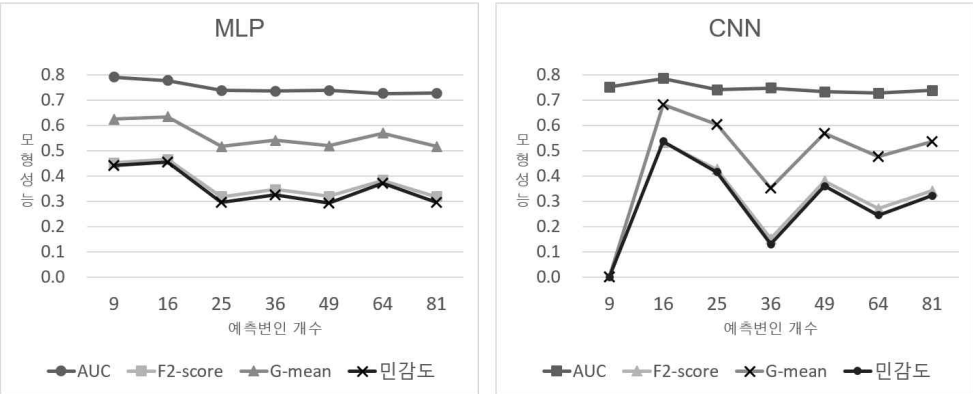
<표 Ⅳ-17> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-Original)

모형	평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.791	**0.777**	0.738	0.736	0.739	0.726	0.727
	F2-score	0.452	**0.465**	0.318	0.346	0.319	0.382	0.318
	G-mean	0.625	**0.634**	0.517	0.541	0.519	0.569	0.517
	정분류율	0.791	**0.791**	0.774	0.777	0.788	0.767	0.774
	민감도	0.441	**0.456**	0.296	0.326	0.293	0.370	0.296
	특이도	0.885	**0.881**	0.902	0.897	0.920	0.873	0.902
	McNemar-test		4.1*	13.5*	8.6*	8.9*	15.3*	10.8*

모형	평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
CNN	AUC	0.752	**0.786**	0.741	0.748	0.733	0.728	0.739
	F2-score	-	**0.532**	0.426	0.151	0.380	0.270	0.340
	G-mean	0.000	**0.681**	0.604	0.352	0.569	0.476	0.535
	정분류율	0.789	**0.794**	0.781	0.784	0.788	0.783	0.769
	민감도	0.000	**0.537**	0.415	0.130	0.359	0.244	0.322
	특이도	1.000	**0.862**	0.879	0.958	0.902	0.927	0.888
	McNemar-test		35.9*	17.2*	21.5*	20.0*	12.0*	8.2*

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-13] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교 (수학-Original)

표집방법을 적용하여 모형 훈련 시 소수집단 정보를 다수집단만큼 반영하였을 때, 예측변인의 개수와 심층인공신경망 종류에 따른 모형 성능은 <표 Ⅳ-18>, [그림 Ⅳ-14]와 같다. 각 모형에서 모형 성능이 가장 높았던 모형(볼드체)을 살펴보면, 표집방법에 관계없이 기초미달 학생 예측률이 64~73%(민감도 0.637~0.726)로 증가하였으며, 이때의 AUC,

F2-score, G-mean, 역시 Original 모형에 비해 값이 증가하였다. 그러나 ROS를 적용한 CNN 모형을 제외하고, 예측변인 81개를 활용한 경우에는 Original 모형에서 모형 성능이 좋은 모형보다(<표 Ⅳ-17>의 볼드체 참고) 예측 성능이 낮은 모습이 나타나기도 하였다. 이는 불균형 정도를 조정해주는 경우에도 예측변인 활용 및 모형 설정에 따라 결과가 달라질 수 있음을 보여준다.

예측변인의 개수에 따른 결과를 살펴보면, 수학에서도 예측변인의 개수가 9~25개 정도일 때 모형 성능이 높게 나타나는 경향을 보였으며(<표 Ⅳ-18>의 볼드체 참고), 예측변인이 49개 이상으로 늘어나는 경우에는 성능이 감소하는 모습이 나타났다.

심층인공신경망 종류에 따른 모형 성능 결과에서는 국어, 영어 예측 모형 결과와 유사하게 MLP, CNN 모형 적용에 따른 예측 성능의 차이는 크지 않았고, 예측변인 개수의 따른 모형 성능에서는 차이를 보였다. <표 Ⅳ-18>에서 각 모형에서 성능이 높게 나타난 모형(볼드체)을 기준으로 비교하였을 때, ROS, B-SMOTE, ADASYN 모두 MLP 모형을 적용한 모형의 성능이 더 좋은 모습이 나타났지만, MLP, CNN 간 각 모형지표의 차이가 0.001~0.056 정도로 다소 적은 차이를 보임을 확인할 수 있었다. 하지만 예측에 활용하는 변인 개수에 따른 MLP, CNN 모형 성능은 차이가 있었다. [그림 Ⅳ-14]와 같이, MLP 모형은 예측에 활용하는 변인의 개수 36개 이상으로 늘어남에 따라 모형 성능이 크게 감소하는 모습이 나타난 반면, CNN 모형에서는 예측변인의 개수가 늘어나도 비슷한 수준의 성능이 나타나거나 감소하는 정도가 적은 모습이었다.

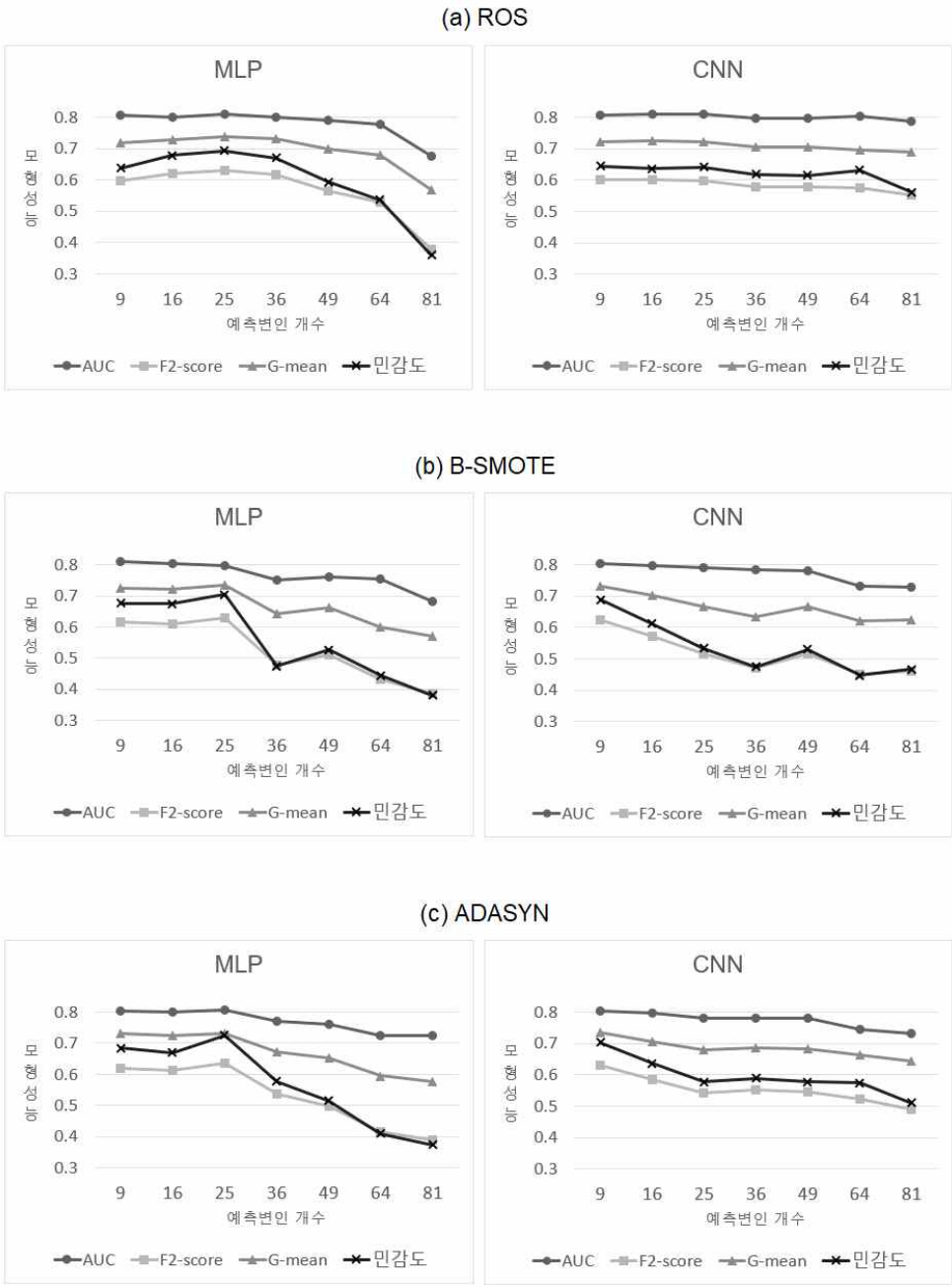
<표 Ⅳ-18> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-표집방법)

모형		평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ROS	MLP	AUC	0.807	0.801	**0.811**	0.801	0.791	0.776	0.677
		F2-score	0.597	0.619	**0.631**	0.617	0.565	0.530	0.379
		G-mean	0.720	0.730	**0.739**	0.731	0.700	0.679	0.568
		정분류율	0.777	0.763	**0.768**	0.770	0.777	0.790	0.784
		민감도	0.637	0.678	**0.693**	0.670	0.593	0.537	0.359
		특이도	0.814	0.786	**0.788**	0.797	0.826	0.857	0.898
		McNemar-test		2.6	4.2*	5.1*	9.4*		
	CNN	AUC	0.808	**0.810**	0.809	0.796	0.796	0.803	0.787
		F2-score	0.601	**0.602**	0.599	0.579	0.578	0.575	0.553
		G-mean	0.722	**0.725**	0.721	0.706	0.707	0.697	0.689
		정분류율	0.773	**0.786**	0.775	0.766	0.771	0.740	0.780
		민감도	0.644	**0.637**	0.641	0.619	0.615	0.632	0.561
		특이도	0.808	**0.826**	0.811	0.806	0.813	0.769	0.846
		McNemar-test		4.8*	3.9*	5.1*	1.0		
B-SMOTE	MLP	AUC	0.809	0.803	**0.797**	0.751	0.760	0.753	0.682
		F2-score	0.616	0.610	**0.630**	0.478	0.512	0.431	0.387
		G-mean	0.725	0.722	**0.734**	0.642	0.663	0.600	0.571
		정분류율	0.756	0.752	**0.752**	0.786	0.771	0.733	0.754
		민감도	0.676	0.674	**0.704**	0.474	0.526	0.444	0.381
		특이도	0.778	0.773	**0.765**	0.869	0.837	0.810	0.853
		McNemar-test		0.6*	7.9*	15.6*	19.3*		
	CNN	AUC	**0.805**	0.798	0.790	0.783	0.782	0.731	0.729
		F2-score	**0.625**	0.573	0.517	0.471	0.515	0.452	0.460
		G-mean	**0.733**	0.702	0.666	0.635	0.665	0.620	0.625
		정분류율	**0.761**	0.766	0.770	0.770	0.771	0.771	0.759
		민감도	**0.689**	0.611	0.533	0.474	0.530	0.446	0.467
		특이도	**0.780**	0.807	0.833	0.850	0.836	0.862	0.837
		McNemar-test		6.8*	12.2*	8.9*	7.5*		

모형		평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
			9개	16개	25개	36개	49개	64개	81개
ADASYN	MLP	AUC	0.805	0.800	**0.806**	0.771	0.761	0.726	0.726
		F2-score	0.621	0.613	**0.636**	0.536	0.499	0.417	0.391
		G-mean	0.730	0.726	**0.732**	0.673	0.652	0.595	0.577
		정분류율	0.758	0.763	**0.736**	0.740	0.761	0.766	0.781
		민감도	0.685	0.670	**0.726**	0.578	0.515	0.411	0.374
		특이도	0.777	0.787	**0.739**	0.783	0.827	0.860	0.890
		McNemar-test		0.9	14.6*	14.2*	9.8*		
	CNN	AUC	**0.804**	0.798	0.782	0.780	0.780	0.744	0.732
		F2-score	**0.632**	0.585	0.543	0.552	0.545	0.523	0.490
		G-mean	**0.735**	0.706	0.680	0.686	0.683	0.662	0.644
		정분류율	**0.755**	0.752	0.754	0.755	0.759	0.724	0.748
		민감도	**0.704**	0.637	0.578	0.589	0.578	0.574	0.511
		특이도	**0.768**	0.783	0.801	0.800	0.808	0.764	0.812
		McNemar-test		18.0*	12.3*	1.9	0.8		

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-14] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교
(수학-표집방법)

ROS, B-SMOTE, ADASYN 표집방법별 모형 성능이 높은 모형을 정리한 <표 Ⅳ-19>를 살펴보면, MLP 모형에서는 ROS, ADASYN을 적용한 경우, CNN 모형에서는 ADASYN 모형을 적용한 경우의 모형 성능이 높게 나타났다. 하지만 [그림 Ⅳ-14]를 살펴보면 표집방법에 따른 결과 차이보다는 예측변인의 활용이나 심층인공신경망 모형에 설정에 따른 차이가 더 클 수 있는 모습[39]을 볼 수 있고, 국어와 영어 과목에서는 표집방법에 따른 모형 성능 차이를 확인하기 어려웠다는 점에서 <표 Ⅳ-19>의 결과를 표집방법에 따른 결과로 해석하기에는 어려운 면이 있었다.

<표 Ⅳ-19> 표집방법에 따른 MLP, CNN 모형 성능 비교 (수학)

신경망 모형	MLP			CNN		
소수집단 정보 반영방법	ROS	B-SMOTE	ADASYN	ROS	B-SMOTE	ADASYN
예측변인 수	25	25	25	16	9	9
AUC	0.811	0.797	0.806	0.810	0.805	0.804
F2-score	0.631	0.630	0.636	0.602	0.625	0.632
G-mean	0.739	0.734	0.732	0.725	0.733	0.735
정분류율	0.768	0.752	0.736	0.786	0.761	0.755
민감도	0.693	0.704	0.726	0.637	0.689	0.704
특이도	0.788	0.765	0.739	0.826	0.780	0.768

주. 동일한 신경망 모형 중 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

39) [그림 Ⅳ-14]의 그래프를 살펴보면, MLP 모형에서는 예측변인 개수가 25개 이상일 때 모형 성능이 점차 감소하고, CNN 모형에서는 상대적으로 예측에 활용하는 변인이 9개, 16개일 때 모형 성능이 높은 경향을 확인할 수 있었으며, 이러한 경향이 ROS, B-SMOTE, ADASYN 적용에 상관없이 유사하게 나타났다. 따라서 표집방법에 비해 예측변인의 개수나 심층인공신경망 종류에 따른 결과 차이가 클 수 있음을 추측해볼 수 있다.

다음으로, 비용민감방법을 적용하였을 때의 예측변인의 개수와 심층인공신경망 모형 종류에 따른 모형 성능은 <표 Ⅳ-20>, [그림 Ⅳ-15]와 같다. 비용민감방법으로 불균형 정도를 조정해준 경우에도 Original 모형에 비해 모형 성능이 향상되는 모습이 나타났다. 전반적으로 민감도 0.644~0.747로 기초미달 학생 예측 비율이 높아졌고, 민감도와 특이도 간 균형을 보여주는 G-mean 값은 0.698~0.746으로 집단별 예측 균형이 조정된 모습이 나타났다. 또한 표집방법과는 달리 비용민감방법을 적용한 경우에는 예측변인 개수에 상관없이 모든 모형에서 Original 대표모형(<표 Ⅳ-17> 볼드체 참고)보다 모형 성능이 높았다.

예측변인 개수에 따른 모형 성능을 살펴보면([그림 Ⅳ-15]), 표집방법을 적용한 모형에 비해 비용민감방법을 적용한 경우에는 예측모형의 성능 이 감소하는 정도가 매우 적었다. MLP 모형에서 예측변인의 개수가 64개 이상 커질 때 성능이 낮아지긴 하였지만 그 정도가 크지 않았다.

심층인공신경망 모형 종류에 따른 성능은 예측변인 25개를 활용한 MLP 모형의 성능이 가장 좋은 모습을 보였으나, CNN 모형과의 두드러진 성능 차이가 있다고 판단하기는 어려웠다. 또한 예측변인 개수에 따른 성능 차이 역시 비슷하게 나타났다. 특히, CNN 모형은 9~81개 예측변인을 활용할 때의 모형 성능이 매우 유사하게 나타났다.

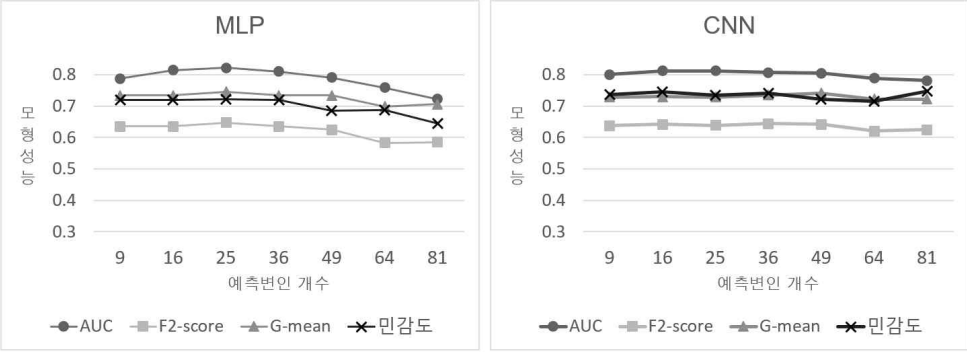
<표 Ⅳ-20> 예측모형 상세화 방법에 따른 모형 성능 비교 (수학-비용민감방법)

모형	평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
MLP	AUC	0.787	0.814	**0.822**	0.810	0.791	0.758	0.722
	F2-score	0.635	0.635	**0.647**	0.635	0.625	0.583	0.585
	G-mean	0.734	0.734	**0.746**	0.734	0.734	0.698	0.706
	정분류율	0.743	0.743	**0.761**	0.743	0.765	0.705	0.747
	민감도	0.719	0.719	**0.722**	0.719	0.685	0.687	0.644
	특이도	0.750	0.750	**0.771**	0.750	0.786	0.709	0.774
	McNemar-test		0.3	6.3*	3.8	6.2*		

모형	평가지표	기초미달 여부 예측에 활용한 설명변인의 개수						
		9개	16개	25개	36개	49개	64개	81개
CNN	AUC	0.800	0.812	0.812	0.807	0.804	0.788	0.781
	F2-score	0.637	0.642	0.638	0.644	0.642	0.620	0.625
	G-mean	0.728	0.730	0.728	0.735	0.741	0.722	0.722
	정분류율	0.723	0.723	0.724	0.732	0.752	0.726	0.707
	민감도	0.737	0.744	0.735	0.741	0.722	0.715	0.747
	특이도	0.720	0.717	0.721	0.730	0.759	0.729	0.697
	McNemar-test		2.0	3.1	4.2*	4.9*		

주1. 각 McNemar-Test 결과는 왼쪽 모형과의 비교 결과이며, 유의확률 .05 이하이면 두 모형의 예측 성능이 차이가 있음을 의미한다. * : $p < 0.05$

주2 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-15] 예측변인 개수에 따른 MLP, CNN 모형의 성능 비교
(수학-비용민감방법)

소수집단 정보 반영방법별 모형 성능이 우수하게 나타난 모형을 비교한 결과는 <표 Ⅳ-21>과 같다. 수학 기초미달 예측에서는 ROS, ADASYN, CS 방법을 적용한 경우에는 예측변인 25개를 활용한 MLP 모형의 성능이 높게 나타났다. SMOTE 방법을 적용한 경우에는 예측변인 25개를 활용한 MLP 모형과 예측변인 9개를 활용한 CNN 모형의 성능이 높았으며, McNemar 검증 결과 두 모형 간 차이가 유의미하지 않았다. 하지만 ROS, SMOTE, ADASYN, CS 대표모형 간 차이는 유의수준 0.01에

서 다른 모형으로 구분할 수 있었다. 이 중 모형 성능이 가장 좋은 모형은 CS 대표모형이었으며, 정분류율이 약 76%일 때, 기초미달 학생의 약 72%, 기초이상 학생의 약 77%를 정확하게 예측하는 결과를 확인할 수 있었다.

<표 Ⅳ-21> 소수집단 정보 반영방법별 대표모형의 모형 성능 비교 (수학)

소수집단 정보 반영방법	ROS	SMOTE		ADASYN	CS
신경망 모형	MLP	MLP	CNN	MLP	MLP
예측변인 수	25	25	9	25	25
AUC	0.811	0.797	0.810	0.806	0.822
F2-score	0.631	0.630	0.637	0.636	0.647
G-mean	0.739	0.734	0.741	0.732	0.746
정분류율	0.768	0.752	0.764	0.736	0.761
민감도	0.693	0.704	0.704	0.726	0.722
특이도	0.788	0.765	0.780	0.739	0.771

주. 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였다.

수학 기초미달 예측모형을 살펴본 결과를 종합하면 다음과 같다. 수학에서는 불균형 정도를 조정해주지 않은 Original 모형도 절반 정도의 기초미달 학생을 제대로 예측하는 모습이 나타났다. 그러나 표집방법 및 비용민감방법을 적용하였을 때에는 민감도가 0.693~0.726까지 향상되었다. 모형 설정 방법에 따른 결과를 살펴보면, 전반적으로 예측변인의 개수가 일정 수준 이상 늘어나면 모형 성능이 감소하는 경향이 나타났으며, CNN 모형에 비해 MLP 모형에서 예측변인 개수 증가에 따른 모형 감소폭이 더 크게 나타났다. 또한 표집방법에 비해 비용민감방법을 적용한 경우에 상대적으로 예측변인 개수에 상관없이 모형 성능이 안정적인 모습을 보였다.

라. 예측모형 성능 비교 결과 종합

국어, 영어, 수학 기초미달 학생 예측모형을 예측변인의 수, 소수집단 정보 반영방법, 심층인공신경망 종류에 따라 상세화하고 모형 성능을 살펴본 결과를 종합하면 다음과 같다.

불균형 정도를 조정해주지 않은 Original 모형은 모든 과목에서 기초미달 학생보다 기초이상 학생을 예측하는 데 치우친 경향을 확인할 수 있었다. 특히 불균형 정도가 높은 수준이었던 국어에서는 거의 모든 학생을 기초이상으로 예측하는 모습이 나타났으며, 상대적으로 불균형 정도가 덜했던 영어와 수학에서는 기초미달 학생 중 약 33~53% 정도를 예측하였다. 따라서 불균형 정도를 조정할 필요가 있었으며, 표집방법과 비용민감방법을 적용한 후에는 국어 약 70%, 영어 약 76%, 수학 약 72%의 기초미달 학생을 예측하고, 민감도와 특이도의 예측 불균형도 조정된 모습이 나타났다.

과목별 예측변인 개수 및 소수집단 정보 반영방법에 따른 기초미달 학생 MLP, CNN 예측모형 성능 비교 결과는 〈표 Ⅳ-22〉, 〈표 Ⅳ-23〉, [그림 Ⅳ-16], [그림 Ⅳ-17]과 같다. 단, 과목별 결과를 살펴보았을 때, 3가지 표집방법 중 B-SMOTE, ADASYN 방법과 ROS 간 결과 차이가 크지 않았고, 오히려 ROS 결과가 상대적으로 안정적으로 나타나, 3가지 표집방법 중 ROS 결과를 대표로 제시하였다.

예측변인의 개수에 따른 모형 성능 결과를 살펴보면, 모든 과목에서 예측에 활용하는 변인의 개수가 일정 수준 이상으로 늘어남에 따라 모형 성능이 감소하는 경향이 나타났다. 특히 국어의 MLP 모형에서는 모형 성능이 큰 폭으로 감소하는 모습이 확인되기도 하였다. 이는 예측변인이 많다는 것은 활용할 수 있는 정보가 많다는 것을 의미할 수 있지만, 종속변인과 관련이 높은 변인을 중심으로 모형을 설정하는 것이 필요할 수 있음을 보여주는 부분이다.

소수집단 정보 반영방법에 따른 모형 성능 결과를 살펴보면, 모든 과

목에서 표집방법에 비해 비용민감방법을 적용한 경우의 모형 성능이 더 좋게 나타난 모습을 확인할 수 있었지만, 그 차이가 크다고 판단하기는 어려웠다. 모형 성능에서는 큰 차이를 보이진 않았지만, 비용민감방법은 표집방법에 비해 예측에 많은 변인을 활용하여도 모형 성능이 감소하는 정도가 적었으며, 일부 모형에서는 거의 비슷한 수준의 성능을 유지하는 것을 확인할 수 있었다. 구체적으로 예측변인의 개수가 9~36개 정도일 때에는 표집방법과 비용민감방법에 따른 성능 차이가 크지 않았지만, 예측변인의 개수가 49개 이상인 경우에는 비용민감방법의 성능이 좋은 경향이 나타났다.

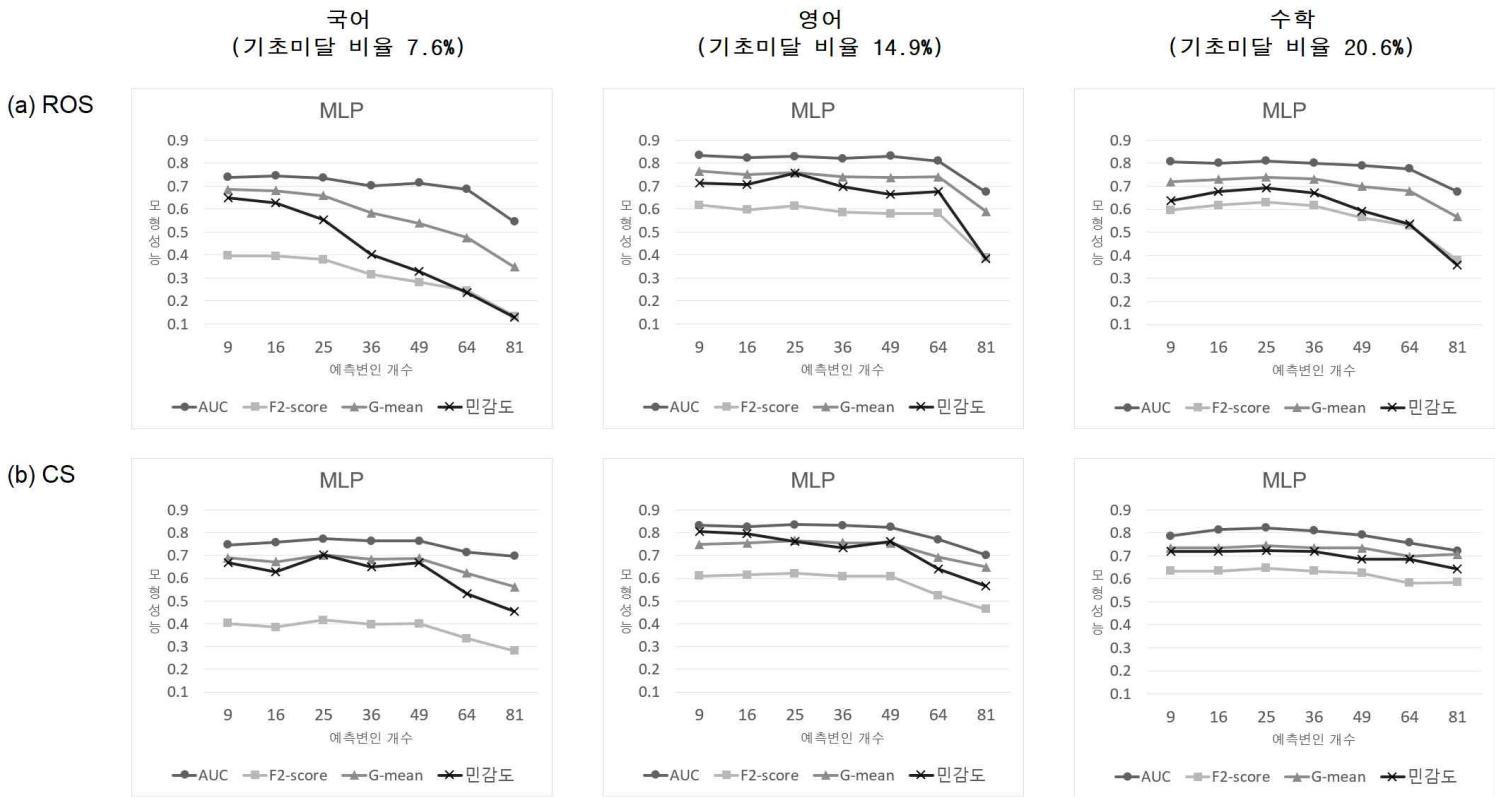
또한 표집방법 3가지 모형 성능을 비교한 결과, 표집방법에 따른 성능의 우위가 나타나진 않았으며, B-SMOTE와 ADASYN이 불균형 자료의 특징을 고려하여 소수집단 사례를 다수집단만큼 생성하는 방법임에도 ROS에 비해 우수한 성능을 보이진 않았다.

마지막으로 심층인공신경망 모형 종류에 따른 결과를 살펴보면, 모든 과목에서 CNN 모형에 비해 MLP 모형의 예측 성능이 더 좋게 나타났으나, 성능의 차이가 크다고 판단하기는 어려웠다. 하지만 CNN 모형의 경우에는 MLP 모형에 비해 소수집단 정보 반영방법에 관계없이 예측변인 개수가 많아져도 모형 성능이 덜 감소하는 모습이 나타났다.

<표 Ⅳ-22> 예측변인 개수 및 소수집단 정보 반영방법에 따른 과목별 기초미달 학생 예측모형 성능 (MLP)

모형	평가지표	국 어							영 어							수 학						
		예측변인의 개수							예측변인의 개수							예측변인의 개수						
		9	16	25	36	49	64	81	9	16	25	36	49	64	81	9	16	25	36	49	64	81
ROS	AUC	**0.739**	0.746	0.736	0.702	0.714	0.687	0.546	**0.835**	0.824	0.830	0.820	0.831	0.810	0.674	0.807	0.801	**0.811**	0.801	0.791	0.776	0.677
	F2-score	**0.398**	0.396	0.381	0.315	0.283	0.246	0.135	**0.617**	0.598	0.615	0.586	0.580	0.582	0.389	0.597	0.619	**0.631**	0.617	0.565	0.530	0.379
	G-mean	**0.685**	0.681	0.659	0.583	0.539	0.475	0.349	**0.765**	0.751	0.760	0.741	0.738	0.740	0.590	0.720	0.730	**0.739**	0.731	0.700	0.679	0.568
	정분류율	**0.717**	0.730	0.768	0.808	0.841	0.898	0.891	**0.805**	0.783	0.763	0.775	0.798	0.791	0.832	0.777	0.763	**0.768**	0.770	0.777	0.790	0.784
	민감도	**0.649**	0.628	0.553	0.404	0.330	0.237	0.128	**0.714**	0.708	0.757	0.697	0.665	0.676	0.384	0.637	0.678	**0.693**	0.670	0.593	0.537	0.359
	특이도	**0.723**	0.739	0.785	0.840	0.882	0.953	0.952	**0.820**	0.795	0.763	0.788	0.820	0.810	0.908	0.814	0.786	**0.788**	0.797	0.826	0.857	0.898
CS	AUC	0.747	0.758	**0.774**	0.764	0.764	0.714	0.698	0.833	0.826	**0.836**	0.832	0.825	0.770	0.703	0.787	0.814	**0.822**	0.810	0.791	0.758	0.722
	F2-score	0.402	0.386	**0.416**	0.397	0.400	0.335	0.281	0.610	0.616	**0.622**	0.609	0.609	0.524	0.465	0.635	0.635	**0.647**	0.635	0.625	0.583	0.585
	G-mean	0.689	0.673	**0.703**	0.684	0.688	0.623	0.562	0.749	0.756	**0.766**	0.757	0.755	0.693	0.649	0.734	0.734	**0.746**	0.734	0.734	0.698	0.706
	정분류율	0.706	0.716	**0.704**	0.715	0.704	0.715	0.676	0.713	0.730	**0.768**	0.773	0.749	0.735	0.716	0.743	0.743	**0.761**	0.743	0.765	0.705	0.747
	민감도	0.670	0.628	**0.702**	0.649	0.670	0.532	0.455	0.805	0.795	**0.762**	0.735	0.762	0.640	0.568	0.719	0.719	**0.722**	0.719	0.685	0.687	0.644
	특이도	0.709	0.723	**0.704**	0.720	0.707	0.729	0.694	0.697	0.720	**0.769**	0.780	0.747	0.751	0.742	0.750	0.750	**0.771**	0.750	0.786	0.709	0.774

주 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.

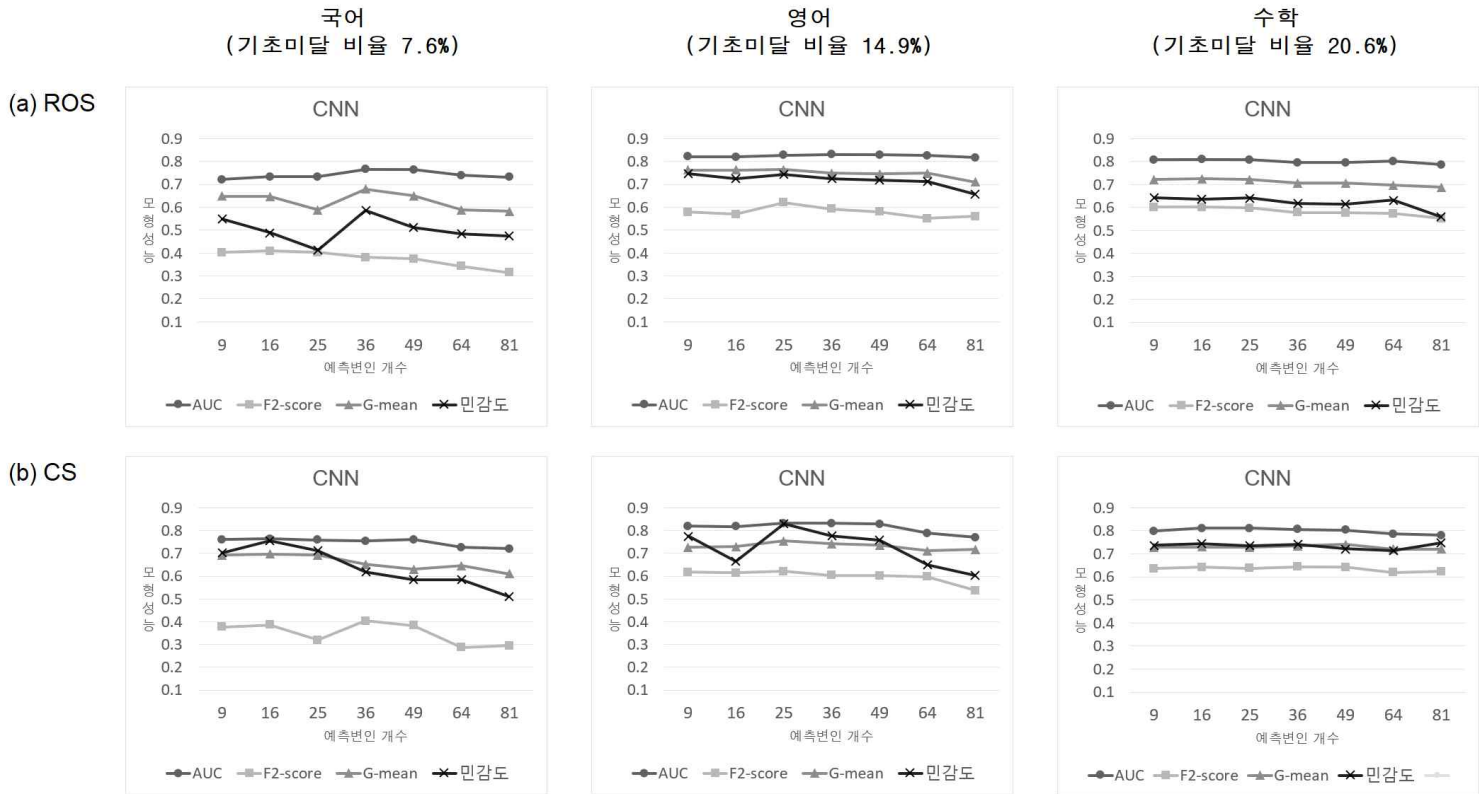


[그림 Ⅳ-16] 과목별 기초미달 학생 예측모형 성능 비교 (MLP)

<표 Ⅳ-23> 예측변인 개수 및 소수집단 정보 반영방법에 따른 과목별 기초미달 학생 예측모형 성능 (CNN)

모형	평가지표	국 어							영 어							수 학						
		예측변인의 개수							예측변인의 개수							예측변인의 개수						
		9	16	25	36	49	64	81	9	16	25	36	49	64	81	9	16	25	36	49	64	81
ROS	AUC	0.721	0.733	0.734	**0.766**	0.765	0.740	0.732	**0.822**	0.821	0.828	0.831	0.830	0.826	0.818	0.808	**0.810**	0.809	0.796	0.796	0.803	0.787
	F2-score	0.377	0.387	0.320	**0.404**	0.384	0.288	0.296	**0.618**	0.615	0.622	0.604	0.602	0.598	0.537	0.601	**0.602**	0.599	0.579	0.578	0.575	0.553
	G-mean	0.649	0.647	0.589	**0.680**	0.651	0.588	0.583	**0.763**	0.763	0.765	0.749	0.747	0.750	0.711	0.722	**0.725**	0.721	0.706	0.707	0.697	0.689
	정분류율	0.750	0.827	0.805	**0.774**	0.807	0.701	0.699	**0.780**	0.792	0.780	0.768	0.767	0.778	0.754	0.773	**0.786**	0.775	0.766	0.771	0.740	0.780
	민감도	0.550	0.489	0.415	**0.585**	0.511	0.483	0.474	**0.746**	0.724	0.745	0.724	0.720	0.714	0.657	0.644	**0.637**	0.641	0.619	0.615	0.632	0.561
	특이도	0.767	0.854	0.836	**0.789**	0.831	0.717	0.717	**0.785**	0.804	0.787	0.776	0.776	0.789	0.769	0.808	**0.826**	0.811	0.806	0.813	0.769	0.846
CS	AUC	0.761	**0.765**	0.760	0.755	0.762	0.728	0.721	0.821	0.819	**0.833**	0.832	0.830	0.789	0.771	0.800	0.812	0.812	0.807	0.804	0.788	0.781
	F2-score	0.403	**0.409**	0.404	0.381	0.375	0.343	0.315	0.580	0.571	**0.621**	0.593	0.581	0.551	0.561	0.637	0.642	0.638	0.644	0.642	0.620	0.625
	G-mean	0.692	**0.698**	0.693	0.652	0.631	0.646	0.610	0.727	0.731	**0.755**	0.743	0.736	0.712	0.718	0.728	0.730	0.728	0.735	0.741	0.722	0.722
	정분류율	0.684	**0.654**	0.677	0.681	0.673	0.703	0.712	0.695	0.784	**0.705**	0.719	0.720	0.759	0.814	0.723	0.723	0.724	0.732	0.752	0.726	0.707
	민감도	0.702	**0.755**	0.713	0.619	0.585	0.586	0.511	0.776	0.665	**0.832**	0.778	0.758	0.651	0.604	0.737	0.744	0.735	0.741	0.722	0.715	0.747
	특이도	0.683	**0.646**	0.674	0.642	0.681	0.712	0.727	0.681	0.804	**0.684**	0.709	0.714	0.778	0.853	0.720	0.717	0.721	0.730	0.759	0.729	0.697

주 동일한 모형에서 불균형 자료의 모형평가에 활용되는 AUC, F2-score, G-mean, 민감도 값이 가장 높은 모형에 음영표시 하였으며, McNemar-test 결과를 참고하여 최종적으로 모형 성능이 가장 좋다고 판단되는 모형에 볼드처리 하였다.



[그림 Ⅳ-17] 과목별 기초미달 학생 예측모형 성능 비교 (CNN)

최종적으로 과목별 모형 성능이 우수하게 나타난 예측모형과 혼동행렬은 <표 Ⅳ-24>와 같다. 모든 과목에서 예측변인 25개를 활용하고 비용민감방법을 적용한 MLP 모형의 모형 성능이 가장 좋았다. 각 과목의 민감도를 살펴보았을 때, 기초미달 학생 중 국어는 약 70%, 영어는 약 76%, 수학은 약 72%를 정확하게 예측하는 모형임을 확인할 수 있었다. 세 모형 모두 기저모형인 Original 모형에 비해 모형 성능이 향상된 모습이었으나, 전체 정확도가 70~76% 정도이고 모형에서 예측한 기초미달 학생 중 기초이상 학생 비율이 상대적으로 많다는 점에서, 실제 교육현장에서 기초미달 학생 예측에 활용하기 위해서는 모형 성능이 더 향상될 필요가 있었다.

<표 Ⅳ-24> 과목별 최종모형의 모형 성능과 혼동행렬

과 목	모형지표	값			예측	
					기초이상(N)	기초미달(P)
국어 25MLPCS	AUC	0.774				
	F2-score	0.416	실제	기초이상(N)	835	351
	G-mean	0.703				
	정분류율	0.704		기초미달(P)	28	66
	민감도	0.702				
	특이도	0.704				
영어 25MLPCS	AUC	0.836			기초이상(N)	기초미달(P)
	F2-score	0.622	실제	기초이상(N)	842	253
	G-mean	0.766				
	정분류율	0.768		기초미달(P)	44	141
	민감도	0.762				
	특이도	0.769				
수학 25MLPCS	AUC	0.822			기초이상(N)	기초미달(P)
	F2-score	0.647	실제	기초이상(N)	779	231
	G-mean	0.746				
	정분류율	0.761		기초미달(P)	75	195
	민감도	0.722				
	특이도	0.771				

V. 결론

1. 요약

기초학력 미달(이하 기초미달) 학생의 학습 결손은 누적되고 심화되는 특성이 있음을 고려하여, 이 연구에서는 향후 기초미달이 될 학생을 예측하는 심층인공신경망 모형을 구축해보고자 하였다. 이때 기초미달 학생이 소수집단임을 고려하여 이를 조정해 줄 수 있는 방법과 예측변인의 개수 및 심층인공신경망 종류에 따른 성능 결과를 살펴봄으로써 보다 나은 예측모형을 구축하는 방안을 모색해보고자 하였다.

심층인공신경망이 많은 수의 변인을 한 모형에 활용할 수 있는 방법이긴 하나, 학업성취와 관련이 없는 변인의 활용은 오히려 예측모형의 성능을 떨어뜨릴 수 있다는 논의에 따라, 심층인공신경망 예측모형을 상세화하기기에 앞서, 과목별 학업성취도와 관련이 높은 설명변인들을 확인하였다. 이 결과를 바탕으로 예측변인의 개수를 늘려가며 모형 성능을 확인하였다. 이후 주요 설명변인을 예측변인으로 활용하여 다층퍼셉트론 모형과 합성곱신경망 모형으로 예측모형을 상세화하였으며, 이때 불균형 자료 문제를 완화할 수 있는 표집방법(임의과대표집, B-SMOTE, ADASYN)과 비용민감방법을 예측모형에 적용하였다. 최종적으로 과목별 모형 성능이 우수한 예측모형을 최종모형으로 선정하였다.

분석에는 별도의 기초학력 검사를 통해 학업성취 수준을 측정하고, 학생들의 인지적・정의적 발달 상황과 가정과 학교에서의 교육경험 정보를 포괄적으로 수집한 한국교육종단연구2013를 활용하였다. 구체적으로 중학교 2학년 정보를 활용하여 중학교 3학년 시기의 국어, 영어, 수학 기초미달 여부를 예측하는 모형을 구현하고자 하였으며, 각 과목의 기초미달 비율은 국어 7.6%, 영어 14.9%, 수학 20.6%로 전체 학생 대비 적은 비율을 보이고 있었다. 이상의 연구를 수행하기 위하여 다음과 같은 연

구문제를 상정하였다.

연구문제 1. 기초학력 미달 학생을 예측하기 위한 심층인공신경망 모형에서 예측변인으로 활용가능한 주요 변인은 무엇인가?

연구문제 2. 기초학력 미달 학생을 예측하기 위한 심층인공신경망 모형은 무엇이며, 그 성능은 어떠한가?

2-1. 예측변인의 수에 따라 예측모형의 성능은 어떻게 나타나는가.

2-2. 소수집단에 속하는 기초학력 미달 학생의 정보를 훈련 모형에 반영하는 방법에 따라 예측모형의 성능은 어떻게 나타나는가.

2-3. 심층인공신경망 종류에 따라 예측모형의 성능은 어떻게 나타나는가.

연구문제에 따라 연구결과를 요약하면 다음과 같다.

먼저, 심층인공신경망 예측모형을 상세화하기에 앞서, 다수의 변인 중 종속변인과 관련 있는 변인을 선택하는 동시에 각 변인의 계수 추정이 가능한 LASSO 분석 방법을 활용하여 주요 설명변인을 확인하였다. 학생, 학부모, 학교 361개 변인 중 국어에서는 91개, 영어에서는 84개, 수학에서는 83개 변인이 각 과목의 학업성취도와 관련이 있는 설명변인으로 나타났다. 하지만 선정된 변인의 계수를 확인한 결과 계수의 절대값이 0에 가까워 종속변인과의 관련성이 상당히 낮은 변인들이 다수 포함되어 있었다. 이에 따라 학업성취도와 관련 높은 변인을 중심으로 예측변인의 수를 늘려가며 예측모형을 구축해보고자 하였다.

각 과목에서 선정된 주요 설명변인 중 상위 30개 변인을 살펴보면, 공통적으로 수업이해도 및 수업태도 관련 변인과 진로 관련 변인이 주요 상위 설명변인으로 나타났다. 과목별 주요 설명 변인으로는 영어와 수학에서는 선행학습 정도, 학원강의 및 과외 참여 관련 변인, 진학계획 등이 주요 설명변인으로 나타났고, 국어에서는 독서활동 관련 변인이 다수 확인되었다.

다음으로 국어, 영어, 수학 과목의 기초미달 학생을 예측하기 위하여, 예측변인의 개수, 소수집단 정보 반영방법, 심층인공신경망 종류에 따라 심층인공신경망 예측모형을 상세화하고 그 성능을 살펴본 결과는 다음과 같다. 첫째, 모든 과목에서 예측변인의 개수가 일정 수준 이상 늘어나면 예측모형의 성능이 낮아지는 경향이 나타났으며, 국어에서는 상대적으로 성능 감소가 큰 모습을 보였다. 전반적으로 예측변인이 9~25개 정도인 경우에는 모형 성능이 비슷한 수준이거나 오히려 예측변인이 많아짐에 따라 성능이 향상되는 모습이 나타나기도 하였으나, 36~81개로 예측에 활용하는 변인이 늘어나면 모형 성능은 점차 감소하는 모습을 보였다.

둘째, 모든 과목에서 불균형 정도를 조정하지 않은 Original 모형에서는 대부분의 학생을 기초이상 학생으로 예측하는 모습이 나타났다. 불균형 정도가 상대적으로 낮은 수학에서는 Original 모형에서도 민감도가 MLP 0.456, CNN 0.537로 나타나기도 하였으나, 불균형 정도가 높았던 국어에서는 민감도가 MLP 0.096, CNN 0.064으로 거의 모든 학생을 기초이상으로 예측하는 모습이 나타났다. 상대적으로 모형 훈련 시 소수집단 정보를 다수집단만큼 반영해준 경우에는 민감도와 함께, AUC, F2-score 및 기초미달과 기초이상 예측 비율 간 균형 정도를 보여주는 G-mean 값도 상승하는 경향을 보였다. 소수집단 정보 반영방법 간 성능을 확인하였을 때에는 표집방법에 비해 비용민감방법을 적용한 경우의 모형 성능이 전반적으로 더 좋게 나타났다. 하지만 예측변인의 개수가 9~25개 정도일 때에는 소수집단 정보 반영방법에 따른 차이가 크지 않은 반면, 예측변인이 49개 이상으로 커짐에 따라 차이가 커지는 모습을 확인할 수 있었다.

셋째, 심층인공신경망 종류에 따른 성능은 모든 과목에서 CNN 모형에 비해 MLP 모형의 예측 성능이 더 좋게 나타났다. 하지만 그 차이가 크다고 판단하긴 어려웠으며, CNN 모형에서는 소수집단 정보 반영방법에 관계없이 예측변인 개수에 따른 모형 성능 감소폭이 적은 경향이 나타나기도 하였다.

최종적으로는 모든 과목에서 예측변인 25개를 활용하고, 비용민감방

법을 적용한 MLP 모형의 성능이 가장 좋게 나타났다. 과목별 최종모형의 성능은 AUC 0.774~0.836, F2-score 0.416~0.647, G-mean 0.703~0.766, 민감도 0.703~0.762으로 나타났다. 하지만 이는 전체 학생을 제대로 예측한 비율이 70~79%일 때, 기초미달 학생 중 70~76% 정도를 제대로 예측하는 모형으로, 예측모형을 실제 활용할 수 있을 정도의 예측 성능이라 판단하기는 어려웠다.

2. 논의 및 시사점

연구의 결과를 바탕으로 논의점을 정리하면 다음과 같다.

첫째, 예측에 활용하는 변인은 설명력이 높은 변인을 위주로 모형을 구축할 필요가 있다. 예측변인 개수에 따른 예측 성능 비교 결과를 살펴보면, 모든 과목에서 예측변인의 개수가 일정 수준 이상 늘어남에 따라 예측 성능이 감소하는 모습이 나타났으며, 25개 예측변인을 활용하는 경우에 모형 성능이 가장 높았다. 예측변인 개수가 많아진다는 것은 활용할 수 있는 정보가 많아지는 것을 의미하지만, 종속변인과 관계 없는 정보가 많은 경우에는 오히려 자료의 패턴을 파악하기 어려울 수 있음을 보여주고 있는 부분이다. 이는 심층인공신경망을 포함한 기계학습 모형에서 종속변인과 관계 없는 정보가 많은 경우에는 오히려 모형 성능이 감소한다는 연구 결과와 일치하는 부분이기도 하다(García et al., 2016; Snieder et al., 2019).

또한 연구에서 활용한 교육패널자료는 학생의 다양한 영역에 대한 정보를 포괄적으로 수집하기 때문에 학업성취와 직접적인 관련이 없을 수 있는 정서적, 환경적 변인들이 다수 포함되어 있는 자료이다. 아울러 정서적 영역의 경우에는 자기보고식으로 응답하기 때문에 학업성취 수준에 따른 특징이 분명하게 구분되지 않을 수 있다. 상대적으로 중간·기말고사 개별 문항의 응답 정보는 응답 값이 객관적인 기준으로 구분되어 있

고, 각 문항의 정보와 학업성취와의 관련성이 높은 자료일 것이라 추측해볼 수 있다. 심층인공신경망 모형은 변인의 수와 종류를 활용하는 데 유연하다는 특성이 있어, 활용할 수 있는 변인을 최대한으로 모형에 투입할 수 있다. 하지만 패널자료의 특성을 고려하였을 때, 예측에 활용할 예측변인을 선정하는 것 또한 예측모형 구축 단계의 중요한 단계임을 인지하여, 예측모형을 구축할 때에 변인 선정부터 심도 있는 고민이 필요할 것이다.

둘째, 소수집단의 예측 과정에서 모형 훈련 시 소수집단의 정보를 반영할 때에는 비용민감방법을 적용해볼 수 있다. 중학교 학생들의 국어, 영어, 수학 기초미달 예측모형 성능을 확인해 본 결과, 모든 과목에서 표집방법에 비해 비용민감방법이 기초미달 학생 예측 성능이 좋은 경향을 보였다. 이러한 경향은 불균형 정도가 상대적으로 높은 국어에서 더 뚜렷하게 나타났다. 과목 간 차이를 기초미달 비율로만 설명할 순 없으나, 선행연구를 살펴보면 특히 불균형 정도가 높은 경우 표집방법으로는 소수집단의 정보를 반영하는 것이 한계가 있으며, 손실함수 조정을 통해 알고리즘 자체를 조정해주는 것이 필요하다는 제언을 확인할 수 있었다(유병주, 2021; Mienye & Sun, 2021). 이는 표집방법과 같이 사례를 복제하거나 유사한 사례를 만듦으로써 소수집단 수를 늘려주는 방법은 다소 간접적으로 정보를 반영해주는 반면, 손실함수의 가중치를 조정하는 방법은 상대적으로 모형에 직접적인 조정을 가해주는 것이기 때문일 수 있다(Zhang et al., 2011).

하지만 예측변인의 개수가 9~25개 정도인 경우에는 표집방법과 비용민감방법의 차이가 크지 않았고, 예측변인이 49개 이상으로 커짐에 따라 두 방법 간 차이가 커지는 모습이 나타난 점에 비추어볼 때, 집단 간 특징 차이를 잘 드러낼 수 있는 예측변인을 선정하는 것이 더 중요할 수 있음을 추측해 볼 수 있다.

셋째, 교육패널자료에는 MLP 모형의 활용을 고려해 볼 수 있다. 이 연구에서는 CNN 모형이 자료의 핵심적인 특징을 잘 포착해내며, 불균형 문제에서 좋은 성능을 보였다는 점에서(Yan et al., 2015), MLP 모형과

함께 예측모형에 활용하고 그 결과를 비교하였다. 과목별 일부 모형에서는 CNN 모형이 MLP 모형에 비해 좋은 성능을 보이기도 하였으나, 패널자료를 CNN에 모형에 적용하기 위해서는 자료를 인위적으로 이미지 자료와 같이 변환하여야 하고, 그에 따라 예측에 활용할 수 있는 변인의 수도 제한적인 면이 있었다. MLP 모형은 연구에서 활용한 기본적인 형태를 넘어서 은닉층의 노드를 매우 적게 줄인 후 다시 늘리는 등의 과정을 통해 자료에서 중요하지 않은 정보를 제외하는 등 다양하게 모형 구성이 가능하다. 아울러 교육패널자료와 같은 정형데이터는 이미지, 음성, 문자 자료에 비해 변인 간 상관이 상대적으로 적다는 특징이 있다는 점을 고려할 때(Sankhe et al., 2021), 각 변인을 독립적인 정보로 입력해주는 MLP 모형을 적용하는 것이 보다 적절할 수 있다.

이 연구가 지니는 시사점은 다음과 같다.

첫째, 패널자료의 특징을 고려하여 심층인공신경망 모형의 적용을 탐색하였다는 점에서 의의가 있다. 심층인공신경망 모형은 이미지, 음성, 문자 자료를 주로 활용하는 경향이 있다. 그러나 교육에서는 인간을 대상으로 하는 연구이기 때문에, 개인정보보호 등의 이유로 학생들의 서술형 평가 정보, 개인 상담자료, 수업 중 발화 녹음 자료 등과 같은 자료 수집이 어려운 편이며, 이를 시・도 또는 전국 단위로 수집하는 것은 더욱 쉽지 않다. 따라서 심층인공신경망 모형을 각 시도에서 이미 수년 간 축적된 패널자료에 적용해봄으로써, 전국 단위, 시・도 단위 연구 및 학교급에 따른 연구도 가능하다.

이에 더해, 심층인공신경망은 활용 가능한 데이터가 다양하고 종류가 다른 자료를 조합하여 분석에 활용할 수 있는 분석 방법이다. 하만석, 안현철(2019)의 연구에서는 패널자료와 유사한 정형데이터와 강사의 상담일지와 같은 비정형데이터를 동시에 예측변인으로 활용하여 직업훈련 중도 탈락을 예측하기도 하였다. 연구 결과, 정형데이터만 활용한 모형에 비해 비정형 데이터를 함께 고려했을 때 예측의 정확도가 20%까지 향상되는 모습을 확인하였다. 이처럼 교육패널자료를 활용한 심층인공신

경망 모형을 기반으로 향후 교육자료의 활용 폭이 넓어지면 활용 범위를 넓어질 것이라 기대해볼 수도 있을 것이다.

둘째, 개별 학생 예측이라는 관점에서 소수집단에 속하는 학생을 예측하는 심층인공신경망 모형 적용해보고 모형 설정에 따른 결과를 상세하게 제시하였다는 점에서 의의가 있다. 기초미달 학생들을 살펴본 선행연구에서는 문제가 발생하기 이전 또는 발생 초기에 확인을 할 수 있는 시스템이 필요함을 지속적으로 이야기하였다(김태은 외, 2018; 이화진 외, 2009). 또한 기초미달 학생을 비롯하여 학교 중도 탈락, 심각한 정서적 문제 행동 등 예방적이고 개별적인 지원이 필요한 학생들은 소수집단인 경우가 많다는 점을 고려할 때 이분형 종속변인이 높은 불균형을 보이는 상황을 다룬 연구가 필요하다. 교육자료를 활용하여 심층인공신경망 모형을 적용하는 연구는 상대적으로 근래에 시작되었다고 할 수 있으며(박미현, 허균, 2021; 이창묵 외, 2020), 그중에서도 소수집단 예측 연구는 찾아보기 어려웠다. 따라서 이 연구에서 다룬 모형 설정에 대한 논의 및 적용 과정과 그 결과는 향후 예측 연구를 위한 기초자료의 역할을 할 수 있을 것으로 기대한다.

이상의 논의 및 한계점을 바탕으로 후속 연구에 대한 제언은 다음과 같다.

첫째, 다양한 교육패널자료와 여러 코호트를 대상으로 심층인공신경망 모형 적용 가능성을 살펴보는 연구가 필요하다. 이 연구에서는 과목별 기초미달 예측모형을 구축하고 모형 상세화에 따른 예측 성능 경향성이 유사하게 나타나는지 살펴보았으나, 하나의 자료만을 활용하였다는 한계가 있다. 특히 심층인공신경망 모형은 주어진 자료의 특징을 반복적으로 학습하여 모형을 구축해나가는 모형이기 때문에 교육자료의 변인 구성, 코호트의 특징에 따라 결과가 상이하게 나타날 가능성이 높다고 할 수 있다. 따라서 모형 설정에 따른 결과가 자료의 특성에 따라 일정한 경향성을 보이는 지 확인하는 것은 연구 결과의 일반화를 위하여 필수적인 부분이라 할 수 있다.

둘째, 적절한 예측변인 선정을 위한 연구가 요구된다. 이 연구에서는 여러 변인선택 방법 중 종속변인 및 설명변인의 종류 및 특징에 따른 변화가 크지 않으며(Fu et al., 2019), 심층인공신경망 예측모형에서 좋은 결과(Zhang et al. 2019)는 점에서 LASSO 분석방법을 적용하였다. 하지만 변인선택 연구를 살펴보면, 변인선택 방법에 따라 중요하게 나타나는 변인들이 다르게 선택되기도 하며, 그에 따라 모형 성능에도 영향을 미치는 모습이 나타났다(Hamer & Dupont, 2021). 교육패널자료는 이분변수, 서열변수, 연속변수 등 다양한 종류의 변인으로 구성되어 있고, 학생 및 학교 수준의 다층구조로 되어 있다는 특징이 있다. 따라서 변인선택 방법에 따라 주요 변인이 어떻게 달라지며, 예측모형에 미치는 영향이 얼마나 되는 지를 살펴볼 필요가 있다.

더 나아가 기초미달 학생을 비롯하여 개별적, 선제적 지원이 필요한 학생들은 소수집단인 경우가 많다는 점을 고려할 때, 불균형 정도를 고려한 예측변인 선택에 대한 연구도 함께 이루어져야 할 것이다. 이 연구에서도 중2 기초미달 여부와 관련이 높은 설명변인을 찾는 과정에서 불균형 자료로 인한 문제가 발생하여, 대안적 방법으로 중2 학업성취도 점수와 관련 높은 변인을 예측변인으로 선정하였다. 하지만 기초미달 학생이 기초학력 수준에도 못 미치는 학생이라는 점을 고려할 때, 전체 학생의 학업성취도를 높이고 낮추는 데 영향을 미치는 변인과 기초미달 여부를 구분하는 데 영향을 미치는 변인은 상이할 수 있으며, 기초미달 여부에 보다 적절한 변인을 선정할 수 있었다면 모형 성능 또한 더 향상되었을 수 있다.

마지막으로 기초미달 학생 예측에 적합한 자료에 대한 고민이 필요하다. 이 연구에서는 국가 및 시·도 수준에서 체계적으로 수집된 교육패널자료의 활용 가능성을 탐색하였으나, 예측 성능이 우수하지 않은 모습을 보였다. 이는 학업성취 수준 예측을 위해서는 실질적으로 학업성취와 관련이 높은 자료를 구축하고, 이를 활용하는 방안을 모색하는 것이 더 적절할 수 있음을 의미할 수 있다. 교육패널자료에도 학업성취도 정보가 있으나 과목별 총점으로 제시되어 있어, 학생들의 성취 정도를 상세하게 파악하기에는 어

려운 면이 있다. 또한 인지적, 정의적 영역의 응답 정보들은 자기보고식 응답이기 때문에 개별 학생마다 응답하는 기준이 다를 수 있다. 이는 집단 간 구분을 불명확하게 하여 예측에 어려움을 초래할 수도 있는 부분이기도 하다. 또한 심층인공신경망 모형이 분석자료에 종속적인 특징이 있다는 점에서도 심층인공신경망 모형 설정 자체에 대한 연구와 함께 활용하는 자료의 질을 높이기 위한 연구 또한 함께 이루어져야 할 것이다.

참 고 문 헌

강은영, 김진구(2018). 성취집단 수준별 전년도 학업성취와 학습풍토 및 학교풍토가 학업성취에 미치는 효과의 차별적 특징 분석. 학습장애연구, 15(1), 93-112. DOI: 10.47635/KJLD.2018.15.1.93

곽현주, 백종면, 양예슬, 송인발(2016). 학교 학업성취수준에 영향을 미치는 학교특성 요인: 학업고성취 및 저성취수준에 미친 요인비교를 중심으로. 교육과정평가연구, 19(1), 333-358.

교육과학기술부(2009). 학력향상 중점학교 지원 기본계획.

교육부(2015). 2015년 초・중・고 사교육비조사 결과 발표.

교육부(2015). 기본교육과정. 교육부고시 제2015-81호 별책 3.

교육부(2016). 2017년 기초학력 향상 지원 사업 기본 계획.

교육부(2021). 「교육회복 종합방안」 기본계획 발표.

권재기(2013). 일반계 고등학교의 기초학력미달 변화에 따른 학교유형 분류 및 종단적 변화. 교육평가연구, 26(5), 1037-1066.

길혜지, 윤지윤(2015). 고등학생 유형 결정요인 탐색: 진로성숙도와 성취도를 중심으로. 직업교육연구, 24(6). 73-91.

김경근, 성열관, 김정숙(2007). 학력부진 아동의 특징 및 발생 원인에 대한 면담 연구. 교육사회학연구, 17(3), 27-52. DOI: 10.32465/ksocio. 2007.17.3.002

김경희, 임은영, 신진아(2013). 학업성취도 평가 결과에 나타난 초・중학생의 정의적 특성에 대한 종단분석과 예측. 교육평가연구, 26(5), 981-1014.

김기태, 이보미, 김종우(2017). 이진 분류문제에서의 딥러닝 알고리즘의 활용 가능성 평가. 지능정보연구, 23(1), 95-108.

김대헌, 김현주(2003). 입학초기 중학생의 삶. 교육과정연구, 21(3), 1-24.

김동일(2001). 학습부진아 교육체계 발전방안 연구. 청소년상담연구, 9(1), 226-250.

김동일, 홍성두(2009). 학습장애 진단을 위한 데이터 마이닝 기법간 비교: 회귀분석, 의사결정나무분석, 신경망분석을 중심으로. 특수교육연구, 16(1), 321-339.

김선, 김경옥, 김수동, 이신동, 임혜숙, 한순미(2001). 학습부진아의 이해와 교육. 서울: 학지사.

김성숙, 송미영, 최인봉, 김희경, 김성훈, 박서홍, 김진화(2010). 2009년 국가수준 학업성취도 평가 전수 분석 결과-고등학교 1학년(연구보고 RRE 2010-7-4). 한국교육과정평가원.

김성훈, 김우진, 장연주, 김현철(2021). 설명 가능한 AI 학습 지원 시스템 개발. 컴퓨터교육학회 논문지, 24(1), 107-115. DOI : 10.32431/kace.2021.24.1.012

김양분, 남궁지영, 박경호, 최인희, 강호수, 김미숙, 이규민, 최보미, 이영주, 송승원(2016). 2016 한국교육종단연구(KELS) 2013(Ⅳ): 조사개요보고서(기술보고 TR 2016-55-01). 한국교육개발원.

김의중(2016). 인공지능, 머신러닝, 딥러닝 입문. 경기: 위키북스.

김정아(2021). 초・ 중학생 기초학력 미달의 종단적 변화 유형과 영향요인 분석. 교육학연구, 59(4), 131-161.

김지은, 김지하(2012). 학력향상 중점학교 사업의 기초학력 향상 효과 분석. 초등교육연구, 25(2), 117-139.

김태은, 오상철, 우연경, 권서경(2018). 초・중학교 학습부진 학생의 성장과정에 대한 연구(Ⅱ)(연구보고 RRI 2018-4). 한국교육과정평가원.

김태은, 오상철, 우연경, 권서경, 김영빈, 서덕희(2017). 초・중학교 학습부진 학생의 성장과정에 대한 연구(Ⅰ)(연구보고 RRI 2017-6). 한국교육과정평가원.

김태은, 이화진, 오상철, 노원경(2012). 초등학생의 기초학력 미달 여부에 영향을 미치는 변인들의 효과 분석. 교육심리연구, 26(2), 505-521.

김한용, 이우주 (2017). 불균형적인 이항 자료 분석을 위한 샘플링 알고리즘들: 성능비교 및 주의점. 응용통계연구, 30(5), 681-690.

김희경, 김완수, 김수진, 정혜경, 김미림, 김성훈(2019). 국가수준 학업성취도 평가 점수 체제 개선 및 결과 활용도 제고 방안(연구보고 RRE 2019-3). 한국교육과정평가원.

남궁지영, 김혜자, 박경호, 송승원, 김양분, 이규민(2018). 한국교육종단연구(KELS)2013(Ⅵ): 조사개요보고서(기술보고 TR 2018-137). 한국교육개발원.

남나라(2021). 혼합효과 랜덤 포레스트 기법을 활용한 중학생의 학업성취도 예측 요인 분석. 서울대학교 박사학위논문.

남창우, 전훈(2012). 교사와 학생의 ICT 활용과 학생의 자기주도학습이 저성취 고등학생의 학업성취도에 미치는 영향. 한국교원교육연구, 29(1), 77-97. DOI: 10.24211/tjkte.2012.29.1.77

박경호, 권희경, 김정아, 남궁지영, 최인희, 송승원, 김위정(2020). 2020 한국교육종단연구: 고등학생의 교육 경험과 성장(Ⅱ)(연구보고 RR 2020-27). 한국교육개발원.

박미현, 허균(2021). 머신러닝 분류기법을 적용한 중학생의 학습 부진 예측모형 개발연구: 대구교육종단자료를 중심으로. 교육공학연구, 37(3), 627-648. DOI: 10.17232/KSET.37.3.627

방성완, 김재오(2020). 불균형 자료의 분류분석을 위한 가우스 혼합 군집모형을 이용한 샘플링 기법. Journal of The Korean Data Analysis Society (JKDAS), 22, 565-574.

백병부(2010). 중학교 단계에서의 학습부진 결정 요인 분석. 한국교육, 37(4), 73-102. DOI: 10.22804/jke.2010.37.4.004

백병부(2011). 수준별 하반 편성과 특별보충수업이 백병부중학교 학습부진학생들의 교과목 효능감에 끼치는 영향. 한국교육학연구, 17(2), 47-72.

손윤희, 박현정, 박민호(2020). 랜덤포레스트를 활용한 읽기소양 수준에 따른 집단 결정 요인 분석: PISA2018 데이터를 중심으로. 아시아교육연구, 21(1), 191-215. DOI: 10.15753/aje.2020.03.21.1.191

송연주, 이상수, 김대현, 함경애, 이동훈(2014). 학업성취 수준과 가정의

경제적 수준에 따른 우울, 불안, 스트레스의 차이. 교육종합연구, 12(1), 29-50.

시기자, 박인용, 구남욱, 김완수, 구슬기, 임현정, 김준엽(2015). 국가수준 학업성취도 평가 종단자료에 기반한 학업성취도 변화 추이 분석 (연구보고 RRE 2015-1). 한국교육과정평가원.

신종호(2002). 저성취 학습부진아동을 위한 학습전략프로그램에 대한 분석적 고찰. 아시아교육연구, 3(1), 63-88.

신종호, 최재원(2019). 학습분석 기반 대학 신입생 대상 학습부진 위험학생 조기예측 모델 개발 및 군집별 특성 분석. 한국교육공학학회, 35(2), 425-454. DOI: 10.17232/KSET.35.2.425

안영은(2019). 고등학생의 최소 기준학력 도달에 미치는 학교수준 영향요인 분석: 학습조직풍토 요인을 중심으로. 교육행정학연구, 37(4), 86-110. DOI: 10.22553/keas.2019.37.4.87

오일석(2017). 기계학습. 서울: 한빛아카데미.

오지영, 이수정(2008). 신경망을 이용한 초등학생 컴퓨터 활용 능력 예측. 정보교육학회논문지, 12(3), 267-274.

오창석(2018). 딥러닝을 위한 인공신경망. 서울: 내하출판사.

유미경, 엄준용(2016). 중학교 기초학력 성취도 차이에 미치는 영향 요인 분석. 한국교원교육연구, 33(2), 97-126. DOI: 10.24211/tjkte.2016.33.2.97

유병주(2021). 불균형 자료의 분류분석 방법별 성능 비교와 접근 전략 연구. 한국자료분석학회, 23(1), 195-207. DOI : 10.37727/jkdas.2021.23.1.195

유진은(2016). 데이터 마이닝 기법을 통한 교육 패널데이터 분석: 벌점회귀모형과 KYPS 자료. 아시아교육연구, 17(3), 1-19.

유진은(2019). 기계학습: 대용량/패널자료와 학습분석학 자료 분석으로의 활용. 교육공학연구, 35(2), 313-338.

이규민, 김경성, 반재천, 권정아, 박성희, 정송희, 성춘영, 정종식, 최형권, 정미순, 이정은, 박지혜(2015). 한국교육종단연구2013(Ⅲ): 중학교

학업성취도검사 개발 및 수직 연계 연구(기술보고 TR 2015-96). 한국교육개발원.

이규민, 반재천, 곽춘옥, 남진, 박동훈, 박승용, 박하영, 신혜진, 이천희, 이형래, 조두경(2013). 초등학교 기초능력검사 도구 개발 및 수직 연계 연구(기술보고 TR 2013-85). 한국교육개발원.

이대식(2015). 학습부진학생의 기초학력 향상을 위한 정부 지원 사업의 특징과 발전 방향. 학습장애연구, 12(3), 101-132

이대식, 임건순(2018). 학습부진 학생 지원 실태에 대한 초등교사들의 인식. 교육문화연구, 2(4), 151-170. DOI : 10.24159/joec.2018.24.4.151

이대식, 황매향(2011). 학습부진학생의 이해와 지도. 파주: 교육과학사.

이성은, 천세영, 이상호(2012). 방과후학교 지원비의 기초학력향상 효과 분석. 교육종합연구, 10(3), 211-233.

이종회(2019). 불균형 데이터 환경에서 가계부채 상환연체 분류를 위한 머신러닝의 활용. 소비자학연구, 30(6), 97-118. DOI: 10.35736/JCS.30.6.5

이창묵, 이현숙, 나우열(2020). 순환신경망을 활용한 개별 학생 학업성취도의 종단적 예측. 교육평가연구, 33(1), 161-189.

이현진(2018). 오토인코더에 기반한 딥러닝을 이용한 사이버대학교 학생의 학업 성취도 예측 분석 시스템 연구. 한국디지털콘텐츠학회, 19(6), 1115-1121. DOI: 10.9728/dcs.2018.19.6.1115

이현철, 정시영, 최성보(2015). 초・중학생의 교사관계 인식의 종단적 변화 양상과 예측 요인. 교육문화연구, 21(3), 57-82.

이화진, 김민정, 이대식, 손승현(2009). 학습부진학생 지도·지원의 실효성 제고를 위한 대안 탐색: 학습부진학생 지도・지원 종합 계획(안) 제안을 중심으로(연구보고 RRC 2009-13). 한국교육과정평가원.

이화진, 김태은, 오상철, 김성열, 박소영(2011). 기초학력 미달 여부에 영향을 미치는 요인 분석-학력향상 중점학교 유형별 비교. 교육과정평가연구, 14(1), 151-176.

이효정, 김광주, 김성훈(2018). 중학교 기초학력 미달률에 미치는 학교환

경과 교장공모제의 효과. 교육과정평가연구, 12(1), 173-195. DOI: 10.22799/jce.2018.21.1.008

임현정(2009). 학습부진에 영향을 미치는 학생과 학교의 교육맥락적 요인 탐색(현안보고 OR 2009-3-7). 한국교육개발원.

장선희, 정제영(2018). 학교 교육 활동이 일반고등학교의 기초학력 미달 비율에 미치는 영향 분석. 중등교육연구, 66(2), 551-573. DOI: 10.25152/ser.2018.66.2.551

전명남(2017). 중학교 기초학력미달 학생으로 살아가기에 대한 현상학적 연구. 복지상담교육연구, 6(1), 1-33. DOI: 10.20497/jwce.2017.6.1.1

전명남, 양명희 (2015). 중학교 기초학력미달 학생의 학습실태와 요구 조사. 교육학논총, 36(2), 21-39.

정혜경(2013a). 방과후학교 프로그램이 '기초학력미달' 고등학교 학생의 학업성취도에 미치는 효과 연구. 한국교육, 40(3), 46-69. DOI: 10.22804/jke.2013.40.3.003

정혜경(2013b). 학습부진학생 지도를 위한 전담교사 배치 정책 효과 연구. 교육학연구, 51(1), 55-75.

정혜원, 김아름, 백예은(2022). 혼합효과 랜덤 포레스트를 적용한 중학생의 학업성취도 설명 변인 탐색. 교육평가연구, 35(1), 107-136.

조난심, 권점례, 김도남, 서근원, 이신동, 이혜영(2009). 기초학력 증진을 위한 정책개발 기초 연구(연구보고 RRI 2009-2). 한국교육과정평가원.

천세영, 이성은(2013). 교육지원청 수준에서 기초학력 영향변인 연구. 교육학연구, 15(2), 37-62.

최재식(2019). 설명가능 인공지능 연구동향. 정보과학회지, 37(7), 8-14. DOI: 10.22733/JITAE.2022.12.01.005

최효식, 신종호, 민지연, 김남희(2013). 중・고등학생의 학업성취도, 부호화 및 인출 전략 간의 관계 분석. 아시아교육연구, 14(1), 331-352. DOI: 10.15753/aje.2013.14.1.013

최희열, 민윤홍(2015). 특집명 01: 지능형 정보 시스템: 딥러닝 소개 및

주요 이슈. 정보처리학회, 22(1), 7-21.
하만석, 안현철(2019). 정형 데이터와 비정형 데이터를 동시에 고려하는 기계학습 기반의 직업훈련 중도탈락 예측 모형. 한국콘텐츠학회 논문지, 19(1), 1-15. DOI : 10.5392/JKCA.2019.19.01.001
한국교육평가학회(1998). 교육평가 용어사전. 학지사.
한광희, 류용준, 김태순, 허준행(2010). 상호정보량 기법을 적용한 인공신경망 입력자료의 선정. 한국수자원학회, 43(1), 81-94.
허유성, 김민성, 박주성(2014). 중학교 학습장애 위험군 학생의 정서·행동 문제 수준과 학교환경변인의 영향. 특수아동교육연구, 16(4), 21-242. DOI: 10.21075/kacsn.2014.16.4.215
허유성, 남창우(2009). 학습부진 프로그램, 수준별 이동수업 및 사교육이 저성취 중학생의 학업성취도와 자기조절학습능력 변화에 미치는 영향. 특수교육학연구, 44(3), 327-353.
허은영, 유현실(2012). 서울시 중학생의 학습부진에 대한 방과후 학습활동 및 심리적 변인의 영향. 학습장애연구, 9(3), 37-52.
홍유정, 박현정(2021). 기초학력 미달 학생 예측을 위한 심층인공신경망 모형의 활용: 불균형 자료의 예측모형 구축을 중심으로. 교육평가연구, 34(4), 615-647.
홍종선, 장동환(2020). ROC 곡선 길이의 판단기준. 한국데이터정보과학회지, 31(5), 851-863. DOI: 10.7465/jkdi.2020.31.5.851
Ahsan, M. M., Mahmud, M. A., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of data scaling methods on machine learning algorithms and model performance. Technologies, 9(3), 52. DOI: 10.3390/technologies9030052
Artzi, M., Redmard, E., Tzemach, O., Zeltser, J., Gropper, O., Roth, J., Shofry, B., Kozyrew. D. A., Constantini, S., & Ben-Sira, L. (2021). Classification of pediatric posterior fossa tumors using convolutional neural network and tabular data. IEEE, 9, 91966-91973. DOI: 10.1109/ACCESS.2021.3085771

Baranyi, M., Nagy, M., & Molontay, R. (2020). Interpretable deep learning for university dropout prediction. In Proceedings of the 21st Annual Conference on Information Technology Education, 13-19. DOI: 10.1145/3368308.3415382

Barros, T. M., Neto, S., Plácido, A., Silva, I., & Guedes, L. A. (2019). Predictive Models for Imbalanced Data: A School Dropout Perspective. Education Sciences, 9(4), 275. DOI: 10.3390/educsci9040275

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2013). A review of feature selection methods on synthetic data. Knowledge and information systems, 34(3), 483-519. DOI: 10.1007/s10115-012-0487-8

Bouten, M., Schietse, J., & Van den Broeck, C. (1995). Gradient descent learning in perceptrons: A review of its possibilities. Physical Review E, 52(2), 1958.

Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32. DOI: 10.1023/A:1010933404324

Butkowsky, I. S., & Willows, D. M. (1980). Cognitive-motivational characteristics of children varying in reading ability: Evidence for learned helplessness in poor readers. Journal of Educational Psychology, 72(3), 408-422. DOI: 10.1037/0022-0663.72.3.408.

Cai, J., Luo, J., Wang, S., & Yang, S. (2018). Feature selection in machine learning: A new perspective. Neurocomputing, 300, 70-79. DOI: 10.1016/j.neucom.2017.11.077

Castro, C. L., & Braga, A. P. (2013). Novel cost-sensitive approach to improve the multilayer perceptron performance on imbalanced data. IEEE transactions on neural networks and learning systems, 24(6), 888-899.

Chaplot, D. S., Rhim, E., & Kim, J. (2015). Predicting student attrition in

MOOCs using sentiment analysis and neural networks. 17th International Conference on Artificial Intelligence in Education (AIED), Madrid, Spain.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321-357. DOI: 10.1613/jair.953

Ciaburro, G., & Venkateswaran, B. (2017). Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles. Birmingham: Packt.

Davis, J., & Goadrich, M. (2006). The relationship between Precision-Recall and ROC curves. In Proceedings of the 23rd international conference on Machine learning, 233-240. ACM Press. DOI: 10.1145/1143844.1143874

Du, L., Gao, F., Chen, X., Jia, R., Wang, J., Zhang, J., Han, S., & Zhang, D. (2021). TabularNet: A Neural Network Architecture for Understanding Semantic Structures of Tabular Data. In Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 322-331. DOI: 10.1145/3447548.3467228

El Naqa, I., Li, R., & Murphy, M. J. (Eds.). (2015). Machine learning in radiation oncology: theory and applications. New York: Springer.

Fernández, A., García, S., del Jesus, M. J., & Herrera, F. (2008). A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets. Fuzzy Sets and Systems, 159(18), 2378-2398. DOI: 10.1016/j.fss.2007.12.023

Friedman J, Hastie T, Tibshirani R (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." Journal of

Statistical Software, 33(1), 1-22. DOI: 10.18637/jss.v033.i01, https://www.jstatsoft.org/v33/i01/.

Fu, G. H., Yi, L. Z., & Pan, J. (2019). LASSO-based false-positive selection for class-imbalanced data in metabolomics. Journal of Chemometrics, 33(10), 3177. DOI: 10.1002/cem.3177

García, S., Luengo, J., & Herrera, F. (2016). Tutorial on practical tips of the most influential data preprocessing algorithms in data mining. Knowledge-Based Systems, 98, 1-29. DOI: 10.1016/j.knosys.2015.12.006

Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F., & Pedreschi, D. (2018). A survey of methods for explaining black box models. ACM computing surveys (CSUR), 51(5), 1-42. DOI: 10.1145/3236009

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of machine learning research, 3, 1157-1182.

Hagan, M. T., Demuth, H. B., & Beale, M. (2018). 신경망 설계 (윤성진 역) (pp. 771-807). 서울: 에이콘. (원서출판 2014).

Hamer & Dupont. (2021). An Importance Weighted Feature Selection Stability Measure. Journal of Machine Learning Research, 22, 1-57.

Han H., Wang WY., Mao BH. (2005) Borderline-SMOTE: a new over-sampling method in imbalanced data sets learning. In Huang DS., Zhang XP., Huang GB. (Eds), Advances in Intelligent Computing, 878-887. Berlin: Springer. DOI: 10.1007/11538059_91

Hastie, T., Tibshirani, R., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (2nd ed.). New York: Springer.

He, H. & Ma, Y. (2013). Imbalanced Learning: Foundations, Algorithms,

and Applications, 1st ed. New York: Wiley-IEEE Press.

He, H. and Garcia, E. A. (2009). Learning from imbalanced data. IEEE Transactions on Knowledge and Data Engineering, 21, 1263-1284. DOI: 10.1109/TKDE.2008.239

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In 2008 IEEE international joint conference on neural networks, 1322-1328. DOI: 10.1109/IJCNN.2008.4633969

Henderson, G., & Bryan, W. V. (2004). Psychosocial aspects of disability. Illinois: Charles C Thomas Publisher.

Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554. DOI: 10.1162/neco.2006.18.7.1527

Hussna, A. U., Trisha, I. I., Ritun, I. J., & Alam, M. G. R. (2021). COVID-19 impact on students' Mental Health: Explainable AI and Classifiers. In 2021 International Conference on Decision Aid Sciences and Application (DASA) (pp. 847-851). IEEE. DOI: 10.1109/DASA53625.2021.9682371

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning. New York, NY: Springer.

Jeni, L. A., Cohn, J. F., & De La Torre, F. (2013). Facing imbalanced data-recommendations for the use of performance metrics. 2013 Humaine association conference on affective computing and intelligent interaction, 245-251. DOI: 10.1109/ACII.2013.47

Johnson, J. M., & Khoshgoftaar, T. M. (2019). Survey on deep learning with class imbalance. Journal of Big Data, 6(1), 1-54. DOI: 10.1186/s40537-019-0192-5

Joseph, M. P. (2005). A PD validation framework for Basel II internal ratings-based systems. Credit Scoring and Credit Control XI.

Khosravi, H., Shum, S. B., Chen, G., Conati, C., Tsai, Y. S. , Kay, J., Knight, S., Martinez-Maldonado, R., Sadiq, S., Gašević, D. (2022). Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence, 3, 100074. DOI: 10.1016/j.caeai.2022.100074

Li, Y., Chen, C. Y., & Wasserman, W. W. (2016). Deep feature selection: theory and application to identify enhancers and promoters. Journal of Computational Biology, 23(5), 322-336. DOI: 10.1089/cmb.2015.0189

Longadge, R., & Dongre, S. (2013). Class imbalance problem in data mining review. arXiv preprint arXiv:1305.1707.

Lunardon, N., Menardi, G., & Torelli, N. (2014). ROSE: a package for binary imbalanced learning. R journal, 6(1). DOI: 10.32614/RJ-2014-008

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. Advances in neural information processing systems, 30.

Mercer, C. D., Mercer, A. R., & Pullen, P. C. (2010). Students with learning disabilities (8th Ed.). London: Merrill/Prentice-Hall.

Mienye, I. D., & Sun, Y. (2021). Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. Informatics in Medicine Unlocked, 25, 100690. DOI: 10.1016/j.imu.2021.100690

Minsky, M., & Papert, S. A. (2017). Perceptrons: An introduction to computational geometry. Massachusetts: MIT.

Mitchell, T. M. (1999). Machine learning and data mining. Communications of the ACM, 42(11), 30-36.

Molnar. C. (2019). Interpretable Machine Learning: A Guide For Making Black Box Models Explainable. [Online]. Available: https://arxiv.

org/abs/1804.06620

Nabil, A., Seyam, M., & Abou-Elfetouh, A. (2021). Prediction of Students' Academic Performance Based on Courses' Grades Using Deep Neural Networks. 9, 140731-140746. DOI: 10.1109/ACCESS.2021.3119596

Nagy, M., & Molontay, R. (2018). Predicting dropout in higher education based on secondary school performance. In 2018 IEEE 22nd international conference on intelligent engineering systems (INES), 000389-000394. DOI: 10.1109/INES.2018.8523888

Naqa, I. E., Li, R., & Murphy. M. J. (2015). Machine learning in radiation oncology. Switzerland: Springer. DOI: 10.1007/978-3-319-18305-3

Oommen, T., Baise, L. G., Vogel, R. M. (2011). Sampling bias and class imbalance in maximum-likelihood logistic regression, Mathematical Geosciences, 43(1), 99-120. DOI: 10.1007/s11004-010-9311-8

Owen, A. B. (2007). Infinitely imbalanced logistic regression, The Journal of Machine Learning Research, 8, 761-773.

Priddy, K. L., & Keller, P. E. (2005). Artificial neural networks: an introduction . SPIE press. DOI: 10.1117/3.633187

Rachburee, N., & Punlumjeak, W. (2021). Oversampling technique in student performance classification from engineering course. International Journal of Electrical & Computer Engineering , 11(4), 2088-8708. DOI:10.11591/ijece.v11i4.pp3567-3574

Rasamoelina, A. D., Adjailia, F., & Sinčák, P. (2020). A review of activation function for artificial neural network. In IEEE World Symposium on Applied Machine Intelligence and Informatics (SAMI), 281-286. DOI: 10.1109/SAMI48414.2020.9108717

Reis, S. M. (2007). The underachievement of gifted students: Multiple

frustrations and few solutions. Kay, K., Robson, D., & Brenneman, J. F. (Eds), High IQ kids: Collected insights, information, and personal stories from the experts, 124-142. Minnesota: Free Spirit.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. Psychological review, 65(6), 386-408. DOI: 10.1037/h0042519

Sankhe, P., Khabiri, E., Agrawal, B., & Li, Y. (2021). TableNN: Deep Learning Framework for Learning Domain Specific Tabular Data. In IEEE International Conference on Big Data, 4097-4102. DOI: 10.1109/BigData52589.2021.9671972

Shapley, L. S. (1953), A Value for n-person Games. Tucker, A. W., H. Kuhn (Eds.), Contributions to the Theory of Games II(pp. 307-317), Princeton University Press. DOI: 10.1515 /9781400881970-018

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In International conference on machine learning, 3145-3153.

Siriseriwan, W. (2019). A collection of oversampling techniques for class imbalance problem based on SMOTE. https://rdrr.io/cran/ smotefamily/

Snieder, E., Shakir, R., & Khan, U. T. (2019). A comprehensive comparison of four input variable selection methods for artificial neural network flow forecasting models. Journal of Hydrology, 538, 124299. DOI: 10.1016/j.jhydrol.2019.124299

Sun, K., Huang, S. H., Wong, D. S. H., & Jang, S. S. (2016). Design and application of a variable selection method for multilayer perceptron neural network with LASSO. IEEE transactions on neural networks and learning systems, 28(6), 1386-1396. DOI:

10.1109/TNNLS.2016.2542866

Tang, Y., Zhang, Y. Q., Chawla, N. V., Krasser, S. (2008). SVMs modeling for highly imbalanced classification, IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 281-288. DOI: 10.1109/TSMCB.2008.2002909

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society, 58(1), 267-288. DOI: 10.1111/j.2517-6161.1996.tb02080.x

Wang, Q., Luo, Z., Huang, J., Feng, Y., Liu, Z. (2017). A novel ensemble method for imbalanced data learning: bagging of extrapolation-SMOTE SVM, Computational Intelligence and Neuroscience, 1-11. DOI: 10.1155/2017/1827016

Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. Journal of the American statistical association, 58(301), 236-244. 10.2307/2282967

Weng, C. G., & Poon, J. (2008). A new evaluation measure for imbalanced datasets. In Proceedings of the 7th Australasian Data Mining Conference, 87, 27-32. DOI: 10.1109/ijcnn.2011.6033267

Wu. G., & Chang, E. Y. (2005). "KBA: Kernel boundary alignment considering imbalanced data distribution". IEEE Transactions on Knowledge and Data Engineering, 17(6), 786-795. DOI: DOI: 10.1109/TKDE.2005.95

Yan, Y., Chen, M., Shyu, M. L., & Chen, S. C. (2015). Deep learning for imbalanced multimedia data classification. In IEEE international symposium on multimedia (ISM), 483-488. DOI: 10.1109/ISM.2015.126

Zhang, H., Wang, J., Sun, Z., Zurada, J. M., & Pal, N. R. (2019). Feature selection for neural networks using group lasso

regularization. IEEE Transactions on Knowledge and Data Engineering, 32(4), 659-673. DOI: DOI: 10.1109/TKDE.2019.2893266

Zhang, J. W., Lu, H. J., Chen, W. T., & Lu, Y. (2011). A comparison study of cost-sensitive learning and sampling methods on imbalanced data sets. Advanced Materials Research, 271, 1291-1296. DOI: 10.4028/www.scientific.net/AMR.271-273.1291

Zhao, P., & Yu, B. (2006). On model selection consistency of Lasso. The Journal of Machine Learning Research, 7, 2541-2563.

Zhou, Z. H., & Liu, X. Y. (2005). Training cost-sensitive neural networks with methods addressing the class imbalance problem. IEEE Transactions on knowledge and data engineering, 18(1), 63-77. DOI: 10.1109/TKDE.2006.17

부 록

1. 주요 설명변인 분석 결과(상위 31개 이후)

<표 부록 1-1> 중2 국어 학업성취도 주요 설명변인(상위 31~91개)

순위	설문 대상	영 역	변인명(문항내용)	계수
31	학교	교육활동	SCH수행평가비율(국어)	0.028
32	학생	학교생활과 적응	규칙준수(화장실이나 급식실에서 차례를 잘 지킨다)	0.027
33	학교	학교크기	학교크기(학생수)	0.027
34	학부모	학교교육	학교만족도(기초능력 향상)	0.027
35	학생	비인지적 특성	자아개념(다른 사람들과 어울리기를 좋아한다)	-0.026
36	학생	여가 및 방과후 시간	선행학습 정도(수학)	0.025
37	학생	학교의 교육 및 환경	수업방식(모르는 것을 질문하면 친절하게 설명해 주신다)	0.024
38	학부모	교육기대 및 계획	PA진학계획(국/공립대)	0.023
39	학교	학교정책 및 운영	SCH혁신학교 운영	0.022
40	학생	부모-자녀 관계	학업지원(평소 생활을 확인하고 일정을 관리하신다)	-0.022
41	학생	여가 및 방과후 시간	여가활동(방과후학교에 참여한다)	0.021
42	학생	비인지적 특성	직업관(힘든 일이라도 내가 좋아하는 일이라면 선택할 것이다)	0.021
43	학생	학교생활과 적응	학폭경험(SNS나 블로그 등에서 욕설이나 비난을 들었다)	-0.021
44	학생	수업참여	수업태도(수업시간에 졸거나 잔다)	-0.021

순위	설문 대상	영 역	변인명(문항내용)	계수
45	학생	학교생활과 적응	학교 의사결정(학생들에게 학교규칙/규정과 변경사항을 알려 준다)	0.020
46	학생	학교생활과 적응	규칙준수(학급 규칙을 잘 지킨다)	0.020
47	학생	학교생활과 적응	진로프로그램 도움정도(고등학교 또는 대학교 방문 및 학과체험)	-0.020
48	학생	비인지적 특성	공동체의식(모둠활동(조별학습)을 수행하는 것이 즐겁다)	-0.019
49	학부모	부모-자녀 관계	부모와의 상호작용(학교 활동에 참가-부모 창며 체험활동, 자녀 운동회 등)	0.019
50	학교	교육활동	SCH수행평가비율(수학)	0.018
51	학생	부모-자녀 관계	소외(나는 부모님이 생각하시는 것보다 더 많은 스트레스를 받고 있다)	-0.017
52	학생	비인지적 특성	공동체의식(나의 학습자료나 필기도구를 다른 친구들에게 빌려준다)	0.017
53	학생	수업참여	수업분위기(수업에 집중한다)	0.017
54	학교	학교배경	SCH특별시	0.017
55	학생	여가 및 방과후 시간	문화활동(미술관, 박물관에 간다)	0.016
56	학생	비인지적 특성	공동체의식(소외된 친구가 있으면 내가 먼저 따뜻하게 대해 준다)	0.016
57	학생	비인지적 특성	자기관리(숙제보다 TV를 보거나 노는 것을 먼저 하는 경우가 많다)	-0.016
58	학부모	교육기대 및 계획	PA진학계획(자사고)	0.016
59	학생	학교생활과 적응	학교내폭력정도(욕설을 하는 학생이 있다)	0.016
60	학생	경제수준	월평균 가정소득	0.015
61	학생	비인지적 특성	정신건강(마음이 편하다)	0.014
62	학생	진학계획	진학계획(자사고)	0.014
63	학생	비인지적 특성	다문화수용성(나와 문화적 배경이 다른 사람을 이웃으로 받아들일 수 있다)	0.013

순위	설문 대상	영 역	변인명(문항내용)	계수
64	학생	수업참여	수업집중도(국어)	0.013
65	학부모	교육기대 및 계획	PA진학계획(특성화고)	-0.013
66	학생	부모-자녀 관계	상호작용(책을 읽고 나와 함께 대화를 나누신다)	0.013
67	학생	여가 및 방과후 시간	휴대전화중독(휴대전화가 없으면 불편해서 살 수가 없다)	0.012
68	학생	학교의 교육 및 환경	수업방식(나의 능력에 맞게 과제를 내 주신다)	0.012
69	학생	비인지적 특성	자기관리(내 책상을 스스로 정리하고 청소한다)	0.011
70	학생	학교생활과 적응	교사관계(나에 대해서 잘 알고 계신다)	0.010
71	학생	여가 및 방과후 시간	여가활동(학습활동 및 여가활동-5)텔레비전을 시청한다)	-0.010
72	학생	여가 및 방과후 시간	여가활동(학교 숙제를 한다)	0.010
73	학생	비인지적 특성	직업관(직업은 단지 돈을 벌기 위한 수단일 뿐이다)	-0.010
74	학교	학교정책 및 운영	진로상담교사 배치, 진로진학 상담실 운영	0.010
75	학생	진학계획	진학계획(예고/체고)	-0.009
76	학생	부모-자녀 관계	정서적지원(나를 격려하신다)	0.009
77	학생	여가 및 방과후 시간	선행학습정도(영)	0.008
78	학생	비인지적 특성	진로성숙도(학교 공부 중 어떤 것들은 나의 진로와 관련해 특별히 중요한 의미가 있다)	0.007
79	학생	비인지적 특성	다문화수용성(나와 문화적 배경이 다른 학생과 가장 친한 단짝친구가 될 수 있다)	0.007
80	학교	교육활동	학교구성원학교운영(학교구성원들이 학교의 비전과 목표에 대해 논의한다)	0.006
81	학생	비인지적 특성	공동체의식(어려운 교과 내용은 친구들과 상의하여 이해한다)	0.006
82	학생	학교생활과 적응	진로프로그램 도움정도진(모의직업체험)	-0.006

순위	설문 대상	영 역	변인명(문항내용)	계수
83	학생	비인지적 특성	자아개념(나의 몸매에 만족한다)	-0.005
84	학생	비인지적 특성	공동체의식(학교 공부가 뒤처진 친구들을 돕는다)	0.005
85	학생	비인지적 특성	자기관리(내일 학교에 가져가야 할 준비물을 스스로 챙긴다)	0.004
86	학생	여가 및 방과후 시간	미디어사용(학습 이외의 정보 검색 및 자료 이용)	0.004
87	학생	부모-자녀 상호작용	상호작용(내가 학교생활을 어떻게 하는지 관심 있게 물어 보신다)	0.003
88	학생	학교생활과 적응	학교내폭력정도(특정 학생을 고의로 따돌리는 일이 일어난다)	-0.002
89	학생	학교생활과 적응	학폭경험(친구나 선/후배에게 심하게 맞았다)	-0.002
90	학교	학교배경	SCH여학교	0.001
91	학생	학교생활과 적응	학업스트레스(공부가 잘 되지 않으면 스트레스를 받는다)	0.001

<표 부록 1-2> 중2 영어 학업성취도 주요 설명변인(상위 31~84개)

순위	설문 대상	영 역	변인명(문항내용)	계수
31	학부모	학교교육	학교만족도(기초능력 향상)	0.025
32	학생	여가 및 방과후 시간	미디어사용(문자, 채팅, 메신저, 이메일, 통화)	-0.025
33	학교	교육활동	동아리참여인원(자율동아리)	0.025
34	학생	여가 및 방과후 시간	여가활동(집안일을 돕는다)	-0.024
35	학생	비인지적 특성	자아개념(단체생활을 잘한다)	-0.023
36	학생	여가 및 방과후 시간	문화활동(미술관, 박물관에 간다)	0.021
37	학생	부모-자녀 관계	상호작용(책을 읽고 나와 함께 대화를 나누신다)	0.021
38	학부모	교육기대 및 계획	PA진학계획(국/공립대)	0.020
39	학생	비인지적 특성	공동체의식(나의 학습자료나 필기도구를 다른 친구들에게 빌려준다)	0.020
40	학생	부모-자녀 관계	소외(나는 부모님이 생각하시는 것보다 더 많은 스트레스를 받고 있다)	0.018
41	학교	학교정책 및 운영	학부모지원(교사의 보조 역할로 봉사한다)	0.018
42	학교	학교배경	SCH여학교	0.018
43	학생	학교생활과 적응	진로프로그램 도움정도(현장체험)	-0.017
44	학생	학교의 교육및 환경	수업방식(모르는 것을 질문하면 친절하게 설명해 주신다)	0.017
45	학생	비인지적 특성	직업관(직업은 단지 돈을 벌기 위한 수단일 뿐이다)	-0.017
46	학생	비인지적 특성	자아개념(나의 용모는 매력적인 편이다)	-0.017
47	학생	여가 및 방과후 시간	여가활동(학습활동 및 여가활동-5)텔레비전을 시청한다)	-0.016
48	학부모	교육기대 및 계획	PA진학계획(자사고)	0.016

순위	설문 대상	영 역	변인명(문항내용)	계수
49	학생	부모-자녀 관계	소외(부모님은 나에게 관심이 없다)	0.016
50	학부모	교육기대 및 계획	PA진학계획(과고/영재고)	0.015
51	학생	학교의 교육 및 환경	수업방식(나의 능력에 맞게 과제를 내 주신다)	0.014
52	학생	비인지적 특성	직업관(직업 자체가 내 인생에서 의미 있는 일은 아니다)	-0.013
53	학생	비인지적 특성	자기관리(오늘 해야 할 일을 미루지 않는다)	0.013
54	학생	비인지적 특성	다문화수용성(한국에 사는 외국 사람들도 우리나라 사람들과 똑같은 대우를 받아야 한다)	0.012
55	학교	학교정책 및 운영	SCH교육복지우선지원사업 운영	0.012
56	학교	학교배경	SCH중소도시	-0.012
57	학생	학교생활과 적응	교사특성(모든 학생이 과제를 다 해 오도록 강조하신다)	0.012
58	학교	학교정책 및 운영	SCH혁신학교 운영	0.011
59	학생	비인지적 특성	정신건강(죽고 싶은 생각이 든다)	-0.010
60	학생	학교생활과 적응	학폭경험(SNS나 블로그 등에서 욕설이나 비난을 들었다)	-0.010
61	학부모	방과후 교육활동	PA사교육참여이유(상급 학교 준비를 위해)	0.010
62	학부모	교육기대 및 계획	PA진학계획(국/공립 전문대학)	-0.010
63	학생	학교생활과 적응	교사관계(나에 대해서 잘 알고 계신다)	-0.009
64	학생	수업참여	수업분위기(시끄럽고 무질서하다)	-0.009
65	학생	학교생활과 적응	학교내폭력정도(욕설을 하는 학생이 있다)	-0.008
66	학교	학교정책 및 운영	진로상담교사 배치, 진로진학 상담실 운영	0.008
67	학부모	가정생활 및 교육지원	교육적지원(학부모들 모임에 나가 정보를 얻는다)	-0.007

순위	설문 대상	영 역	변인명(문항내용)	계수
68	학부모	가정생활 및 교육지원	교육적지원(우리 아이가 가고 싶어하는 고등학교 유형을 알고 있다)	0.007
69	학생	비인지적 특성	정신건강(마음이 편하다)	0.006
70	학생	진학계획	진학계획(외고/국제고)	0.006
71	학생	진학계획	진학계획(자사고)	0.006
72	학생	수업참여	수업분위기(수업에 집중한다)	0.005
73	학생	여가 및 방과후 시간	독서활동(다른 사람과 책에 대해 이야기하는 것을 좋아한다)	0.005
74	학교	교육활동	학교구성원학교운영활동(학교구성원들이 각자가 맡은 역량을 공유하며 서로 협력한다)	0.004
75	학생	학교생활과 적응	교우관계(내 친구들에게 내 고민을 털어 놓는다)	-0.004
76	학생	여가 및 방과후 시간	공부시간	0.003
77	학교	수업참여	수업분위기(과제를 잘 해 온다)	0.002
78	학생	비인지적 특성	공동체의식(모둠활동(조별학습)을 수행하는 것이 즐겁다)	-0.002
79	학부모	부모-자녀 관계	부모와의 상호작용(학교 활동에 참가-부모 창며 체험활동, 자녀 운동회 등)	0.002
80	학생	학교생활과 적응	규칙준수(화장실이나 급식실에서 차례를 잘 지킨다)	0.002
81	학생	비인지적 특성	자아개념(운동신경이 발달되어 있다)	-0.001
82	학교	수업참여	수업분위기(선생님 말씀을 제대로 듣지 않는다)	-0.001
83	학부모	교육기대 및 계획	PA진학계획(사립 전문대학)	-0.001
84	학생	여가 및 방과후 시간	문화활동(영화 보러 간다)	0.001

<표 부록 1-3> 중2 수학 학업성취도 주요 설명변인(상위 31~83개)

순위	설문 대상	영 역	변인명(문항내용)	계수
31	학생	학교생활과 적응	규칙준수(학급 규칙을 잘 지킨다)	0.023
32	학부모	학교교육	학교만족도(기초능력 향상)	0.023
33	학생	비인지적 특성	공동체의식(소외된 친구가 있으면 내가 먼저 따뜻하게 대해 준다)	0.022
34	학생	여가 및 방과후 시간	여가활동(학습활동 및 여가활동-5)텔레비전을 시청한다)	-0.022
35	학교	교육활동	SCH수행평가비율(수학)	0.021
36	학생	비인지적 특성	직업관(직업 자체가 내 인생에서 의미 있는 일은 아니다)	-0.021
37	학생	학교생활과 적응	진로프로그램 도움정도(현장체험)	-0.021
38	학생	부모-자녀 관계	상호작용(책을 읽고 나와 함께 대화를 나누신다)	0.021
39	학부모	교육기대 및 계획	PA진학계획(국/공립대)	0.019
40	학생	학교생활과 적응	학폭경험(친구나 선/후배에게 심하게 맞았다)	-0.019
41	학부모	가정생활 및 교육지원	교육적지원(우리 아이가 가고 싶어하는 고등학교 유형에 들어가기 위해 어느 정도의 성적이 필요한지 알고 있다)	0.018
42	학생	비인지적 특성	정신건강(죽고 싶은 생각이 든다)	-0.018
43	학생	비인지적 특성	직업관(어떤 직업을 가지는가가 미래의 나의 인생에 중요한 영향을 미친다)	0.016
44	학교	교육활동	SCH선다형비율(국어)	0.016
45	학부모	학교 교육	PA학교만족도(적성에 따른 진로지도)	-0.016
46	학교	교육활동	동아리수(자율동아리)	0.015
47	학생	여가 및 방과후 시간	공부시간	0.015
48	학생	비인지적 특성	창의성(다른 친구들이 생각하지 못하는 새로운 생각을 잘한다)	-0.015

순위	설문 대상	영 역	변인명(문항내용)	계수
49	학생	여가 및 방과후 시간	미디어사용(게임 및 오락)	-0.014
50	학교	학교통계	학교크기(학생 수)	0.014
51	학생	학교생활과 적응	교우관계(내 친구들에게 내 고민을 털어 놓는다)	-0.013
52	학생	비인지적 특성	다문화수용성(우리 학교에 여러 인종과 문화를 지닌 학생들이 들어와 학생들이 다양해지면 좋겠다)	0.013
53	학부모	학교 교육	PA학교만족도(자녀 수준에 맞는 교육)	0.011
54	학생	학교생활과 적응	교사특성(학생들이 열심히 공부하기를 원하신다)	0.011
55	학교	학교배경	SCH중소도시	-0.011
56	학생	진학계획	진학계획(고교해외유학)	-0.011
57	학생	여가 및 방과후 시간	독서활동(서점이나 도서관에 가는 것을 좋아한다)	0.011
58	학생	진학계획	진학계획(외고/국제고)	0.010
59	학부모	교육기대 및 계획	PA진학계획여부(외고/국제고)	0.010
60	학생	비인지적 특성	다문화수용성(한국에 사는 외국 사람들도 우리나라 사람들과 똑같은 대우를 받아야 한다)	0.009
61	학부모	교육기대 및 계획	PA진학계획(자공고)	0.009
62	학생	비인지적 특성	다문화수용성(외국인이 늘어나면 범죄가 더 많이 발생할 것이다)	-0.008
63	학생	학교생활과 적응	학교내폭력정도(욕설을 하는 학생이 있다)	-0.008
64	학생	비인지적 특성	자기관리(숙제보다 TV를 보거나 노는 것을 먼저 하는 경우가 많다)	-0.008
65	학생	비인지적 특성	다문화수용성(나와 문화적 배경이 다른 학생과 가장 친한 단짝친구가 될 수 있다)	0.008
66	학교	학교정책 및 운영	진로상담교사 배치, 진로진학 상담실 운영	0.008
67	학부모	교육기대 및 계획	PA진학계획(예고/체고)	-0.008

순위	설문 대상	영 역	변인명(문항내용)	계수
68	학생	학교의 교육및 환경	수업방식(모르는 것을 질문하면 친절하게 설명해 주신다)	0.008
69	학부모	가정생활 및 교육지원	PA교육적 지원(자녀의 학교 공부와 숙제를 확인한다)	-0.007
70	학교	학교크기	학교크기(학생수)	0.007
71	학생	학교생활과 적응	교사특성(모든 학생이 과제를 다 해 오도록 강조하신다)	0.006
72	학생	학교생활과 적응	학폭경험(SNS나 블로그 등에서 욕설이나 비난을 들었다)	-0.005
73	학생	학교생활과 적응	진로프로그램 도움정도(고등학교 또는 대학교 방문 및 학과체험)	-0.004
74	학생	학교생활과 적응	학폭경험(다른 학생들이 나를 놀렸다)	-0.004
75	학생	여가 및 방과후 시간	문화활동(미술관, 박물관에 간다)	0.004
76	학생	비인지적 특성	공동체의식(모둠활동(조별학습)을 수행하는 것이 즐겁다)	-0.003
77	학생	학교생활과 적응	교육프로그램 도움정도(수학)	-0.002
78	학생	부모-자녀 관계	소외(나는 부모님이 생각하시는 것보다 더 많은 스트레스를 받고 있다)	0.002
79	학교	교육활동	역량기반교육활동(교육과정 재구성)	-0.002
80	학교	교육활동	학교구성원학교운영활동(학교구성원들이 교사들의 수업을 참관함으로써 수업을 개선할 수 있도록 돕는다)	0.002
81	학생	비인지적 특성	자아개념(나는 편안한 마음으로 수업에 참여한다)	0.002
82	학생	학교생활과 적응	학교의사결정(학교 운영에 학생회 의견을 반영한다)	0.001
83	학생	학교생활과 적응	교사특성(시험에서 우리 반이 상위권에 들 것을 강조하신다)	-0.001

Abstract

Prediction of Low Academic Achievers Using a Deep Neural Network

Yujung Hong
Department of Education
The Graduate School
Seoul National University

Low academic achievers are students below basic academic ability, and their learning deficit tends to gradually accumulate and deepen over time. The learning deficit causes negative emotions about studies or makes it difficult to adapt to school life. Since the causes of learning deficits are very diverse and complex, individualized and preventive support are needed to provide practical help to low academic achievers. Therefore, if a system that can predict the future academic performance of individual students can be established, it will be helpful to provide educational support before problems arise or at an early stage.

However, the previous studies that predicted the academic achievement of low academic achievers focus on examining the relationship between educational context variables and academic

achievement rather than predicting individual students' academic achievement. This is because educational research aims to diagnose the current situation by examining major variables that affect academic achievement.

Also, among the studies that predicted academic achievement, there were few studies focused on low academic achievers. At the school or office of education level, many studies have investigated variables affecting the rate of low academic achievement, or analyzed students with below basic level by integrating with low-education students who were not below basic. One of the reasons for this is that low achievers are a very small proportion to the total number of students, and they are statistically imbalanced data. As a result, it is not easy to meet basic statistical assumptions, and it is difficult to secure the number of cases that can be used for the analysis.

In order to predict low academic achievers, a Deep Neural Network (DNN) model can be applied. The DNN does not assume a specific distribution of the data, but predicts each case by capturing the patterns within the analyzed data. In addition, it can prevent the omission of information on minority groups while training predictive models by sampling or cost-sensitive techniques.

Therefore, this study aimed to explore the DNN model that predicts low academic achievers belonging to a minority group. The education panel data that has been systematically collected over the years is used. Specifically, Korean Education Longitudinal Study 2013 was used, and the model for predicting low academic achievers in the subjects of Korean, English, and mathematics in the third year of middle school was detailed using students' information in the second year of middle school.

Although DNN is a method that can utilize many variables as

predictors, when variables not related to academic achievement are used, the performance of the predictive model can be lowered. Considering this, before actually constructing the DNN predictive model, we investigated the variables highly related to academic achievement among the 361 available variables. Based on these results, we tried to figure out the predictive models' performance by increasing the number of predictive variables.

Next, in order to construct the predictive model, the prediction performance were compared according to the methods of reflecting minority group information as much as the majority group during predictive model training and the DNN types. Random over sampling, Borderline-Synthetic Minority Over-sampling Technique (B-SMOTE), Adaptive Synthetic sampling approach (ADASYN), and cost-sensitive methods were applied to reflect minority group information. Multi-Layer Perceptron (MLP) and Convolutional Neural Network (CNN) were considered for DNN types. Based on the comparison results, the final predictive model for each subject was identified, and ways to improve the predictive models' performance were sought.

The main research results are as follows.

First, significant explanatory variables were identified using the Least Absolute Shrinkage and Selection Operator (LASSO), which can estimate the coefficients of each predictor while identifying the dependent variable among several variables. As a result of the analysis, it was found that 91 variables in Korean, 84 variables in English, and 83 variables in mathematics out of a total of 361 variables were related to academic achievement in each subject. However, as a result of checking the coefficients of the selected variables, it was confirmed that variables judged to have a low correlation with the dependent variables are still included. Therefore, a predictive model was

constructed by increasing the number of predictors centering on variables with high absolute values of coefficients.

Next, to predict low academic achievers in Korean, English, and mathematics, the DNN prediction model is implemented according to the number of predictors, the method of reflecting minority group information, and the type of DNN, and the results are as follows. First, when the number of predictor variables in all subjects increased by more than a certain level, the performance of the predictive model tended to decrease. In particular, a relatively significant decrease in performance was observed in Korean, where the degree of imbalance is large. This demonstrates the need to consider the selection of appropriate predictors when constructing the DNN prediction model.

Second, the Original model, which did not adjust the degree of imbalance, showed inappropriate results for predicting low academic achievers. In Korean, almost all students were predicted to be non-low academic achievers. Relatively, when the DNN model was trained by reflecting the information of the minority group as much as the majority group, AUC, F2-score, G-mean, and specificity tended to increase. In addition, comparing the performance according to the method reflecting minority group information, the model applying the cost-sensitive method than the sampling method performed better overall. In particular, as the number of predictors increased to 49 or more, the difference between the sampling and cost-sensitive methods increased. Based on this, it is confirmed that it is necessary to apply a method to adjust the degree of imbalance when predicting imbalanced data, and it is suggested that a cost-sensitive method can be applied when implementing a prediction model using unbalanced data.

Third, the prediction performance of the MLP model was better than that of the CNN model, but it was difficult to judge that the difference

was considerable. However, in the CNN model, regardless of sampling and cost-sensitive method, the model's performance was less degraded even if the number of predictors increased. Therefore, the CNN model can be considered in a situation where many predictors are used, but considering the characteristics of the panel data and the difficulty of setting the CNN model, it was discussed that the MLP model could be applied preferentially.

Finally, the MLP model, which used 25 predictors in all subjects and applied the cost-sensitive method, showed the best performance. However, when the accuracy was 0.704 to 0.768, the sensitivity was 0.703 to 0.762, so it was not easy to judge that the prediction performance was sufficient to use the prediction model. Therefore, based on this, we made suggestions for the future use and research of the DNN model, along with matters to be considered in the DNN model using panel data.

keywords : low academic achiever, prediction of academic performance, education panel data, imbalanced data, deep neural network

Student Number : 2018-37446